Pattern Recognition 47 (2014) 1891–1903

Contents lists available at ScienceDirect

Pattern Recognition

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

Spectra of shape contexts: An application to symbol recognition

Su Yang*

Shanghai Key Laboratory of Intelligent Information Processing, College of Computer Science and Technology, Fudan University, Shanghai 201203, China

ARTICLE INFO

Article history:
Received 30 September 2012
Received in revised form
26 October 2013
Accepted 18 November 2013
Available online 1 December 2013

Keywords:
Symbol recognition
Shape matching
Histogram-based descriptor
Shape context
FFT
Histogram comparison
Point pattern matching
Maximum clique
Image retrieval
Object recognition

ABSTRACT

The pixel-level constraint (PLC) histograms are known for robustness and invariance in symbol recognition but limited in $O(N^3)$ complexity. This paper proves that matching two PLC histograms can approximately be solved as matching the power spectra of the corresponding shape contexts. As a result, spectra of shape contexts (SSC) inherit robustness and invariance from PLC while the computational cost can be reduced. Moreover, a maximum clique based scheme is proposed for outlier rejection. The theoretical and experimental validation justifies that SSC possesses the desired properties for symbol recognition, that is, robustness, invariance, and efficiency. It outperforms PLC in terms of robustness and time efficiency, and shape context in terms of rotation invariance.

1. Introduction

Shape matching and recognition is one of the fundamental problems in computer vision. Shape context is one of the well-known shape descriptors in the literature [7]. However, its rotation invariance is not reliable, which is based on taking the tangent at each point of interest as the reference direction. In case the estimated tangents do not adhere to the true tangents exactly, the rotation invariance will not be guaranteed. Besides, for some applications like symbol recognition, no tangent can be obtained. In view of such weakness of shape context, the author proposes a descriptor for symbol recognition referred to as pixel-level constraint (PLC) histogram [10]. PLC makes use of virtual vectors to figure out the constraints among feature points so as to guarantee rotation and scale invariance. The experiments justify that it not only promises rotation and scale invariance but also tolerates degradation of image quality and shape deformation. However, the $O(N^3)$ complexity limits its usages. Hence, seeking a shape descriptor that can promise robustness, invariance, and efficiency at the same time is always the goal for symbol recognition and shape matching. In [22], we prove that the 2-dimensional Fast Fourier Transform (2D-FFT) of shape context is rotation invariant. However, this is just an initial effort and not enough to lead to a sound shape descriptor. In this paper, it is demonstrated that spectra of shape context (SSC) should be a descriptor possessing the aforementioned properties that we desire. The main contributions of this work are: (1) through theoretical analysis, it is revealed that PLC histogram can be approximately regarded as autocorrelation of shape context. Correspondingly, the frequency-domain representation of such autocorrelation is spectra of shape context. Then, it is proved that the similarity between two PLC descriptors can be totally solved in frequency domain as the inner product of the two corresponding SSC descriptors, the computational complexity of which relies mainly on the $O(N\log_2 N)$ complexity in computing FFT. Since SSC can approximately be regarded as frequency-domain representation of PLC and the similarity between SSCs is roughly equal to that between PLCs, SSC inherits all the advantages from PLC, that is, robustness and invariance. (2) The experiments show that SSC performs better than PLC in terms of robustness while overcomes the shortcoming of SC in rotation invariance. (3) The time cost in matching SSC is much lower than that in matching PLC and SC. (4) A maximum clique based outlier rejection method is proposed to filter out the erroneous matches resulting from the descriptor-level matching. Following outlier rejection, the preserved matches are geometrically consistent with each other. The final matching score is then computed in the sense of point set matching, where a similar transformation computed from the refined matches is applied to align the two point sets.

---

* Tel.: +86 21 5135552; fax: +86 21 51355558.
E-mail address: suyang@fudan.edu.cn

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.11.020

The proposed solution is validated through two symbol recognition benchmarks, GREC 2003 and GREC 2011, and the CMU house benchmark, which is widely used for evaluating point set matching.

As for GREC 2003 benchmark, SSC achieves 100% accuracy in recognizing symbols with degradation and deformation while rotation and scale invariance can be promised to some extent, above 90%. The combination of SC with the proposed outlier rejection method also leads to as good as close to 100% accuracy in recognizing deformed and degraded symbol images but it misses almost all rotated symbols. So, SSC gains advantage over SC in the sense of rotation invariance. In comparison with PLC, the recognition accuracy of SSC is a little bit better in the presence of deformation or degradation but the rotation and scale invariance performance of SSC is inferior to that of PLC. Nevertheless, the main advantage of SSC over PLC lies in the reduced computational cost by applying FFT.

As for GREC 2011 benchmark, the overall performance of SSC in recognizing symbols with degradation, rotation, and scaling is comparable to the best performance reported in the literature [18,25] but SSC performs better for large data set.

To further compare SSC with SC in the sense of descriptor-level discriminative power, CMU house data is used to evaluate how sensitive the two descriptors are subject to position variation of the key points caused by 3-dimensional (3D) motion of the object of interest. The experiments show that SSC performs better than SC.

The rest of the paper is organized as follows. In Section 2, the related works are reviewed. The proposed method is presented in Section 3. In Section 4, the experimental results are provided. The conclusion is made in Section 5.

2. Related works

2.1. Overview

Symbol recognition plays an important role in graphics recognition like architectural drawing understanding [1]. According to the pattern primitive applied, the existing methods can be sorted into three classes: vector-based methods [2–6], pixel-based methods [7–13,24], and the methods combining both vector and pixel-based descriptors [14,15]. For vector-based approaches, in general, the binary images of interest should undergo the so-called vectorization procedure at first to render a graph-based pattern representation, where the vertices of the graph are usually the vectors resulting from vectorization while the edges of the graph reflect the topological or geometrical relations among such vectors. Then, graph matching or grammar-based parsing is applied to achieve final recognition. The major advantage of vector-based approaches is that rotation and scale invariance is guaranteed with ease. For pixel-based approaches, the pattern descriptor is usually a statistical model that figures out the distribution of the feature points. The primary advantage of pixel-based approaches is that such methods do not rely on exact pattern matching as vector-based approaches do so that robustness can be promised. Yet, many problems remain unsolved for both methodologies. For vector-based methods, the challenge is how to obtain robust vectorization. For pixel-based methods, in general, the descriptor itself does not possess rotation invariance. Due to the limits of both vector and pixel-based methods, some investigations trying to combine both methodologies are conducted [14,15]. Yet, simply pooling the two kinds of descriptors into one feature or switching between them based on some heuristic clues cannot overcome the drawback of either kind of descriptor. Although machine learning can be applied as an alternative way for improving the performance of symbol recognition [16,17,23,24], in case no enough training samples can be obtained, the key problem returns to how discriminative the applied descriptor is. In contrast to the aforementioned endeavors, the pixel-level constraint (PLC) histogram proposed previously by the author [10] is unique in that it makes use of virtual vectors to figure out statistically the constraints among pixels. This accounts for why it can guarantee invariance while be robust for symbol recognition. However, PLC suffers from the high computational complexity of $O(N^3)$.

To reach reasonable computational complexity while preserving the desired properties of PLC, in this study, a new descriptor based on spectra of shape context is proposed. Note that FFT is also employed in [8] to represent the shape of symbols, which is the so-called F-signature. In contrast, the distinct contribution of this study is: a more promising shape descriptor coupled with a well-fitted similarity measure is developed on a theoretically rigorous basis.

Following descriptor-level coarse match, the author proposes to filter out inconsistent matches via maximum clique in terms of graph theory, and has developed an approximate algorithm to solve it. After the outlier rejection, a similarity/affine transformation is applied to align the two feature point sets. As a result, the similarity between two images can be obtained at the point set matching level. Compatible with our point of view, point-to-point matching has been applied as a promising scheme for both symbol recognition and symbol spotting in [18]. Note that maximum clique has already been applied in early pattern recognition researches for point pattern matching [19]. In contrast, this work is unique in that maximum clique is performed prior to affine/similar transform in order to find out a reasonable affine/similar transform that can align the two point sets of interest to the best extent. In [19], however, maximum clique is applied following the affine transformation and only two matched point pairs are used to judge the local consistency. Here, the dilemma lies in that if we can obtain a well-fitted transformation, it is unnecessary to apply any correction scheme like maximum clique.

2.2. Pixel-level constraint histograms

In [10], the author proposes a pattern descriptor referred to as pixel-level constraint (PLC) histograms for symbol recognition. The procedure to construct the descriptor is as follows: first, interesting points are extracted from the binary symbol image of interest by applying a thinning and sparse sampling scheme. Suppose that there are in total K interesting points extracted, which are denoted as $P_0,P_1,\ldots,P_{K-1}$. Then, let every point act as the reference point alternately to construct a pattern descriptor in association with it. Without loss of generality, as shown in Fig. 1, let P_0 be the reference point as an example: (1) connect P_0 to every point in $\{P_1,P_2,\ldots,P_{K-1}\}$ to form $K-1$ vectors and denote the length and the angle of every vector $V_i=\overrightarrow{P_0P_i}$ as r_i and α_i, $i=1,2,\ldots,K-1$. (2) Compute the length ratio and the angle difference between every two vectors

$$r_{ij}=r_i/r_j \tag{1}$$

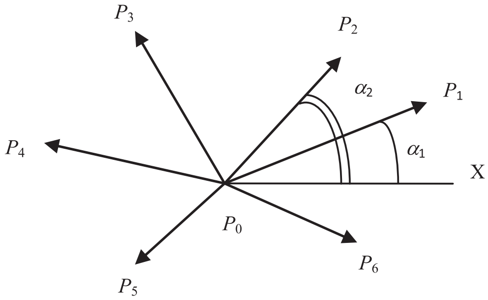

Fig. 1. An example of vector-represented pixel-level constraints.