Neurocomputing 190 (2016) 70–81

Contents lists available at ScienceDirect

Neurocomputing

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

Metric learning based object recognition and retrieval

Jianyu Yang *, Haoran Xu

School of Urban Rail Transportation, Soochow University, Suzhou, China, 8 Jixue Road, Xiangcheng District, Suzhou, Jiangsu, China

ARTICLE INFO

Article history:
Received 5 November 2015
Received in revised form
22 December 2015
Accepted 5 January 2016
Communicated by Jianbin Qiu
Available online 18 February 2016

Keywords:
Metric learning
Object recognition
Object retrieval
Robot learning
Intelligent analysis

ABSTRACT

Object recognition and retrieval is an important topic in intelligent robotics and pattern recognition, where an effective recognition engine plays an important role. To achieve a good performance, we propose a metric learning based object recognition algorithm. To represent the invariant object features, including local shape details and global body parts, a novel multi-scale invariant descriptor is proposed. Different types of invariant features are represented in multiple scales, which makes the following metric learning algorithm effective. To reduce the effect of noise and improve the computing efficiency, an adaptive discrete contour evolution method is also proposed to extract the salient feature points of object. The recognition algorithm is explored based on metric learning method and the object features are summarized as histograms inspired from the Bag of Words (BoW). The metric learning methods are employed to learn object features according to their scales. The proposed method is invariant to rotation, scale variation, intra-class variation, articulated deformation and partial occlusion. The recognition process is fast and robust for noise. This method is evaluated on multiple benchmark datasets and the comparable experimental results indicate the effectiveness of our method.

1. Introduction

Object recognition is a fundamental and challenging problem in robot learning and pattern recognition with applications such as object retrieval [1,2], robot navigation [3] and multiple camera object recognition based on network control technology [4,5]. The visual feature of object is promising to contain the object information and an intuitive idea is to measure the similarity of two objects by comparing their features. In recent years, there is a fruitful literature in object recognition and retrieval algorithms that reports promising results [6–9]. Most of the researchers in the current object feature analysis community make effort to design a robust and discriminative descriptor which plays an important role in the traditional pairwise matching based methods [2,10–12]. However, it is still an open problem to recognize objects effectively in large scale datasets, due to the unavoidable perturbations, e.g., intra-class variation, articulated deformation, occlusion and noise.

Most of the existing descriptors are proposed to improve the performance of shape matching in two aspects. On one hand, the descriptors are expected to capture the geometric invariance of salient shape features against various variations, e.g, rotation and articulated deformation. The Integral Invariant (II) [13] is a typical method in this aspect. On the other hand, the shape structure information is preferred to represent the shape, which is usually represented by the spatial relations between each pair of the contour points, e.g., the shape context methods [2,10]. However, both of the two types of methods have limits. The learning algorithms are widely used in recent works to solve this problem, but most of the learning algorithms are directly used without considering the property of descriptors, which limits the recognition results. As the object descriptors represent various aspects of salient features, the matching between learning method and feature description is very important for the performance of object recognition. Thus the learning method should be fine selected and designed.

In this paper, we propose an object recognition method based on metric learning algorithm. For feature description, a novel multi-scale invariant descriptor (MID) is proposed with three types of invariants in different scales. The multiple scales of features are well matched with the metric learning method. Besides, we adopt the bag of words (BoW) method to summarize the feature vectors in different scales into histograms which makes the features much suit for the metric learning method. This increases the discrimination of our method. To reduce the intra-class distance, the core idea is using the directly connected region to calculate the MID instead of the II for articulated deformations. To increase the representative of feature and decrease the effect of noise, an evolution method is also proposed to find the salient feature points of object.

* Corresponding author.
E-mail addresses: jyyang@suda.edu.cn (J. Yang), hranxu@163.com (H. Xu).

http://dx.doi.org/10.1016/j.neucom.2016.01.032
0925-2312/

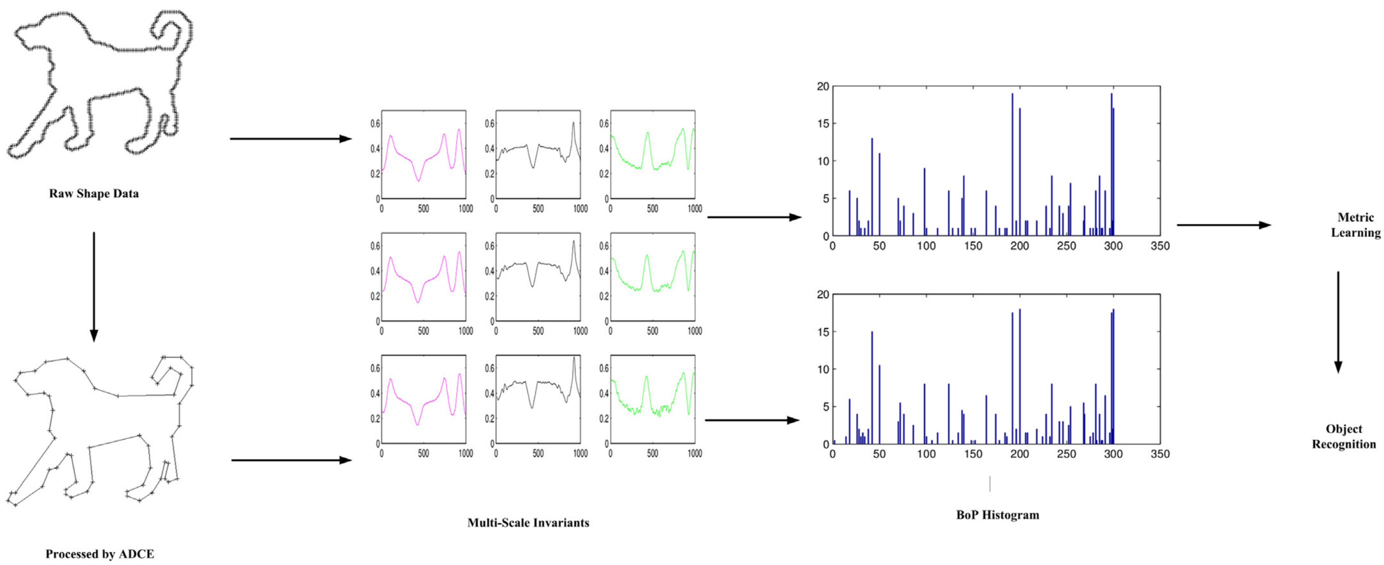

Fig. 1. The pipeline of our method.

The pipeline of our method is shown in Fig. 1. In our framework, the raw data is preprocessed by a new contour evolution method first. In general shape data, there are many redundant contour points which have no contribution to the object representation but introduce extra perturbation to object matching. To extract the salient feature points, we propose an adaptive discrete contour evolution (ADCE) algorithm to process the closed shape contour. After that, the invariants of different types and scales are calculated to represent the shape features. To make the features appropriate for metric learning, a bag of salient points (BoP) method is proposed to represent the shape feature points by histograms. At last, the metric learning algorithm is used according to relevant component analysis.

The proposed method has the following advantages:

- Our method is invariant to salient intra-class variation, partial occlusion and articulated deformation.
- The metric learning based method enlarges the inter-class distance and reduce the intra-class distance.
- Object recognition by computing the metric learned BoP histograms is very fast. The traditional matching algorithms, e.g., DTW, DP, TPS, are not used in our method.
- Our method is robust to noise.

The proposed method is evaluated on multiple benchmark shape datasets, including MPEG-7 dataset [14], Kimia's 99 dataset [15], Kimia's 216 dataset [15] and Articulated dataset [10]. The experiments are carried out on these datasets to evaluate the invariant properties of our method, including: rotation, scale variation, intra-class variation, articulated deformation and occlusion. We also test the capability of our method for shape retrieval comparing with other methods. The robustness of our method to noise is also tested.

The organization of the remainder of this paper is as follows. The next section summarizes the relevant work on object recognition. In Section 3, the multi-scale invariants are presented. The adaptive discrete contour evolution method is presented in Section 4. Section 5 introduces the BoP method and the RCA metric learning for object recognition, and the experiments follow in Section 6. Finally, this work is concluded in Section 7.

2. Related work

In the literature, there are various methods for object recognition. In this section, we mainly discuss two sorts of them: the learning based methods and the matching based methods, which are the most related to the proposed method in this paper. The invariant descriptors [13,16–18] are predicted to be invariant to spatial variations and various deformations. The integral invariants [13,19] use the distance integral invariant and area integral invariant to represent each point of the shape contour. The distance integral invariant is not discriminative for shape matching, while the area integral invariant is a local descriptor (approximation of curvature) which cannot represent the spatial structure and context of shapes. The curvature scale space (CSS) [16] method uses the locations of the curvature zero crossing points on shape contour to represent shapes. But the scale is hard to determine and the convex shapes have no curvature zero crossing points. The Biswas method [17] uses a variety of simple invariants to represent shape features for shape indexing and retrieval.

The structure learning methods make use of the spatial relation among contour points [2,9,11,20,21]. The most referred work is the Shape Context (SC) [2] which represent each contour point by the spatial relations with all other points as a histogram. This method can well characterize the spatial distribution of contour points and performs well on rigid objects, whereas it is sensitive to articulated deformation. The triangle area based methods [11,20] use the spatial relationship of three contour points to represent the shape feature. Alajlan et al. [11] propose the triangle-area representation (TAR) method for shape description where the length of triangle base is changing. But the feature extraction is complicated and sensitive to noise. El et al. [20] propose the multi-scale triangle-area representation (MTAR) method to make use of the zero crossings of the triangle-area at different wavelet scales for matching, which is more robust to noise. Wang et al. [9] use the height function to represent the spatial relationship among contour points. Each contour point is represented by the distances from all other points to the tangent line of it. The fuzzy based [22,23] modeling and learning methods have been widely applied to computer vision system which can also be used in this application.

In most of the recent works, it is hard to distinguish them to any sort because both of the targets are concerned together: represent