Available online at www.sciencedirect.com
SCIENCE DIRECT
Image and Vision Computing 24 (2006) 327–341
www.elsevier.com/locate/imavis

Object detection using spatial histogram features

Hongming Zhang [a,*], Wen Gao [a,b], Xilin Chen [b], Debin Zhao [a]

[a] *Department of Computer Science and technology, Harbin Institute of Technology, No. 92, west Da-zhi street, Harbin 150001, China*
[b] *Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China*

Received 5 April 2005; received in revised form 22 November 2005; accepted 23 November 2005

Abstract

In this paper, we propose an object detection approach using spatial histogram features. As spatial histograms consist of marginal distributions of an image over local patches, they can preserve texture and shape information of an object simultaneously. We employ Fisher criterion and mutual information to measure discriminability and features correlation of spatial histogram features. We further train a hierarchical classifier by combining cascade histogram matching and support vector machine. The cascade histogram matching is trained via automatically selected discriminative features. A forward sequential selection method is presented to construct uncorrelated and discriminative feature sets for support vector machine classification. We evaluate the proposed approach on two different kinds of objects: car and video text. Experimental results show that the proposed approach is efficient and robust in object detection.

Keywords: Object detection; Spatial histogram features; Feature selection; Histogram matching; Support vector machine

1. Introduction

In computer vision community, object detection has been a very challenging research topic. Given an object class of interest *T* (target, such as pedestrian, human face, buildings, car or text) and an image *P*, object detection is the process to determine whether there are instances of *T* in *P*, and if so, return locations where instances of *T* are found in the image *P*. The main difficulty of object detection arises from high variability in appearance among objects of the same class. An automatic object detection system must be able to determine the presence or absence of objects with different sizes and viewpoints under various lighting conditions and complex background clutters.

Many approaches have been proposed for object detection in images under cluttered backgrounds. In most approaches, the object detection problem is solved within a statistical learning framework. First, image samples are represented by a set of features, and then learning methods are used to identify objects of interest class. In general, these approaches can be classified as two categories: global appearance-based approaches and component-based approaches.

Global appearance-based approaches consider an object as a single unit and perform classification on the features extracted from the entire object. Many statistical learning mechanisms are explored to characterize and identify object patterns. Rowley et al. [1] and Carcia and Delakis [2] use neural network approaches as classification methods in face detection. Based on wavelet features, Osuna et al. [3] and Papageprgiou and Poggio [4] adopt support vector machines to locate human faces and cars. Schneiderman and Kanade [5] use Naïve Bayes rule for face and non-face classification. Recently, boosting algorithms are applied to detect frontal faces by Viola and Jones [6], then are extended for multi-view face detection by Li et al. [7] and for text detection by Chen and Yuille [8]. Other learning methods used in object detection include probabilistic distribution [9,10], principal components analysis [11] and mixture linear subspaces [12].

Component-based methods treat an object as a collection of parts. These methods first extract some object components, and then detect objects by using geometric information. Mohan et al. [13] propose an object detection approach by components. In their approach, a person is represented by components such as head, arms, and legs, and then support vector machine classifiers are used to detect these components and decide whether a person is present. Naquest and Ullman [14] use fragments as features and perform object recognition with informative features and linear classification. Agarwal

* Corresponding author. Tel.: +86 10 58858300(313); fax: +86 10 58858301.
E-mail addresses: hmzhang@jdl.ac.cn (H. Zhang), wgao@jdl.ac.cn (W. Gao), xlchen@jdl.ac.cn (X. Chen), dbzhao@jdl.ac.cn (D. Zhao).

0262-8856/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2005.11.010

et al. [15] extract a part vocabulary of side-view cars using an interest operator and learn a Sparse Network of Winnows classifier to detect side-view cars. Fergus et al. [16] and Leibe et al. [17,18] also use interest operators to extract objects' parts and perform detection by probabilistic representation and recognition on many object classes, such as motorbikes, human faces, airplanes, and cars.

As opposed to a majority of the above approaches, the problem of detecting multi-class objects and multi-view objects has been recently gained great attention in computer vision community. Schneiderman and Kanade [5] train multiple view-based detectors for profile face detection and car detection. Lin and Liu [19] propose a multi-class boosting approach to directly detect faces of many scenarios, such as multi-view faces, faces under various lighting conditions, and faces with partial occlusions. Amit et al. [20] use a coarse to fine strategy for multi-class shape detection with an application of reading license plates. There are 37 object classes to be recognized, including 26 letters, 10 digits, and 1 special symbol. Li et al. [21,22] propose methods to learn a geometric model of a new object category using a few examples and detect multi-class objects by a Bayesian approach. To improve efficiency, Torralba et al. [23] introduce an algorithm for sharing features across object classes for multi-class object detection. Tu et al. [24] propose an image parsing framework to combine image segmentation, object detection, and recognition for scene understanding.

One visual task related to object detection is object recognition, whose goal is to identify specific object instances in images. Local descriptor-based methods are increasingly used for object recognition. Schiele [25] proposes to use Gaussian derivatives as local characteristics to create a multi-dimensional histogram as object representation, and then perform the task to recognize many 3D objects. Lowe [26] develops an object recognition system that uses SIFT descriptors based on local orientation histograms. However, these methods are designed to recognize a specific object rather than in generalization to categorize the object class.

Feature extraction for object representation plays an important role in automatic object detection systems. Previous methods have used many representations for object feature extraction, such as raw pixel intensities [1,2,27], edges [28], wavelets [3,4,29], rectangle features [6–8], and local binary pattern [30]. However, what kinds of features are stable and flexible for object detection still remains an open problem.

Motivated by the observation that objects have texture distribution and shape configuration, we propose spatial histogram based features (termed as spatial histogram features) to represent objects. As spatial histograms consist of marginal distributions of an image over local patches, the information about texture and shape of the object can be encoded simultaneously. In contrast to most features previously used, spatial histogram features are specific to the object class, since discriminative information of the object class is embedded into these features through measuring image similarity between the object class and the non-object class. In addition, computation cost of spatial histogram features is low. Our previous work [31] shows that spatial histogram features are effective and efficient to detect human faces in color images.

Based on object representation of spatial histogram features, we present an object detection approach using a coarse to fine strategy in this paper. Our approach uses a hierarchical object detector combining cascade histogram matching and a support vector machine to detect objects, and learns informative features for the classifier. First, we employ Fisher criterion to measure the discriminability of each spatial histogram feature, and calculate features correlation using mutual information. Then, a training method for cascade histogram matching via automatically selecting discriminative features is proposed. Finally, we present a forward sequential selection algorithm to obtain uncorrelated and discriminative features for support vector machine.

Unlike methods which use interest operators to detect parts prior to recognition of the object class, we apply the proposed object detector at anywhere in image scale space. Therefore, our method does not need figure-ground segmentation or object parts localization. In contrast to most systems which are designed to detect a single object class, our method can be applied to any type of object classes with widely varying texture patterns and varying spatial configurations. Extensive experiments on two different kinds of objects (car and video text) are conducted to evaluate the proposed object detection approach.

The rest of the paper is organized as follows. Section 2 outlines the proposed object detection approach. Section 3 describes spatial histogram features for object representation and provides quantitative measurement of spatial histogram features. Section 4 presents the methods of selecting informative features for object detection. Section 5 gives experiment results of car detection and video text detection. Section 6 concludes this paper.

2. Overview of the proposed object detection approach

The proposed approach is designed to detect multiple object instances of different sizes at different locations in an input image. Take car detection as an example, the overall architecture of the object detection approach is illustrated in Fig. 1. One essential component of the proposed approach is an object detector, which uses spatial histogram features as object representation. We call the object detector as spatial histogram features-based object detector (hereinafter referred as 'SHF-based object detector'). In our approach, the SHF-based object detector is formed as a hierarchical classifier which combines cascade histogram matching and support vector machine.

For object detection process, we adopt an exhaustive window search strategy to find multiple object instances in an input image. The process of object detection contains three phases: image pyramid construction (Step 1), object detection at different scales (Step 2), and detection results fusion (Step 3).

Initially, an image pyramid is constructed from the original image in the Step 1. The detector is applied at every location in