Pattern Recognition 60 (2016) 499–514

Contents lists available at ScienceDirect

Pattern Recognition

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

Object recognition using local invariant features for robotic applications: A survey

Patricio Loncomilla, Javier Ruiz-del-Solar*, Luz Martínez

Advanced Mining Technology Center & Dept. of Electrical Engineering, Universidad de Chile, Chile

ARTICLE INFO

Article history:
Received 2 February 2016
Received in revised form
15 March 2016
Accepted 11 May 2016
Available online 24 May 2016

Keywords:
Local invariant features
Object recognition
Local descriptors
Local interest points

ABSTRACT

The main goal of this survey is to present a complete analysis of object recognition methods based on local invariant features from a robotics perspective; a summary which can be used by developers of robot vision applications in the selection and development of object recognition systems. The survey includes a brief description of the main approaches reported in the literature, with more specific analyses of local interest point computation methods, local descriptor computation and matching methods, and geometric verification methods. Different methods are analyzed by considering the main requirements of robotics applications, such as real-time operation with limited on-board computational resources, and constrained observational conditions derived from the robot geometry (e.g. limited camera resolution). In addition, various object recognition systems are evaluated in a service-robot domestic environment, where the final task to be performed by a service robot is the manipulation of objects. It can be concluded from the results reported that (i) the most suitable keypoint detectors are ORB, BRISK, Fast Hessian, and DoG, (ii) the most suitable descriptors are ORB, BRISK, SIFT, and SURF, (iii) the final performance of object recognition systems using local invariant features under real-world conditions depends strongly on the geometric verification methods being used, and (iv) the best performing object recognition systems are built using ORB–ORB and DoG–SIFT keypoint–descriptor combinations. ORB–ORB based systems are faster, while DoG–SIFT are more robust to real-world conditions.

1. Introduction

The recognition of objects under uncontrolled, real-world conditions is of paramount importance in robotics. Object recognition is an essential ability for building object-based representations of the environment, and for the manipulation of objects. In this work, object recognition refers to the recognition of a specific object instance (e.g. my cup), instead of a generic object class/category (e.g. a cup), which is usually called object categorization or generic object recognition. Both object recognition and object categorization are important abilities in robotics, and they are used for solving different tasks. This survey is focused on object recognition, and then the analysis of object categorization techniques (Bag of Visual Words [65], VLAD [21], FLAIR [59], cascaded ensembles of randomized decision trees [4], unsupervised segmentation of unknown objects [5]) is beyond its scope.

In recent years, several approaches to object recognition have been developed. They are usually based on global and/or local descriptions of the objects. Global description based methods model the appearance of an object as a whole, while local description based methods represent objects as a set of local interest points (keypoints), each of them represented by a local invariant feature[1] (or descriptor). Methods based on local features have advantages, such as not needing object segmentation, robustness against occlusions and against changes in the viewpoint (rotation and scale change), and having a near real-time recognition frame rate.

Object recognition methods based on the use of local invariant features have been developed mostly within the computer vision community, and then transferred to the robotics community. However, robot vision applications have different requirements than standard computer vision applications, such as the requirement of real-time operation with limited on-board computational resources, and the constrained observational conditions derived from the robot geometry, limited camera resolution, and sensor/object relative pose. In addition, in many cases the developers of robot vision applications adapt computer vision modules (e.g. the ones available in OpenCV [73]) to their robotics applications,

* Corresponding author.
E-mail addresses: ploncomi@ing.uchile.cl (P. Loncomilla), jruizd@ing.uchile.cl (J. Ruiz-del-Solar), luz.martinez@amtc.cl (L. Martínez).

[1] In this work, the focus is on appearance-based keypoints and features; 3D feature descriptors such as feature histograms obtained from range images are not considered.

http://dx.doi.org/10.1016/j.patcog.2016.05.021
0031-3203/

without analyzing the specific characteristics of the methods and the applications' requirements carefully. Only general criteria, such as "SURF is faster than SIFT", are applied.

In this context, the main motivation of this survey is to present a complete analysis of object recognition methods based on local invariant features from a robotics perspective, which can be used by developers of robot vision applications in the selection and development of object recognition systems. The survey analyzes the main functionalities of popular methods (local interest point computation, local descriptors computation and matching, geometric verification), and presents evaluations in terms of accuracy, robustness and efficiency.

Previous studies have analyzed the performance of object recognition approaches based on local features without considering the full requirements of robotics applications. For instance, in [68] the authors focused on the analysis of the precision of the methods regarding viewpoint angle, scale and affine transformations, but without considering the main robot vision requirements, such as real-time operation. In [50], several interest point detectors are compared, and their runtime and accuracy are evaluated for several image resolutions. However, real world problems, such as changes in illumination, background and partial occlusions are not analyzed. In [36], six object recognition algorithms based on local descriptors are compared in an object recognition task (recognizing objects on a table). Real-world conditions are included in that comparison. The results obtained are included in this survey, and the experiments they performed are extended.

This survey includes a brief description of the main approaches described in the literature, with specific analyses of local interest point computation methods, local descriptor computation and matching methods, and geometric verification methods. In addition, comparisons of the applicability of the methods in robotic applications, based on their accuracy, robustness, and efficiency, are presented.

The use of RGB-D sensors is very popular in the robotics community, and it could be wrongly assumed that the use of local invariant visual features is less relevant than the use of 3D range derived features. This assumption is really incorrect, because (i) the use of local invariant visual features is complementary to the use of 3D range features, (ii) standard RGB-D sensors do not work properly in outdoors and/or when observing black surfaces, restricting their applicability, and (iii) 3D range features require a much higher resolution than visual features to recognize objects, therefore their use impose constraints in robotics applications, e.g. objects can be recognized only at short distances (see the detailed analysis in [36]).

It is also important to explain why this survey does not include object methods based on deep learning. Although object recognition based on the use of the deep learning paradigm is a hot topic in the computer vision community, and its use in robotics applications will increase in the near future, still most of these methods are not able to fulfill the main requirements of robotics applications (real-time operation with limited on-board computational resources). Certainly, the use of both object recognition paradigms (local invariant features and deep-learning) will complement each other in the near future.

This paper is organized as follows: in Section 2, the paradigm of object recognition through the use of local features is presented. In Section 3, several interest point detection algorithms including corner-based and blob-based variants are described. In Section 4, algorithms for computing local descriptors and the standard procedure for matching descriptors are explained. In Section 5, algorithms for finding geometric verification of the matched features are described. In Section 6, a comparison of several object recognition systems in a real robot application is presented. Finally, in Section 7 some conclusions are drawn.

This survey intends to be a guide for developers of object recognition systems for robotics applications. The reader interested in having a practical guide for the use/application of the different algorithms/methods, and not just a description of them, is referred to (sub) Sections 2, 3.4, 4.2, 4.3, 5.3 and 6.

2. Object recognition using local invariant features

A local feature is "an image pattern which differs from its immediate neighborhood" [68]. A local interest point, also called a keypoint, defines the position of a local feature, and a descriptor describes/represents its image pattern. Therefore, the interest points are first searched for in the image under analysis, and then the regions around the interest points are described by the descriptors.

In general terms, object recognition based on local invariant features works according to the following principle: (i) keypoints are extracted independently from both a test image and a reference image (model), and characterized using invariant descriptors, and (ii) the invariant descriptors (features) are matched with each other. Afterwards, (iii) geometric verifications of the matched features are carried out using different procedures. For instance, whether or not the matched features satisfy a similarity or an affine transformation is tested.

The object recognition pipeline includes the following stages (see Fig. 1):

- Object segmentation (optional): In case a depth image I_D is available, objects that are on a planar surface, such as a table or the floor, can be isolated/segmented, and the object recognition method can be applied to only the segmented area.
- Local Interest Point Computation: Interest points (keypoints) are computed from the image I_{RGB} under analysis.
- Descriptors Computation: Local image descriptors are computed around each keypoint. In some methods more than one descriptor can be computed for each keypoint, depending on the local gradients' characteristics.
- Matching: Local descriptors belonging to the image under analysis I_{RGB} and to reference images (training descriptors) are

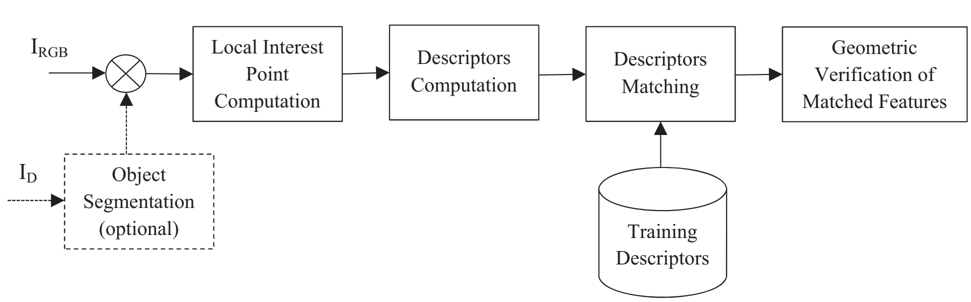

Fig. 1. Pipeline used for Recognizing Objects by using Local Invariant Features. The object can be segmented by using the depth image. Then, interest points and descriptors are computed and compared against those in a database. Extra verifications can be performed for rejecting incorrect detections.