Pattern Recognition 52 (2016) 358–374

Contents lists available at ScienceDirect

Pattern Recognition

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

Hierarchical projective invariant contexts for shape recognition

Qi Jia [a,b], Xin Fan [a,b,*], Yu Liu [c], Haojie Li [a,b], Zhongxuan Luo [a,b], He Guo [a,b]

[a] School of Software Technology, Dalian University of Technology, Dalian, Liaoning, China
[b] Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, China
[c] Leiden Institute of Advanced Computer Science, Leiden University, Leiden, Netherlands

ARTICLE INFO

Article history:
Received 15 May 2015
Received in revised form
1 October 2015
Accepted 7 November 2015
Available online 2 December 2015

Keywords:
Shape descriptor
Perspective deformations
Projective invariants
Coarse to fine

ABSTRACT

Shape descriptors play an important role in various computer vision tasks. Many existing descriptors are typically derived from pair-wise measures, such as distances and angles, which may vary with severe geometrical deformations including affine and projective transformations. In this paper, we propose a new shape descriptor from a newly developed projective invariant, the characteristic number (CN). This new descriptor is invariant to projective (or perspective) transformations by computing CN values on a series of 5 sample points along the shape contour with the intervals varying from coarse to fine. This hierarchical strategy yields a compacter descriptor so that the time complexity for both descriptor construction and shape matching are less or comparable to many existing methods. We also use the derived points out of the contour and the ratio of two invariant values, in order to improve the stability at finer scales and robustness to noise. We demonstrate the performance of the descriptor by comparing with the state-of-the-art on the MCD and other public shape sets with severe perspective transformations and other type variations including noise, missing parts and articulated deformations.

1. Introduction

Shape is one of the most important cues to characterize an object in an image, and shape descriptors play a critical role in various applications including object detection and recognition [1,2], target tracking [3,4] and multimedia retrieval [5]. It still remains a challenging task to design a descriptor that is invariant to various geometrical changes ranging from simple similarity transformations to projective/perspective deformations [6]. Moreover, shape descriptors are typically used as low level features for higher level tasks so that a compact and informative descriptor is desired for efficient recognition and retrieval. In the context of visual search, there exist numerous excellent compact *textural* descriptors that group local ones into a global one [7–10]. In this paper, we focus on the development of a compact *shape* descriptor that is invariant to *projective transformations*, and also stable to noise, part-missing, and articulation.

Traditional shape recognition methods are usually based on *pair-wise* measures between points along the contour [11,12]. These descriptors are able to discriminate subtle shape differences, but resulting in a long feature vector for each point. Some newly developed methods use training or coding strategies [13,14] to generate compact descriptor, but can hardly deal with such severe deformations like affine and perspective transformations. In order to accommodate more degrees of freedoms of transformations, more advanced geometric invariants are introduced into the construction of shape descriptors that are invariant to projective transformations, such as the cross ratio on 4 collinear points [15] and coplanar conic pair [16]. These descriptors are robust to perspective deformations. Nevertheless, most of these methods have the collinear or coplanar constraint, and small deviations on the extraction of contour points may yield mismatches owing to the lack of global contexts.

In this paper, we introduce a novel projective invariant, the characteristic number (CN) [17–19], into the construction of shape descriptors. Naturally, the derived shape descriptors have the invariance to perspective deformations. The geometric invariant CN relaxes such constraints on the underlying geometry like the collinearity of the cross ratio in [15] and the conic pairs of the invariant in [16]. The contributions of this work are listed below:

- We derive a new invariant (FCN) from CN on a series of 5 sample points (minimal number of points to construct a projective invariant) along a shape contour, and develop the robust computation of FCN by (1) deriving a new point out of the contour (named external point) as the substitution of collinear neighboring ones to avoid peak values, and (2) taking the ratio of two FCN values to cancel unstable values. We also provide the theoretical analysis on its invariance and computational robustness.
- We propose a hierarchical strategy to construct a new shape descriptor that recursively calculate the FCN values with

* Corresponding author at: School of Software Technology, Dalian University of Technology, Dalian, Liaoning, China. Tel.: +86 411 62274422.
E-mail address: xin.fan@ieee.org (X. Fan).

http://dx.doi.org/10.1016/j.patcog.2015.11.003
0031-3203/

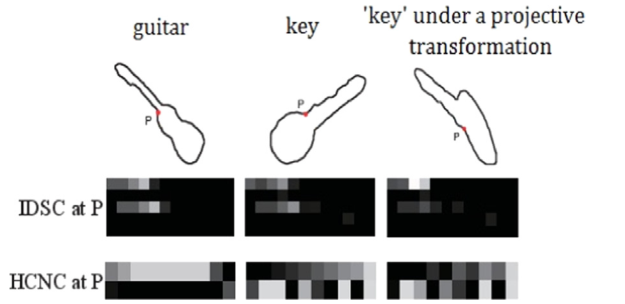

Fig. 1. Feature vectors of HCNC and IDSC at the point P in two similar shapes 'guitar' and 'key' under a projective transformation. The feature values are normalized in a range from 0 to 255 and represented by gray scales with 10 values on each row.

subgroups of 5 contour points by varying intervals from coarse to fine. The descriptor, named hierarchical characteristic number contexts (HCNC), preserves both *global* geometry and *local* contextual information, which balances between accuracy and efficiency. This hierarchical construction can produce a compacter descriptor compared with the pair-wise descriptors including SC [11], IDSC [12], and HF [20]. Also, we are able to deal with both noisy and part-missing shapes.

Fig. 1 gives an example to compare this new descriptor HCNC with the classical IDSC [12] by peering into the features at a contour point. IDSC fails to discriminate the shapes of 'guitar' from 'key' (see the left two columns) and it cannot preserve the features for two versions of 'key' (the right two columns). On the contrary, HCNC presents different features for 'guitar' and 'key', and meanwhile ensures similar features for 'key' under projective deformations. Herein, the dimensionality of our feature vector for each point is 20 (100 sample points in total), about one-third of IDSC.

We employ a simple Euclidean-like similarity metric and the nearest neighborhood (NN) classifier in order to valid the effectiveness of the low level descriptor. Experimental results on the MCD data set [21] show that the calculation of FCN with the construction points and invariant ratios is able to provide accurate recognition with the robustness to noise. We perform more comprehensive comparisons with the state-of-the-art shape descriptors [22,20,13,14] and shape analysis method [6] on the MCD, MPEG-7 and Projective Landmarks data sets [6] with severe perspective deformations. The higher recognition rates of our descriptor built upon the hierarchical strategy demonstrate both the invariance to global transformations and the discriminative power on local shape variations. The experiments on the UMD Activity data set [23] demonstrate that our descriptor has the possibility to handle other deformations than perspective ones. We also validate the effectiveness of our descriptor for part-missing, and articulation. Finally, we provide the comparisons on space and time complexities.

The rest of this paper is organized as follows. We review the related works in Section 2. Then, we provide a modified version of the characteristic number for five points, which is applicable for shape contours without any inner structures in Section 3. Section 4 introduces the definition of hierarchical characteristic number contexts and its computation. The similarity measure on the descriptor for shape matching is given in Section 5. We show the experimental results in Section 6 and Section 7 concludes the paper.

2. Related work

In this section, we list the descriptors derived from the classical shape context and those invariant to affine or projective transformations. Additionally, we provide the algorithms to align contour samples, which are typically necessary for shape matching with a descriptor.

2.1. Shape context based methods

It has been long recognized in computer vision to use measures on contour points of a shape for the characterization of shape information [24,25]. In [11], Belongie et al. develop the shape context (SC) by the histogram of pair-wise distance and orientation between contour points in the log-polar space. The shape context has been extended in various ways in order to accommodate a wider range of shape changes [22]. Ling and Jacobs replace the Euclidean distance in SC with the inner distance (IDSC) that is robust to articulations [12]. Zhao et al. exploit a polynomial fitting-based feature point extraction method as a preprocessing step, so as to enhance the performance of the shape contexts-based descriptor under translation, rotation and scaling [26]. In [20], Wang et al. propose the height function (HF) descriptor, in which one sample point is characterized by the distances of all the other sample points to its tangent line. Some works combine these methods or use triangle-area formed by the sample points in order to improve the performance [27–29]. These descriptors are able to discriminate subtle shape differences, but they represent shapes by *pair-wise* measures between points along the contour, resulting in a long feature vector for each point.

Motivated by the middle level feature 'bag-of-features' in multimedia retrieval, Wang et al. develop the bag of contour fragments (BCF) and generate a *compact* code as the descriptor [13]. In [14], the descriptor is constructed by coded angular structures on clusters of contour points. There are also some other works using curvatures or shock graphs to generate a compact descriptor [30,31]. These methods are likely to provide a compact shape representation robust to noise, but can hardly deal with such severe deformations like affine and perspective transformations.

2.2. Affine or projective invariant shape descriptor

Researchers have proposed features invariant to affine transformations consisting of rotation, scaling, translation and shearing. The Fourier representation of the closed object contour is widely used to achieve affine-invariant shape recognition [32]. Similar descriptors using measures in the transformed domain include covariant conic [33], locally geometric invariants [34], dyadic wavelets [35], and cosine transforms [36]. Raviv and Kimmel [37] add the similarity invariance to an equi-affine invariant, and establish an affine invariant pseudo-metric. In our previous work [38], we present an affine invariant descriptor for symbols with rich inner structures. There are also some region-based methods use texture descriptor or Radon transform to handle the shapes under affine deformations [39,40].

Object shapes typically undergo more general projective transformations in practical applications especially when capturing an object from multiple viewpoints. More advanced geometric invariants are introduced into the construction of shape descriptors that are invariant to projective transformations. Xu et al. devise a set of moment descriptors from multiple integral of geometric quantities [41]. Early in 1998, the cross ratio, a fundamental invariant in projective geometry, is used to construct a projective signature for planar shapes [42]. Li and Tan construct a descriptor robust to severe perspective deformations by using the cross ratio on 4 collinear points [15]. Other projective invariants for a coplanar conic pair are also employed in the construction of shape descriptors in [16]. These descriptors have the invariance for perspective deformations. Nevertheless, small deviations on