Optik 126 (2015) 2912–2916

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Generic visual categorization using composite Gabor and moment features

S. Arivazhagan, R. Ahila Priyadharshini *

Mepco Schlenk Engineering College, Sivakasi, Tamil Nadu, India

ARTICLE INFO

Article history:
Received 26 May 2014
Accepted 4 July 2015

Keywords:
Object recognition
Salient point detector
Patch
Gabor features
Moment features

ABSTRACT

In this paper an effective method to recognize objects from different categories of images which suffer from illumination, variability in shape, occlusion and clutter, based on a combination of spatial and spectral features called new composite features is presented. In domain of object recognition, it is often to classify objects from images that make only limited part of the image. Hence to identify local features and certain region of images, patches are extracted over the interest points detected from the original image using Wavelet based interest point detector. Gabor features and Moment features are computed separately for every patch and classified using SVM classifier. In addition to this, Gabor features are combined with Moment features, so-called new composite features are computed for every patch and its performance is compared with the independent features. The observations revealed that composite features outperform the independent features with less error rate. The experimental evaluation is done using the Caltech database.

1. Introduction

Object recognition is a great task in computer vision. It is the concept of finding objects in an image or video sequence. Normally, Human Visual systems recognize large number of objects from images with little effort even when they have different appearances in different circumstances. But Computer Systems face it with a lot of difficulties as they have to recognize objects that is occluded by other objects and that having different appearances [1]. The performance of computer vision approaches for object recognition is still not comparable with the human biological analogies in these respects: real-time response, level of performance, and the ability to handle thousands of objects.

In order to improve the computer vision, the image data should be represented in a suitable way. Therefore, objects can be described into different types. These include model-based approaches [2–4], shape-based approaches [5,6] and appearance-based approaches [7–10]. Model-based approaches try to represent the object as a collection of three dimensional, geometrical primitives (boxes, spheres, cones, cylinders, generalized cylinders, surface of revolution) whereas shape-based methods represent an object by its shape/contour. In contrast, for appearance-based models only the appearance is used, which is usually captured by different two-dimensional views of the object-of-interest.

The main disadvantages of geometry-based methods (model or shape based methods) are: the dependency on reliable extraction of geometric primitives, the ambiguity in interpretation of the detected primitives (presence of primitives that are not modeled), the restricted modeling capabilities only to a class of objects which are composed of few easily detectable elements, and the need to create the models manually [11]. Appearance-based methods demanded exhaustive set of learning images, taken from densely distributed views and illuminations. Such set is only available when the object can be observed in a controlled environment.

In order to address the above mentioned issues, methods based on matching local features (recognition based on Correspondence of local features) is proposed. Here, objects are represented by a set of local features (patches), which are automatically computed from the training images. The learned features are organized into a database. When recognizing a test image, local features are extracted as in the training images. Similar features are then retrieved from the database and the presence of objects is assessed in the terms of the number of local correspondences. The approach is robust to occlusion, variability in shape and cluttered background. The block diagram of the proposed method is shown in Fig. 1. Our approach addresses the background clutter problem by extracting patches over the interest points which are detected by wavelet based approach. The numbers of salient points, which

* Corresponding author. Tel.: +91 9486647103.
E-mail addresses: s_arivu@yahoo.com (S. Arivazhagan), ahilaprem@gmail.com (R. Ahila Priyadharshini).

http://dx.doi.org/10.1016/j.ijleo.2015.07.035

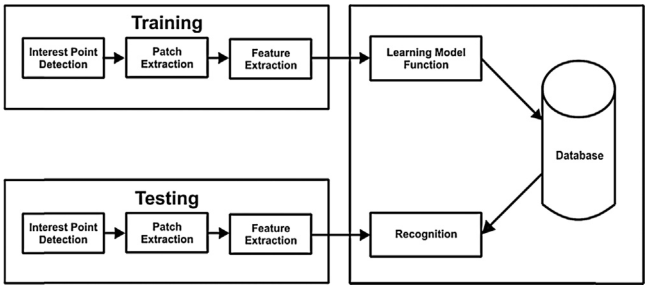

Fig. 1. Block diagram of the proposed method.

Fig. 2. (a) Sample images from Caltech database, (b) most prominent salient points detected.

discriminate the object from the cluttered background, are selected by applying highest priority to their high saliency value. The new set of features known as composite Gabor and moment features are computed for every patch and classified.

2. Interest point detection

Interest points (or) salient points are the points which maximize the discrimination between the objects. Interest point detection plays an important role in content based image retrieval in order to represent the local properties of the image. The salient points are not confined to corners, but show variations that happen at different resolutions in the images [12]. In this paper, the interest points are detected using wavelet transform to detect global variations as well as local ones. The aim is to find a relevant point to represent this global variation by looking at wavelet coefficients at finer resolutions. The algorithm for detecting relevant salient points using Haar wavelet transform is given as follows:

- Calculate the wavelet representation of an image for all scales $j = 1/2, \ldots, 2^{-J_{\max}}$ and spatial orientations $d = 1, 2, 3$, where $J_{\max} = \log_2[\min(m, n)]$, m and n are the width and height of an image.
- For each wavelet coefficient, find the maximum child coefficient.
- Track it recursively in finer resolutions.
- At the finer resolution (½), set the saliency value of the tracked pixel: the sum of the wavelet coefficients tracked.
- Choose the most prominent points based on the saliency value.

The tracked point and its saliency value are computed for every wavelet coefficient. A point related to a global variation has a high saliency value, since the coarse wavelet coefficients contribute to it [13]. A finer variation also leads to an extracted point, but with a lower saliency value. Then it is needed to threshold the saliency value, in relation to the desired number of salient points. Even though the number of salient points obtained varies from image to image, the number of interest points considered for the experiment process here is 250. The most prominent 250 salient points are detected from each and every categories of Caltech database. Fig. 2 shows the most prominent salient points detected from sample images of Caltech database.

3. Feature extraction

The transformation of input data into set of features is called feature extraction. The reduced information that is set of features instead of full size input is used to recognize various complex images with better accuracy. The two kinds of features used here are Gabor and Moment features.

3.1. Review on Gabor features

In the spatial domain 2D Gabor wavelet is a Gaussian kernel modulated by a sinusoidal plane wave [14]. Gabor wavelets provide analysis of the input signal in both spatial and frequency domains simultaneously. Gabor functions provide the optimal resolution in both the spatial and frequency domains, and the Gabor wavelet transform seems to be the optimal basis to extract local features for several reasons such as biological motivation and empiric motivation [15]. Here, Gabor wavelet decomposition at 2 scales and 2 orientations (s = 2; o = 2) and 2 scales and 4 orientations (s = 2; o = 4) is done for every patch extracted from the original image. The 2D Gabor wavelet function is defined in Eq. (1)

$$\varphi(x, y, \omega_0, \theta) = \frac{1}{2\pi\sigma_x\sigma_y}\left[e^{-\left(((x\cos\theta + y\sin\theta)^2/2\sigma^2) + ((-x\sin\theta + y\cos\theta)^2/\sigma_x\sigma_y)\right)}\right] \times \left[e^{i(\omega_0 x\cos\theta + \omega_0 y\sin\theta) - e^{-\omega_0^2\sigma_x\sigma_y/2}}\right] \quad (1)$$

where ω_0 – radial center frequency; x, y – Pixel position in the spatial domain; σ_x, σ_y – standard deviation of Gaussian function along the x and y axis; θ – orientation of the Gabor wavelet.

This Gabor wavelet decomposition results in 4 and 8 filtered images at s = 2; o = 2 and s = 2; o = 4 respectively. Mean and standard deviation, a statistical measure, on these filtered images results in the feature vectors for every patch.

3.2. Review on moment features

It is very important to recognize objects despite their position, size, illumination and orientation in the field of object recognition. Here in addition with Gabor features obtained from Gabor wavelet, Basic Moments are also used as features. Basic moments are invariant under scale, translation and rotation. Each and every category of moments has its own advantages. Among various kinds of moments, one simple method of moments used here is geometric moments. Geometric moments are meant to be a very efficient tool in image analysis [16]. The formula for calculating basic moments is shown in Eq. (2)

$$M = \sum_x \sum_y I(x, y)^\alpha x^p y^q \quad (2)$$

Here $I(x, y)$ represents extracted patch values, α indicates degree of moments, while p and q are order of moments. A total of ten Basic moments using orders 0, 1 and 2 and degrees 1 and 2 are extracted for every patch.