Robotics and Autonomous Systems 60 (2012) 1279–1294

Contents lists available at SciVerse ScienceDirect

Robotics and Autonomous Systems

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

An automated vision based on-line novel percept detection method for a mobile robot

Xiaochun Wang *, Xia Li Wang, D. Mitchell Wilkes

Xian Jiaotong University, People's Republic of China
Changan University, People's Republic of China
Vanderbilt University, USA

ARTICLE INFO

Article history:
Received 11 June 2011
Received in revised form
14 June 2012
Accepted 23 June 2012
Available online 6 July 2012

Keywords:
On-line novelty detection
Computer vision
Statistical approaches
Classification tree
Clustering

ABSTRACT

It is generally agreed that learning, either supervised or unsupervised, can provide the best possible specification of known classes and offer inference for outlier detection by a dissimilarity threshold from the nominal feature space. Novel percept detection can take a step further by investigating whether these outliers form new dense clusters in both the feature space and the image space. By defining a novel percept to be a pattern group that has not been seen before in the feature space and the image space, in this paper, a non-conventional approach is proposed for multiple-novel-percept detection problem in robotic applications. Based on a computer vision system inspired loosely by neurobiological evidence, our approach can work in near real time for highly sparse high-dimensional feature vectors extracted from image patches while maintaining robustness to image transformations. Experiments conducted in an indoor environment and an outdoor environment demonstrate the efficacy of our method.

1. Introduction

Because of our persistent interest in novelty, the human species has been able to succeed in this world. Since a machine learning system can only be trained on data already observed, the ability to continue learning new or novel objects is important in any signal classification scheme, and novelty detection is considered as a key component of any intelligent systems. Being able to identify new or unknown data which a learning system is not aware of during training, novelty detection has a variety of applications, including fault detection, radar target detection, detection of masses in mammograms, statistical process control, etc. [1,2]. Since there is no single best model for every situation, novelty detection has been regarded as a very challenging task.

Typical solutions to this problem aim to model normal data and determine the abnormality by using a similarity measure and a threshold. Generally speaking, the novelty threshold is regarded as one of the most critical parameters in novelty detection tasks and directly affects the false negative rate. Two broad categories of techniques used for novelty detection are statistical approaches [1] and neural network approaches [2]. In statistical approaches, statistical properties of data are employed to generate a model, which can then be used to estimate whether a test sample belongs to the same distribution or not. Generally, there are two main statistical modeling techniques, parametric modeling, which assumes that the data come from a family of known distributions, and nonparametric modeling, in which no assumption on the form of the data distribution is made. The former is fairly efficient in practice, but may often be limited by assumptions about the distributions, while the latter is more flexible but often more computationally expensive, especially for large data sets. Neural-network-based approaches make no *a priori* assumptions on the properties of data and demand a relatively small number of parameters for optimization. However, neural networks have increased computational complexity in training and typically cannot be retrained to learn new units with computational efficiency in both time and space comparable to that of statistical models, particularly when the dimensionality of the data is high.

In this paper, we implement a fast approximate nearest neighbor search tree based novelty filter. First, it identifies novel samples using a statistically obtained threshold. Next, these samples are accumulated and evaluated whether they represent truly novel patterns and form dense clusters in both the feature space and the image space. By this way, the novelty detection problem becomes the identification of novel percepts in the segmented images. Finally, our on-line novelty detection mechanism completes the learning process by inserting the verified novel samples into the

* Corresponding author at: Xian Jiaotong University, People's Republic of China.
E-mail addresses: xiaocchunwang@mail.xjtu.edu.cn, xcawang@yahoo.com (X. Wang), xlwang@chd.edu.cn (X.L. Wang), mitch.wilkes@vanderbilt.edu (D.M. Wilkes).

0921-8890/$ – see front matter
doi:10.1016/j.robot.2012.06.005

search tree of image vocabulary percepts. One important contribution of the proposed method is its significant efficiency in the tree construction over Vocabulary tree for image-patch-based highly sparse high dimensional data while maintaining reasonably good search capability. Another important contribution of the proposed method is the automated strategy for setting the novelty threshold. Finally, any intelligent systems, particularly autonomous ones, have to incorporate a method for automatically learning newly discovered objects. The proposed novelty detection strategy incorporates such capability by using spatial information to cluster the detected novel samples into different categories and by inserting them into the tree on-line to allow for intuitive incremental updating of the knowledge base. This third contribution is an advantage over a preliminary work of this approach presented in [3], in which only a single novel object was meant to be detected.

To demonstrate the soundness of this manuscript, the performance of our proposed filter is compared with that of the well-known Grow-When-Required (GWR) neural network approach for a novelty detection task in an indoor environment, and with that of efficient support vector data description method (E-SVDD) for a novelty detection task in an outdoor environment. While currently we focus on research aspects, the ultimate goal of the research presented here is to relieve human operators by equipping autonomous robots with such automatic inspection capability.

In the following, Section 2 gives an account of related work in visual novelty detection. Section 3 introduces our image-patch-based very high dimensional perceptual learning system. Next, in Section 4, the proposed on-line novel object detection method is described. Experiments and results are presented in Section 5. Finally, in Section 6, conclusions are made and future work is outlined.

2. Related work

Novel object detection is an important ability with numerous applications in both robotics and computer vision. Model selection in novelty detection is a very important research topic due to the lack of examples from the novel targets [4] and requires a synthesis of techniques from computer vision, machine learning and pattern recognition. Therefore, work related to the methods presented in this paper falls into three main categories: visual object learning methods, object recognition methods and novel object detection methods.

2.1. Visual object learning methods

Object discovery using vision is a high-level learning task and can be pursued at different levels of visual properties such as appearance and shape etc. Since recognizing an object by its shape is computationally much more expensive, the goal of this research is to represent object by learning appearance features of objects, such as color and texture (the so-called percept in this paper), in the form of local descriptors which are extracted from image patches and form the basis of many appearance-based object discovery systems.

In order to understand and manipulate the world, labeling local descriptors extracted from training images to target values through machine learning methodology is often a major task in the building of intelligent systems. Recently, because of DARPA sponsored projects, machine learning based approaches, such as supervised learning, semi-supervised learning, self-supervised learning and unsupervised learning (also known as clustering in data mining literature), have gained prominence. Supervised learning uses hand-labeled data for offline training, which limits the scope of the robot's expertise to environments seen during training. To improve this, Teichman and Thrun used tracking-based semi-supervised learning as an effective method to train object recognition systems with a very small amount of hand-labeled data [5]. Veloso et al. proposed a method in which the robot learns to identify objects by seeing people interact with the objects in known tasks [6]. Dahlkamp et al. used self-supervised learning to train a vision-based road detection system for autonomous driving on labels provided by a laser range finder, significantly extending the range of road detection while requiring no extra human supervision [7]. Triebel et al. used unsupervised learning technique to locate similar objects in dense 3D point clouds [8]. To ease the labeling task by grouping unlabeled examples, many approaches have been developed for unsupervised learning, including hierarchical, partition-based, density-based, model-based, and graph-based approaches, and their use strongly depends on the application. An extensive study and comparison of the state-of-the-art unsupervised object discovery techniques is reported in [9]. As a graph-based approach, MST-based clustering algorithms can outperform the classic methods such as k-means clustering when the boundaries of clusters are irregular. To this end, a fast minimum spanning tree based clustering algorithm was proposed in [10], with impressive learning and segmentation results for challenging high-dimensional data obtained from an indoor environment [11]. We use this method in this research.

2.2. Visual object recognition methods

Matching an object model against object data from some other source, object recognition capabilities are a fundamental limitation of many practical robotic systems, particularly when recognition is performed by nearest neighbor search in the presence of a large database of object models. To support efficient k-nearest neighbor (kNN) search, multidimensional index structures, especially various types of tree-based index structures, have been developed to reduce the computation cost.

The best-known and most widely used of these is the K-dimensional tree (Kd-tree [12]), which was proposed in 1979 as a generalization of a binary search tree to higher dimensions with a logarithmic time complexity. To be more efficient, Beis and Lowe proposed to find an approximate neighbor by examining at most Emax leaves [13]. For high-dimensional data, Duch et al. proposed a randomized generalization of Kd-trees (the randomized Kd-trees) through randomization in selecting the partitioning value [14]. To further improve, Silpa-Anan and Hartly proposed a search method based on simultaneous independent searches in multiple randomized Kd-trees with different orientations [15]. However, all variants of Kd-trees have to satisfy two properties: (1) for every node with data point x and discriminate j, any data point y in the left subtree satisfies $y(j) < x(j)$ and any data point y in the right subtree satisfies $y(j) > x(j)$; (2) for any i, its ith attribute is different from the ith attribute of any other data point already in the tree.

Other tree-based index structures, such as R-tree [16], SS-tree [17] and SR-tree [18], rely on the amount of overlap among sibling regions for efficient search. They work well for low dimensional data sets, but dramatically worse for high dimensional large sized data sets.

As a very interesting alternative to kNN search algorithm, decision tree construction is a classic technique for classification of a set of data records of known classes. Standard splitting rules usually involve a single attribute (e.g., applying a threshold to obtain locally optimal decisions at each node) [19]. Branches from the root to the leaf nodes represent conjunctions of features that lead to those class labels. To capture the diversity and richness of high dimensional descriptors, Amit and Geman proposed to simply build multiple weakly dependent random trees of modest depth, each on a small random subset of training examples [20]. For higher dimensional data, Geurts et al.