Journal of Visual Languages and Computing 28 (2015) 56–70

Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

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

View-independent object detection using shared local features ☆

ByoungChul Ko *, Ji-Hun Jung, Jae-Yeal Nam

Department of Computer Engineering, Keimyung University Sindang-dong, Dalseo-gu, Daegu 704-701, Republic of Korea

ARTICLE INFO

Article history:
Received 1 October 2014
Received in revised form
4 November 2014
Accepted 19 December 2014
Available online 30 December 2014

Keywords:
Object detection
Shared local feature
View-independent
Part model
OCS-LBP
Random forest

ABSTRACT

In this study, we developed a novel method for detecting view-independent objects in a cluttered background with partial occlusion using shared features. These shared features are selected as common features among classes while the detectors used for each class are trained jointly rather than independently using shared features, which reduces the number of classifiers. We developed an exhaustive greedy selection method for selecting shared features and training their classifiers using only the shared features. The exhaustive greedy selection method randomly selects an exhaustive set of rectangular local features in a normalized object window and selects n significant shared local features from 12 different viewpoints and their effective shared classifiers using random forests. An integral histogram based on oriented-center symmetric local binary pattern (OCS-LBP) descriptor is used to represent a shared feature and to reduce the feature dimensions effectively. The final score is summed bilinearly using the probabilities of neighboring views to determine the location and viewpoint of the object because each view overlaps with neighboring views. Our proposed algorithm was successfully applied to the PASCAL VOC 2012 dataset and its detection performance was better than other methods.

1. Introduction

One of the main issues affecting computer vision is the detection and recognition of different types of objects. Thus, various applications have been developed for image retrieval, video surveillance, object tracking, augmented reality, and human-computer interaction. Although several object detection tasks such as face and human detection have yielded successful results, object detection remains a considerably difficult issue in the real world, which is typically cluttered, because objects have wide variations in their poses, colors, textures, and shapes, as well as extrinsic variations in lighting, viewpoints, and occlusion.

Conventional object detection methods recognize objects by considering full body and single degree viewpoints using color histograms [1], histograms of oriented gradients (HOGs) [2], and shape features [3]. However, these methods have several limitations in a cluttered background with partial occlusions; therefore, part-based models and local feature-based approaches [5–11] [25–26] have been proposed. Part-based models [4] [13–19] deliver a better object detection performance than a single model. In this type of model, objects are represented as a collection of parts arranged in a deformable configuration, and the part-based models can capture significant variations in appearance. In research on part-based models, different parts are used to capture the local-appearance properties of an object, whereas the deformable configuration of the object is characterized by spring-like connections between specific pairs of parts. In contrast, a single (global) model is often insufficiently expressive to represent a rich object category [4].

☆ This paper has been recommended for acceptance by Shi Kho Chang
* Corresponding author. Tel.: +82 53 580 5235; fax: +82 53 580 6275.
E-mail addresses: niceko@kmu.ac.kr (B. Ko), jihunjung@kmu.ac.kr (J.-H. Jung), jynam@kmu.ac.kr (J.-Y. Nam).

http://dx.doi.org/10.1016/j.jvlc.2014.12.006
1045-926X/

Object detection methods based on part-based models have the following advantages.

- Part models are intrinsically robust against partial inter-object occlusions [5].
- Part models exclude most of the background in the detection window, which avoids any confusion caused by changes in the background [6].
- Although the offline learning of part models requires a specific amount of computational time and training samples to extract discriminative information, object detection can be performed in real time using the trained part-based models.

In a similar manner, local feature-based methods build class-specific clusters of local features with a similar appearance, which are then treated as object parts and combined spatially in a probabilistic manner [7]. Leibe et al. [8] presented a two-stage object detection approach. In the first stage, a codebook of local appearances is learned, which contains information about the local structures that may appear in objects in the target category. Next, an implicit shape model is learned to specify where the codebook entries could occur on the object. Laptev [9] introduced an object detection method for single views by combining AdaBoost learning with a local histogram feature. In this method, an exhaustive set of rectangular regions in the normalized object window is selected, and then, AdaBoost is used to select the histogram features and learn the object classifier. Although these methods produce good detection results for single-viewpoint objects, they still have a few limitations such as detecting view-independent objects and computational runtime.

One problem affecting object detection is how to handle view independent of a class rather than through category-level object detection, which is concerned only with finding single views of an object, such as frontal and profile views of faces and cars. To detect view-independent objects, Torralba et al. [10] presented a multi-task learning procedure based on boosted decision stumps, which reduces the computational and sample complexity by finding common features. These common features can be shared across classes, and the detectors used for each class are trained jointly rather than independently using common features. Leibe et al. [11] introduced an object detection method that learns the appearance and spatial structure of a visual object category to recognize view-independent objects in that category, localizes them in cluttered real-world scenes, and segments them automatically from the background. However, these methods do not consider the deformation of objects. Uijlings et al. [12] designed a three-step object detector that first selects candidate bounding-box size and viewpoint, and then rely on a view-specific classifier to validate these hypotheses and decide whether an object is present.

A mixture of multi-scale deformable part models (DPMs) for view-independent object detection was proposed by Felzenszwalb et al. [4]. This method uses discriminative training with partially labeled data on individual side-views of the objects. Gu and Ren [13] use a mixture of holistic templates and discriminative learning for object viewpoint classification and category detection. Their research discriminatively incorporates a large mixture of templates inspired by the previous study [4] using HOGs, and shows that the templates, which are directly used for viewpoint classification, correspond well to the canonical views of an object. Further, López-Sastre et al. [14] revisited the DPMs and improved the accuracy of object category pose estimation by designing different training strategies using a semi-latent support vector machine (SVM) learning methodology. Gu et al. [15] produced visual clusters by considering multi-view components of the data, which are similar in their appearance and configuration spaces. The authors trained individual classifiers for each component and learned a second classifier, which operated at the category level by aggregating the responses from multiple components. Wang and Lin [16] presented a discriminative part-based model to represent and recognize object shapes using an And/Or graph. The And-Or graph model can handle large intra-class variances and background clutter during object shape detection in images. However, object detection based on DPMs requires additional computational time whenever part models are added to each viewpoint class.

Similar to DPMs, Brox et al. [17] used a poselet-based detector that characterizes object parts rather than global objects. In addition, their detector overcomes the shift and deformation issues that have affected view-independent object detection by non-rigidly aligning each poselet activation to the corresponding edge structures in the image. Khan et al. [18] used color attributes as explicit color representations during object detection. The combination of shape features with compact and computationally efficient color attributes improve the object detection results significantly. Russakovsky et al. [19] proposed an object-centric spatial pooling approach (OCP) for determining the location of an object of interest, which can be useful for image classification. OCP is used to infer the locations of objects, and this location information is then used to pool the foreground and background features. However, these methods need to extend their models to deal with the continuous viewpoint estimation problem.

Recent studies have attempted to detect view-independent objects using 3D object models. Glasner et al. [20] incorporated a category-level detection and viewpoint estimation method for rigid 3D objects from single 2D images. This research uses the voting method for efficient accumulation of evidence, and combines a re-scoring and refinement mechanism using an ensemble of view-specific SVMs. However, this method considerably has only been used for experiments on rigid car data. Zia et al. [21] proposed a method for representing object models with more geometric detail than that provided by previous object class detectors, such as local shape features, discriminative part detectors, and efficient techniques of approximate probabilistic inference. The authors then proved that this geometric richness is a meaningful ingredient for accurate geometric scene-level reasoning. Pepik et al. [22] extended discriminatively trained deformable part models to include estimates of both the