Pattern Recognition Letters 33 (2012) 134–143

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

Shape matching and classification using height functions

Junwei Wang [a], Xiang Bai [a,*], Xinge You [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] Department of Computer and Information Sciences, Temple University, Philadelphia, PA, USA

ARTICLE INFO

Article history:
Received 19 April 2011
Available online 20 October 2011
Communicated by M.S. Nixon

Keywords:
Shape matching
Contour
Height function
Shape retrieval

ABSTRACT

We propose a novel shape descriptor for matching and recognizing 2D object silhouettes. The contour of each object is represented by a fixed number of sample points. For each sample point, a *height function* is defined based on the distances of the other sample points to its tangent line. One compact and robust shape descriptor is obtained by smoothing the height functions. The proposed descriptor is not only invariant to geometric transformations such as translation, rotation and scaling but also insensitive to nonlinear deformations due to noise and occlusion. In the matching stage, the Dynamic Programming (DP) algorithm is employed to find out the optimal correspondence between sample points of every two shapes. The height function provides an excellent discriminative power, which is demonstrated by excellent retrieval performances on several popular shape benchmarks, including MPEG-7 data set, Kimia's data set and ETH-80 data set.

1. Introduction

Shape matching is a very critical problem in computer vision, which has been widely used in many applications such as object recognition (Belongie et al., 2002; Siddiqi et al., 1999), character recognition (Tang and You, 2003; You and Tang, 2007), shape evolution (Lewin et al., 2010), medical image and protein analysis (Wang et al., 2011), robot navigation (Wolter and Latecki, 2004), and topology analysis in sensor networks (Jiang et al., 2009), etc. It is a very difficult problem, as shape instances from the same category, which look similar to humans, are often very different when measured with geometric transformations (translation, rotation, scaling, etc.) and nonlinear deformations (noise, articulation and occlusion). Compared to geometric transformations, the nonlinear deformations are much more challenging for shape similarity measures. Therefore, one key problem of shape matching is to define a shape descriptor which is informative, discriminative, and efficient for matching process. A good shape descriptor should tolerate the geometric differences of objects from the same category, but at the same time should allow to discriminate objects from different shape classes.

As stated in a previous study (Alajlan et al., 2007), shape descriptors with only global or local information may probably fail to be robust enough in these situations. Global descriptors are robust to local deformations, but they cannot capture local details of shape boundary. Local descriptors are precise to represent local shape features, while they are too sensitive to noise. In fact, it is always challenging to distinguish between noise and local details of the shape boundary. Naturally, one solution to this problem is to define a "rich" shape descriptor, which consists of both global and local shape characteristics. By combining local and global shape features, many recent works (Ling and Jacobs, 2007; McNeill and Vijayakumar, 2006; Felzenszwalb and Schwartz, 2007; Alajlan et al., 2007; Xu et al., 2009) achieved excellent performances on the most popular benchmark: MPEG-7 data set (Latecki et al., 2000).

Some other important requirements for a promising shape descriptor include: computational efficiency, compactness, and generality of applications. It is difficult to satisfy all of these requirements. In this paper, we propose a novel shape descriptor that captures both global and local shape features similar to recent works. For each sample point, a height function is defined as a vector of distances of the other sample points to its tangent line. Then, the whole shape contour is represented as a sequence of the height functions. A further process called smoothing is performed on these height functions to make the descriptor more compact and insensitive to local deformations. After the proposed descriptor is calculated, the Dynamic Programming algorithm is employed to accomplish the shape matching task. Experiments in Section 4 demonstrate the excellent discriminative power of this novel shape descriptor.

Using height functions to represent a shape is partly inspired by a recent work (Liu et al., 2008), which provided a novel definition of curvature to discover the extreme points along the curves with the height functions in all the directions. This work demonstrates

* Corresponding author. Tel.: +86 27 87543236; fax: +86 27 87542667.
E-mail addresses: wjw.20012001@163.com (J. Wang), xiang.bai@gmail.com, xbai@hust.edu.cn (X. Bai), youxg@hust.edu.cn (X. You), liuwy@hust.edu.cn (W. Liu), latecki@temple.edu (L.J. Latecki).

0167-8655/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2011.09.042

that the height functions of the sample points of a given contour can represent the geometric changes and deformations. However, the height function was not used in (Liu et al., 2008) as a kind of shape descriptor for shape matching.

1.1. Related work

There are mainly two categories of shape descriptors: contour-based and region-based methods. In general, contour-based shape descriptors exploit only boundary information, and they cannot capture shape interior content. Besides, these methods cannot deal with complex shapes consisting of disjoint parts (Zhang and Lu, 2002). In region-based techniques, shape descriptors are derived using all the pixel information within a shape region. In contrast to contour-based approaches, region-based methods are more reliable for complex shapes such as trademarks, logos and characters (Kim and Kim, 2000). However, (especially some early) region-based methods consist of only global shape characteristics without many important shape details. Therefore, their discriminative power is limited for large databases or complex situations when there are a lot of intra-class variations.

Some well known and recent works for the region-based shape description include Zernike moments (Kim and Kim, 2000), generic Fourier descriptor (Zhang and Lu, 2002), multi-scale Fourier-based description (Direkoglu and Nixon, 2008) and so on. In (Kim and Kim, 2000), a region-based shape descriptor is presented utilizing a set of the magnitudes of Zernike moments. This descriptor has many desirable properties such as rotation invariance, robustness to noise, fast computation and multi-level representation for shape description (Kim and Kim, 2000). Zhang and Lu (2002) propose a generic Fourier descriptor (GFD), which is extracted from spectral domain by applying 2-D Fourier transform on polar raster sampled shape image. Compared with Zernike moments, GFD has no redundant features and allows multi-resolution feature analysis in both radial and angular directions (Zhang and Lu, 2002). In (Direkoglu and Nixon, 2008), the authors believe that the boundary and exterior parts create much more contributions to object recognition than the central part. Based on this observation, they produce new multi-scale Fourier-based descriptors in 2-D space, which represent the boundary and exterior parts of an object and also allow the central part to contribute to shape classification slightly (Direkoglu and Nixon, 2008).

As there have been a lot of contour-based works on shape representation and matching, we only review the most recent approaches. In the last decade, several contour-based shape descriptors have been presented and studied, which are "rich" descriptors with both global and local information. These methods include curvature scale space (CSS) (Mokhtarian et al., 1997), multi-scale convexity concavity (MCC) (Adamek and OĆonnor, 2004), triangle area representation (TAR) (Alajlan et al., 2007, 2008), hierarchical procrustes matching (HPM) (McNeill and Vijayakumar, 2006), shape tree (Felzenszwalb and Schwartz, 2007), contour flexibility (Xu et al., 2009), shape context (SC) (Belongie et al., 2002), inner-distance shape context (IDSC) (Ling and Jacobs, 2007) and so on.

One type of "rich" descriptors is defined in a multi-scale space. curvature scale space (CSS) (Mokhtarian et al., 1997) and multi-scale convexity concavity (MCC) (Adamek and OĆonnor, 2004) are two classical descriptors of this type. In both of them, contour convolution is performed using Gaussian kernel smoothing. By changing the sizes of Gaussian kernels, several shape approximations of the shape contour at different scales are obtained. The descriptors are defined based on these shape approximations. For CSS, a shape feature called Curvature Scale Space image is defined, which records the locations of curvature zero crossings on all evolved curves; the maxima of the curvature zero-crossing contours in the Curvature Scale Space image are used to represent shapes (Mokhtarian et al., 1997). For MCC, the displacements of contour sample points between every two consecutive scale levels are calculated, which are used to represent contour convexities and concavities at different scale levels (Adamek and OĆonnor, 2004). The main limitation for this type of descriptors is that it is difficult to determine the optimal parameter of each scale. Another problem of CSS is that CSS is not a good choice for convex shapes, as there is no curvature zero crossing for convex objects.

Another type of multi-scale descriptors is defined directly on the original shape contours without any preprocessing, including triangle area representation, hierarchical procrustes matching and shape tree. triangle area representation (TAR) (Alajlan et al., 2007, 2008) presents a measure of convexity/concavity of each contour point using the signed areas of triangles formed by boundary points at different scales. The area value of every triangle is a measure for the curvature of corresponding contour point, and the sign of the area is positive, negative or zero when the contour point is convex, concave or on a straight line, respectively. This representation is effective in capturing both local and global characteristics of a shape (Alajlan et al., 2007).

Hierarchical procrustes matching (HPM) (McNeill and Vijayakumar, 2006) and shape tree (Felzenszwalb and Schwartz, 2007) are two classical segment-based shape matching algorithms. In both of them, closed shape contours are divided into curve segments hierarchically, and the matching process is performed by comparing these segments explicitly. In HPM (McNeill and Vijayakumar, 2006), the contour of each shape is divided into overlapped segments with relative arc length percentages 50%, 25% and 12.5% of the whole length. HPM achieves segment matching in a global to local direction, i.e., longer segments that have already been matched together provide an initial match for shorter segments (McNeill and Vijayakumar, 2006). For shape tree, one curve can be broken into two halves by the middle point on it, and each of the two sub-curves can be broken into its halves. This hierarchical description is represented by a binary tree called the *shape tree*. When matching two curves $\boldsymbol{A}$ and $\boldsymbol{B}$, they build a Shape Tree for curve $\boldsymbol{A}$ and search a mapping from points in $\boldsymbol{A}$ to points in $\boldsymbol{B}$ such that the shape tree of $\boldsymbol{A}$ is deformed as little as possible (Felzenszwalb and Schwartz, 2007). Shape tree achieves high retrieval rates on both MPEG-7 data set (87.70%) and Kimia's data set (see in Section 4). However, it suffers from an expensive computational complexity.

Another interesting hierarchical approach is proposed by Payet and Todorovic (2009), which convents the contour matching problem into a graph matching framework for the first time. The shape representation in (Payet and Todorovic, 2009) based on salient contour parts is similar to Felzenszwalb and Schwartz (2007).

Contour flexibility (Xu et al., 2009) is a kind of rich descriptor for planar contours, which depicts the deformable potential at each point along a curve. Contour flexibility provides the information about how extensively the neighborhood of a contour point is connected to the main body and about the deformation tolerance of an object at this point. This method achieves an excellent retrieval result of 89.31% on MPEG-7 data set.

Besides the above descriptors, there is another type of rich descriptors, for which the geometric relationship between contour sample points is utilized. This type of descriptors includes Shape Context and Inner-Distance Shape Context. For every sample point, Shape Context (SC) (Belongie et al., 2002) captures the spatial distribution of all the other sample points relative to it. The spatial distribution is represented by a coarse histogram, and the bins in the histogram are uniform in log-polar space, which makes the descriptor more sensitive to nearby sample points than to points farther away (Belongie et al., 2002). To make contour-based shape descriptors articulation insensitive, Ling and Jacobs proposed one novel