Pattern Recognition 47 (2014) 2116–2125

Contents lists available at ScienceDirect

Pattern Recognition

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

Bag of contour fragments for robust shape classification

Xinggang Wang [a], Bin Feng [a,*], Xiang Bai [a], Wenyu Liu [a], Longin Jan Latecki [b]

[a] Department of Electronics and Information Engineering, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, Hubei Province 430074, PR China
[b] CIS Department, Temple University, 1805 N. Broad St., Philadelphia, PA 19122, USA

ARTICLE INFO

Article history:
Received 22 March 2013
Received in revised form
11 December 2013
Accepted 22 December 2013
Available online 3 January 2014

Keywords:
Shape classification
Shape representation
Bag of contour fragments

ABSTRACT

Shape representation is a fundamental problem in computer vision. Current approaches to shape representation mainly focus on designing low-level shape descriptors which are robust to rotation, scaling and deformation of shapes. In this paper, we focus on mid-level modeling of shape representation. We develop a new shape representation called Bag of Contour Fragments (BCF) inspired by classical Bag of Words (BoW) model. In BCF, a shape is decomposed into contour fragments each of which is then individually described using a shape descriptor, e.g., the Shape Context descriptor, and encoded into a shape code. Finally, a compact shape representation is built by pooling shape codes in the shape. Shape classification with BCF only requires an efficient linear SVM classifier. In our experiments, we fully study the characteristics of BCF, show that BCF achieves the state-of-the-art performance on several well-known shape benchmarks, and can be applied to real image classification problem.

1. Introduction

Shape is an intrinsic feature for image understanding, which is stable to illumination and variations in object color and texture. Because of these advantages, shape is widely considered for object recognition. In particular, with the recent advance in contour detection proposed by Arbelaez et al. in [1], shape based object recognition in natural image is becoming more practical and attracts more attention in computer vision community. Main challenges in shape based object recognition include deformation, occlusion and viewpoint variation of objects. Various shape descriptors have been proposed to address these challenges, *e.g.*, [2–5]. Shape based object recognition is usually considered as a classification problem. Given a set of training shapes and category label of each training shape, we need to determine which category a testing shape belongs to. Traditional shape classification methods are usually based on matching shape descriptors from two different shapes: for every training shape, we find correspondences between its shape descriptors and the shape descriptors in the testing shape using matching algorithms, such as Hungarian algorithm, dynamic programming algorithm; then we compute matching costs according to the matching results; finally, we rank training shapes based on the matching costs and classify the testing shape using the nearest neighbor (NN) classifier. This exemplar-based shape classification strategy has been widely used, for example, in [2,3,6]. However, it has its own limitations. With few training samples, it is difficult to capture the large intra-class variation using these algorithms. For large training samples, it is extremely time consuming to perform shape matching one-by-one.

Different from exemplar-based shape matching, in this paper, we propose a compact shape representation and handle the large intra-class variation by discriminative learning. Inspired by the huge progress in image classification and representation with Bag-of-Words (BoW) [7,8], we decompose shape into contour fragments and quantize the contour fragments into shape codes. The contour fragments under different scales contain both local and global shape information which can be encoded utilizing coding strategies for local descriptors [9,10]. Then, a statistical histogram of shape codes is used to represent each shape and similarity of shapes can be directly computed from these histograms. Matching shapes based on this new shape representation does not explicitly give correspondences between contour fragments. But using a classifier for shape classification is much more efficient than using the typical matching algorithms such as Hungarian, thin plate spline (TPS), dynamic programming, dynamic time warping, and so on. In fact, BoW model is a natural solution for finding correspondences between two sets of features and can be used efficiently for recognition tasks. However, it has seldom been successfully applied to shape analysis, since the popular image descriptors such as SIFT [11] and LBP [12] are mainly designed for describing the local texture/appearance variations. These image features are not good at capturing the intrinsic structure in shape. Toward this end, we directly work on shape contour by decomposing it into contour fragments. We name our method Bag of Contour Fragments (BCF), which can not only provide a compact and informative representation, but also achieve the state-of-the-art classification performance on several popular shape benchmarks.

* Corresponding author.
E-mail addresses: xgwang@hust.edu.cn (X. Wang), fengbin@hust.edu.cn (B. Feng), xbai@hust.edu.cn (X. Bai), liuwy@hust.edu.cn (W. Liu), latecki@temple.edu (L. Jan Latecki).

0031-3203/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.12.008

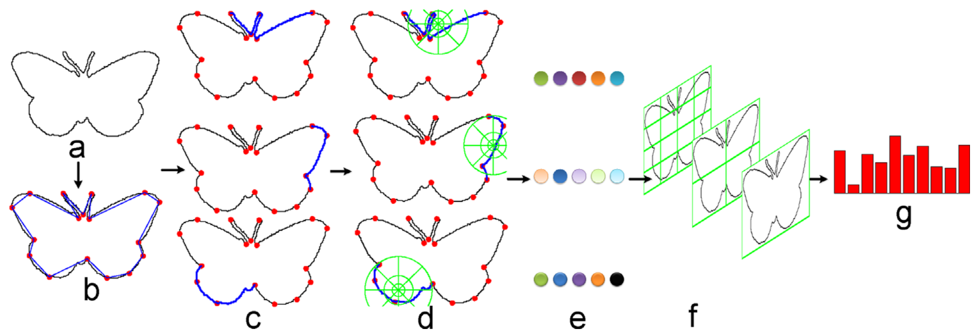

Fig. 1. Pipeline of building shape representation using BCF. (a) Shows contour of a shape; (b) shows critical points detected using DCE method; (c) shows some contour fragments in blue color; (d) shows that we use shape context [2] to describe each contour fragment; (e) shows shape codes; (f) shows we use 1×1, 2×2, and 4×4 spatial pyramid for max-pooling; (g) shows the histogram for shape representation. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

Pipeline of building shape representation in BCF is shown in Fig. 1. The outer contour of each shape is decomposed into salient contour fragments using a well-known contour decomposition method named discrete contour evolution (DCE) [13]. Each contour fragment is then described by collecting the shape context features [2] on its reference points, and encoded into shape codes. Finally, the shape codes are pooled into a compact image representation with spatial pyramid. We utilize the current advances in image classification, such as local-constrained linear coding (LLC) [9] for feature coding and spatial pyramid matching (SPM) [14] in our BCF shape classification framework. Both LLC and SPM are seldom used in shape analysis. LLC utilizes the locality constraints and encodes each descriptor with its local-coordinate system in a codebook. In practice, it first performs k-nearest-neighbor search to find local-coordinates for feature to be encoded, and then solves a constrained least square fitting problem on the local-coordinates. The state-of-the-art performance on PASCAL VOC [15] image classification has shown effectiveness of LLC. SPM is a simple and computationally efficient extension of the orderless BoW model for image representation. It works by partitioning the image into increasingly fine sub-regions and computing histograms of local features found inside each sub-region. Histograms of different sub-regions are concatenated as final image representation. SPM can capture the spatial information in contour fragments which are useful for shape recognition. BCF naturally utilizes LLC and SPM to improve the accuracy of shape classification.

One of the major difficulties involved in shape classification for many shape-matching based algorithms is to directly match two shapes with large deformation since shapes are only partially similar to each other. BCF can easily solve this problem caused by large shape deformation, and is good at classifying shapes with partial similarity. As each shape contour is divided into contour fragments in BCF, the contour fragments contain partial shape information. After coding, a discriminative classifier such as SVM or Adaboost can be used to select the representative and informative contour parts for each shape category. Fig. 4 shows some contour fragments selected by linear SVM in four shape categories in our experiments. We can see that even though contour fragments are parts of the shapes, they are very informative for recognizing shape category. Thus, BCF is able to deal with partial occlusion in shape, especially, in the edge map extracted from real image. Besides, we find that BCF is also robust to noisy contour in our experiments.

In summary, the proposed BCF has several good properties:

1. It provides a very compact shape representation which is a single vector rather than a set of feature vectors used in many other methods.
2. It precisely preserves information of individual shape contour via LLC and spatial layout of contour fragments in one shape via SPM.
3. For shape classification it avoids pairwise matching between local shape descriptors and significantly reduces the time cost.
4. It is robust to the shapes with occlusions or parts missing, and can be easily applied to real image classification.

The rest of the paper is organized as follows. We review the related works in Section 2. Then, we introduce the details of our shape representation with BCF in Section 3, including extracting, encoding and pooling contour fragments, and so on. We evaluate the proposed method on several popular shape benchmarks, illustrate good properties of BCF in applications, and demonstrate its effectiveness in shape classification in various datasets in Section 4. Finally, we conclude this paper in Section 5.

2. Related work

Here, we briefly review the recent progress in shape classification. Sun and Super [16] proposed a shape classification framework for recognizing contour shapes using class contour segments as input features with Bayesian classifier. Bai et al. [6] adopted contour segments and skeleton paths as the input features for shape classification with a Gaussian mixture model. Daliri and Torre [17,18] transformed the contour points into a symbol representation, and then used the edit distance between pair of strings is used for classification with a kernel support vector machine. Wang et al. [19] proposed a tree-union [20] representation as the prototype for each shape category, and performed shape classification is determined by the shape similarity between a test shape and each prototype. Edem and Tari [21] also used a skeletal tree model to represent the prototype of each category, and then used the edit distance between a given shape and each prototype is used as the input feature for a linear SVM. Thus, each prototype in [21] can be considered as a shape codebook. Shape classification by skeleton matching has been studied by [22–25].

Various shape descriptors have been proposed for shape matching and recognition. There are some region-based methods, such as Zernike moments [26] and generic Fourier descriptor [27]. Other methods based on contour include curvature scale space (CSS) [4], multi-scale convexity concavity (MCC) [28], triangle area representation (TAR) [5], hierarchical procrustes matching (HPM) [29], shape-tree [30], contour flexibility [31], shape context (SC) [2], inner-distance shape context (IDSC) [3] and so on. In this paper, we only use shape context to describe contour fragments in BCF. Generally speaking, most of these shape descriptors can be