J. Vis. Commun. Image R. 25 (2014) 1640–1646

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Multi-scale embedded descriptor for shape classification

Chen Huang, Tony X. Han, Zhihai He *

Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211, USA

ARTICLE INFO

Article history:
Received 12 March 2014
Accepted 11 August 2014
Available online 28 August 2014

Keywords:
Shape descriptor
Feature extraction
Running angle
2-D description image
Subband decomposition
Wavelet transformation
Object recognition
Image classification

ABSTRACT

We present a new shape descriptor that are robust to deformation and capture part details. In our framework, the shape descriptor is generated by (1) using running angle to transforming a shape into a 2-D description image in the position and scale space and (2) performing circular wavelet-like sub-band decomposition of this 2-D description image based on its periodic convolution with orthogonal kernel functions. Each sub-band is described by the histogram of its decomposition coefficients. To capture unique and discriminative part, we compare the decomposition coefficients across sub-band to detect singularity in the position and scale space. The singularity information is encoded with a tree of binary bits. The coded feature vectors of all sub-bands and singularity trees are pooled together to form the descriptor of the shape. The shapes are classified with linear SVM. Our performance evaluations on several public datasets, demonstrating that the proposed method significantly outperforms state-of-the-art methods.

1. Introduction

Shape description, matching and classification are fundamental problems in computer vision with important applications like image retrieval, object detection and recognition. Compared to other image features, shape is invariant to lighting conditions and changes in object colors and texture [1]. However, shape contours obtained from object segmentation often exhibits a large degree of intra-class of variations due to different view points, changing illumination, segmentation errors, shape parts articulation variations, non-rigid deformations, etc. [2]. Furthermore, shape contours of objects often have strong inter-class ambiguities. For example, tree leaves or animals of similar species, are often very similar to each other, except some small or even tiny distinguishable features embedded in a large amount of intra-class shape contour variations.

For example, Fig. 1 shows two types (Group-1 and Group-3) of leaves from the Swedish Leaf Database [3]. We can see that the inter-class ambiguity is very strong. Human eyes, even experts, cannot tell them apart.

In the literature, a number of efficient shape descriptors and shape similarity measures have been developed for representing, matching, classifying, and recognizing shapes. One typical approach for shape similarity measure is to construct some physical shape models and then measure the amount of energy required to deform one shape contour into another [4]. Latecki et al. [5] developed an elastic partial shape matching algorithm to model a possible non-rigid shape deformation. A hierarchical matching approach has been developed in [6]. It has been observed that this type of approaches are often very sensitive to strong, local shape variations. Extracting invariant descriptors of the shape is another important approach. In order to capture the local features, Belongie et al. [7] introduce the 2-dimensional non-linear histogram, Shape Context (SC), to describe the distance and angles between contour points. Since SC cannot solve the problem of matching articulated shape, Ling and Jacobs [8] modified SC by using the geodesic distance inner shape instead of Euclidean distance to represent shapes, which is called Inner Distance Shape Context (IDSC). Bending invariants [9] for 2D and 3D shapes can be achieved by using geodesic distances. For both 2D and 3D shape analysis, topology invariants have been developed in [10]. To capture the inherent part structure of the shape contour, skeleton based approaches, particularly the shock graph method in [2], have been developed. Given a shape and its boundary, shocks are defined as curve segments of the medial axis with monotonic flow. The shocks are then organized into a shock graph, which forms a hierarchical representation of the shape and naturally captures its part structure. These invariant descriptors have demonstrated their effectiveness in handling intra-class variations. However, as observed in [11], invariants to larger groups of deformation often come at a price of reduced inter-class discriminative.

* Corresponding author.
E-mail addresses: chenhuang@mail.missouri.edu (C. Huang), hantx@missouri.edu (T.X. Han), HeZhi@missouri.edu (Z. He).

http://dx.doi.org/10.1016/j.jvcir.2014.08.005
1047-3203/

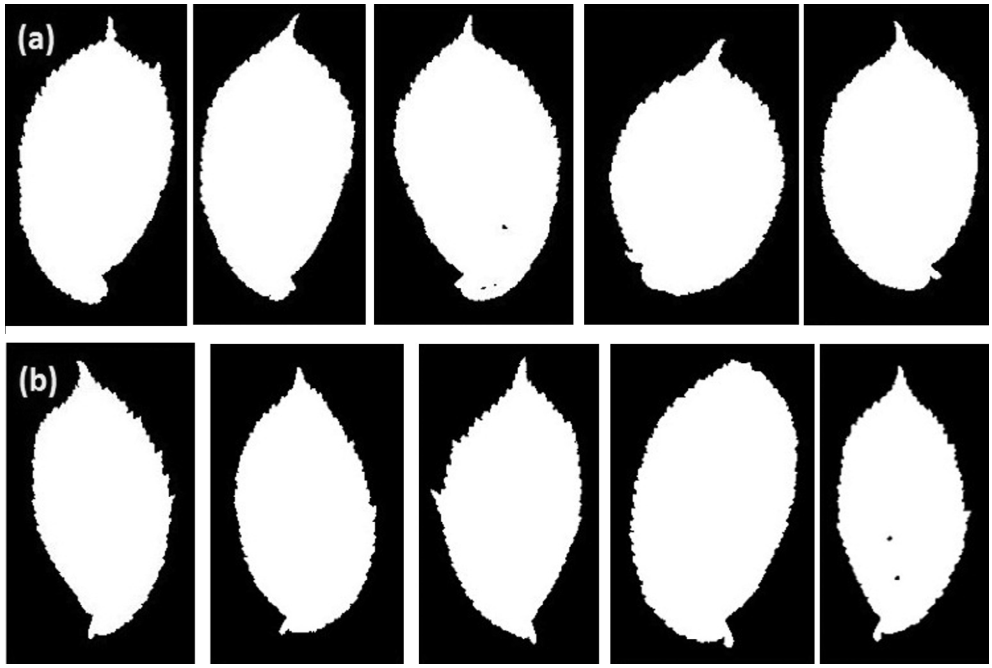

Fig. 1. Two groups of leaves, Group-1 (the first row) and Group 3 (the second row) from the Swedish Leaf Database [3].

Apart from the existing efforts in the literature to construct shape descriptors and similarity measures and then use nearest neighbor approaches for shape classification [2,6,12,8,13], in this work, we propose a global shape descriptor and a learning-based approach for shape classification, aiming to develop trainable classifiers for shape classification and recognition. During the past several years, we have witnessed the success of data-driving and machine learning approaches in computer vision applications, e.g., object detection, classification, and recognition [14,15]. These approaches are able to discover and construct important features to resolve inter-class ambiguities during classification. We believe that this learning and discovery capability is also important and beneficial in shape-based domain. However, we observe that shapes are much different from conventional image or video data for object detection and recognition. Shape contour is basically a 1-D data. An enclosed contour has no starting point. The shape feature should be invariant or insensitive under changes in scale, rotations, and non-rigid deformations. How to extract low-level from shapes for successful learning and training of accurate and robust shape classifiers remains a challenging and interesting research problem.

In this work, we propose to develop a global multi-scale embedded description scheme for shape classification. We are mainly motivated by the following observation: our human vision system often examines and compares shapes at different scales. Some classes of shapes can be well separated by shape features at coarse scales. However, for shapes of some closely related classes, we have to examine or discover detailed features at fine scales. To develop the multi-scale embedded shape description, we construct a dense description of shapes or object contours using a low-level feature called running angle, transforming the whole shape into a 2-D shape description image in the geodesic index and scale space, as Fig. 2. We then perform circular wavelet-like sub-band decomposition of this shape descriptor image based on its periodic convolution with orthogonal kernel functions. We then extract invariant features from these decomposition coefficients and pool them across sub-band to form a dense description of the shape. The shape is classified with linear SVM. Our performance evaluations on public datasets demonstrate that the proposed method significantly outperforms state-of-the-art methods.

The rest of the paper is organized as follows. Section 2 presents the 2-D shape description image. Section 3 presents our global multi-scale shape descriptor and the SVM-based shape classification framework. The experimental results are presented in Section 4. Section 5 concludes the paper.

2. Multi-scale 2-D shape description image and circular sub-band decomposition

In this work, we focus on enclosed shape contours. We uniformly sample the contour and denote this sequence of point samples by

$$\mathbf{C} = \{\mathbf{C}(s) = (x_s, y_s) | 1 \leqslant s \leqslant N\}. \tag{1}$$

s is the geodesic distance along the curve from the starting position. Later on, we will see that our algorithm does not depend on this starting position.

Next, we will compute the so-called running angle at each contour sample. As illustrated in Fig. 3, at point $\mathbf{C}(s)$, we consider a looking-out window of size w, which consists of w samples before $\mathbf{C}(s)$ and w samples after $\mathbf{C}(s)$. We compute the centroids of contour samples at both sides as follows:

$$\mathbf{P}(s,w) = \sum_{[s-w]_N}^{[s-1]_N} \mathbf{C}(s), \quad \mathbf{Q}(s,w) = \sum_{[s+1]_N}^{[s+w]_N} \mathbf{C}(s). \tag{2}$$

The running angle of the contour at index s with a looking-out window size w is defined to be the angle between vectors $\mathbf{P}(s,w) - C(s)$ and $C(s) - \mathbf{Q}(s,w)$. We denote this angle by $\theta(s,w)$. When the window size w is large, the running angle θ will capture more global information around the point $\mathbf{C}(s)$. When w is small, θ will focus on local and more detailed structures of the shape. This represents a 2-D image containing multi-scale information, with each image point taking a value between $[0, 2\pi]$. We refer to this image as the running angle image. Fig. 2 shows the running angle images for three contour samples from the MPEG-7 shape dataset. We can