Information Sciences xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Information Sciences

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

Beyond diffusion process: Neighbor set similarity for fast re-ranking

Q1 Xiang Bai, Song Bai, Xinggang Wang*

Department of Electronics and Information Engineering, Huazhong Univ. of Science and Technology, N1 Hall, 1037 Luoyu Road, Hongshan, Wuhan 430074, PR China

ARTICLE INFO

Article history:
Received 6 January 2014
Revised 20 May 2015
Accepted 4 July 2015
Available online xxx

Keywords:
Shape retrieval
Image retrieval
Re-ranking
Contextual information
Diffusion process

ABSTRACT

Measuring the similarity between two instances reliably, shape or image, is a challenging problem in shape and image retrieval. In this paper, a simple yet effective method called Neighbor Set Similarity (NSS) is proposed, which is superior to both traditional pairwise similarity and diffusion process. NSS makes full use of contextual information to capture the geometry of the underlying manifold, and obtains a more precise measure than the original pairwise similarity. Moreover, based on NSS, we propose a powerful fusion process to utilize the complementarity of different descriptors to further enhance the retrieval performance. The experimental results on MPEG-7 shape dataset, N-S image dataset and ORL face dataset demonstrate the effectiveness of the proposed method. In addition, the time complexity of NSS is much lower than diffusion process, which suggests that NSS is more suitable for large scale image retrieval than diffusion process.

1. Introduction

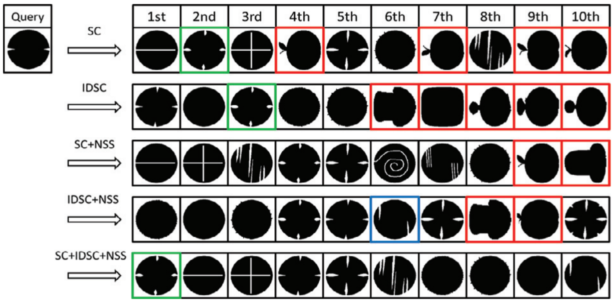
Shape or image retrieval is a fundamental issue in computer vision with many applications. Given a query, all the instances in the database are sorted in an ascending (or descending) order based on their dissimilarity (similarity) to the query. Then, the ranking list of the query is initialized, where the most similar instance occupies the top position. Many researchers have focused on designing robust, informative and discriminative descriptors [3,11,15,18–20,24,28,33] in order to achieve better retrieval performance. However, these basic methods totally ignore the structure of the underlying data manifold, thus cannot generate satisfactory retrieval results.

In order to capture the geometry of the underlying manifold, many context-sensitive similarity measures [9,13,30,31,36–42,47] are proposed to improve the retrieval accuracy. Diffusion process, one of the most representative branches in context-based re-ranking, starts with constructing a weighted graph base on the graph theory [6], and uses the nodes to represent the visual instances. The edge connecting two nodes represents their pairwise similarity. Diffusion process conducts a random walk to spread the similarity through the graph, in which a transition matrix is used. The transition matrix usually interprets the similarity after normalization as the transition probability from one node to another. The computation of transition probability is usually relevant to the local distribution of the data manifold, which makes diffusion process robust to noise and outliers.

It seems that diffusion process is an indispensable tool for improving retrieval performance. However, it also has a disadvantage of computational expensive. Some iterative methods [12,36,37] require many computational steps to converge. These

* Corresponding author. Tel.: +86 27 87543236; fax: +86 27 87543236.
E-mail addresses: xbai@hust.edu.cn (X. Bai), songbai@hust.edu.cn (S. Bai), xgwang@hust.edu.cn, wxghust@gmail.com (X. Wang).

http://dx.doi.org/10.1016/j.ins.2015.07.022
0020-0255/© 2015 Published by Elsevier Inc.



Q2

Fig. 1. Top to bottom: the retrieval results of a certain query from MPEG-7 dataset measured by SC, IDSC, NSS+SC, NSS+IDSC and NSS+IDSC+SC. The false retrieval results are put in red boxes. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

iterative methods are usually time-consuming, and improper for large scale re-ranking. Some other approaches [7,31] although have closed-form solutions, but they require complex operations, such as computing the inverse of matrices whose size is often proportional to the scale of the database. Such operations are computationally prohibited, when the size of database becomes larger. Such a shortcoming limits the usage of many diffusion-based algorithms in real-time search engines. Following the same principal as diffusion process, we propose Neighbor Set Similarity (NSS) to speed up the re-ranking procedure. Unlike the conventional approaches, NSS does not need an iterative process, resulting in much higher efficiency while keeping the re-ranking accuracy.

Besides the algorithms that focus on enhancing one type of similarity measure, some methods [1,30,32,44,45,48,49] are proposed to fuse multiple kinds of similarity measures for re-ranking, since different similarities may be complementary to each other. For example, as two popular shape descriptors, Shape Context (SC) [3] encodes the global information of a shape and generally works well with rigid objects, while Inner Distance Shape Context (IDSC) [15] replaces the Euclidean distance used in SC by the geodesic distance, and is more suitable for non-rigid analysis. It seems difficult to design a generic descriptor that can handle all the properties under different conditions, which inspires us to exploit a framework to fuse multiple complementary similarity measures. It is straightforward that a better performance can be achieved when the complementarity is used in a proper way. In this paper, based on NSS we propose a more powerful fusion method inspired by the co-training algorithm [5], which yields a much more precise retrieval result. However, unlike co-training that assumes views (sets of features) with two conditions, NSS deals with single-view but multiple-input similarity measures for robust re-ranking.

Fig. 1 shows the retrieval results when querying a given shape from MPEG-7 dataset [14], as measured by SC, IDSC, SC+NSS, IDSC+NSS and SC+IDSC+NSS. The false results are surrounded by red boxes. The first two rows show the retrieval results measured by SC and IDSC, and obviously several outliers exist in the ranking list. As the third row shows, NSS with SC as the input measure is more robust to noise compared with using SC only. The fourth result of SC is ranked in the 9th position in the retrieval result of NSS+SC, and some outliers are even excluded. Moreover, NSS is also able to find instances that are not in the original retrieval list. As the fourth row shows, the sixth result in the blue box is newly found, and it is not in the top-10 retrieval results of IDSC at first. What is more important is that NSS with two input measures can improve the retrieval results significantly by utilizing the complementarity of SC and IDSC. The instances that are ranked high in both measures will obtain a higher position in the retrieval result of NSS. For example, the instance in the green box, which holds the second position of SC and the third position of IDSC, is ranked first in the fifth row. This example shows NSS can utilize contextual information as well as multiple features to improve the retrieval performance.

The rest of this paper is organized as follows: In Section 2 we briefly revisit the related works. The motivation and definition of NSS are given in Section 3. A study of the comparison between NSS and diffusion process is given in Section 4. In Section 5, we conduct some experiments on several benchmark datasets to demonstrate the advantages of the proposed method again. Conclusions are given in Section 6.

2. Related work

In this section, we provide an overview of basic descriptors, diffusion process, feature fusion and kNN selection algorithms.

2.1. Descriptors

Many shape descriptors have been proposed in the literatures recently. Shape Context (SC) proposed in [3] works well for rigid objects, and Inner Distance Shape Context (IDSC) proposed in [15] is better at dealing with articulated shapes. Gopalan et al. [11] propose Articulation-Invariant Representation (AIR) by modeling an articulating shape as a combination of approximate convex parts connected by non-convex junctions. In [33], the contour of each shape is represented by a fixed number of sample points, and a height function is defined based on the distances of the other sample points to its tangent line. A more complicated matching method is introduced in [8], where SC is used to find the correspondence with dynamic programming,

Please cite this article as: X. Bai et al., Beyond diffusion process: Neighbor set similarity for fast re-ranking, Information Sciences (2015), http://dx.doi.org/10.1016/j.ins.2015.07.022