Graphical Models 98 (2018) 33–42

Contents lists available at ScienceDirect

Graphical Models

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

Class-sensitive shape dissimilarity metric

Manyi Li[a], Noa Fish[b], Lili Cheng[a], Changhe Tu[*,a], Daniel Cohen-Or[b], Hao Zhang[c], Baoquan Chen[a]

[a] Shandong University, China
[b] Tel Aviv University, Israel
[c] Simon Fraser University, Canada

ARTICLE INFO

Keywords:
Dissimilarity metric
Fine-grained classes
Distinctive attributes

ABSTRACT

Shape dissimilarity is a fundamental problem with many applications such as shape exploration, retrieval, and classification. Given a collection of shapes, existing methods develop a consistent global metric to compare and organize shapes. The global nature of the involved shape descriptors implies that overall shape appearance is compared. These methods work well to distinguish shapes from different categories, but often fail for fine-grained classes within the same category. In this paper, we develop a dissimilarity metric for fine-grained classes by fusing together multiple distinctive metrics for different classes. The fused metric measures the dissimilarities among inter-class shapes by observing their unique traits. We demonstrate the advantage of using our approach in several applications.

1. Introduction

Shape dissimilarity is one of the most fundamental problems in computer graphics and various other fields, with numerous applications such as shape exploration, retrieval and classification [37]. Recent advances in shape analysis call for the development of semantic-aware metrics and fine-grained dissimilarity measures [18,24].

Many shape dissimilarity methods compare instances based on global shape descriptors [3,7,26,32], which encode and capture overall shape geometry. These global descriptors are designed for high-level classification and are generally insensitive to fine-grained dissimilarities. Other methods [12,18,25,28] deliberately emphasize fine-grained distinction, but they are class-oblivious in that they apply the same measure to all classes without taking into account the possibly unique characteristics of each particular class.

The premise of this work is that different classes share different commonalities (*e.g.*, parts, geometric features, *etc.*), and hence require different fine-grained dissimilarity measures to better capture their fine differences. Therefore, the dissimilarity measures employed for inter-class and intra-classes distinctions should be tailored according to the observed commonalities within the various classes. Our key idea is therefore to compose an inter-class dissimilarity metric based on common class characteristics, and fuse multiple fine-grained metrics together into one general dissimilarity metric. This metric then garners the unique properties of the individual metrics, and is able to perform high-level differentiation of inter-class shapes.

The rest of the paper is organized as follows. Section 2 contains an overview of previous related works on the subject of shape dissimilarity. Our approach is presented in detail in Section 3, which starts with a short method overview, and continues to discuss our chosen shape representation, class-specific metric and metric fusion. Results and comparisons are shown in Section 4 and we conclude in Section 5.

2. Related works

Shape dissimilarity is a central problem in computer graphics with many applications and therefore a rich body of previous work. A common thread connecting the majority of these works is the reliance on shape descriptors that are designed and computed to capture important characteristics of the shapes such that they are conveniently comparable to one another [37]. Some shape descriptors are composed of geometric features or signatures [16,17,31]. For instance, the light field descriptor (LFD) [10,29] aggregates features from multi-view images to represent a shape. The scale-invariant heat kernel signature (SI-HKS) [8] extracts local and global geometric properties of points on the 3D shape. These feature descriptors put more focus on shape geometry such that shape topology is somewhat overlooked. On the other hand, a shape descriptor can also be computed from different shape representations, such as graph representation [6,20] and skeleton representation [36], to preserve the topology or the connections between shape components. Shape descriptors aim to represent and capture the overall appearance of a shape, but often do not delve deeper into its

* Corresponding author.
E-mail address: chtu@sdu.edu.cn (C. Tu).

https://doi.org/10.1016/j.gmod.2018.06.002
Received 19 October 2017; Received in revised form 9 June 2018; Accepted 22 June 2018
Available online 23 June 2018
1524-0703/

semantic meaning.

With the expansion of readily available shape repositories containing a wide variety of different types of shapes, attention and emphasis put on semantics as a tool for analysis has grown. Correspondence between shapes and their parts is a strong cue for shape dissimilarity that aligns with human intuition. Many methods first employ a shape matching process to align shape parts and then compute shape dissimilarity by summing up the differences between part descriptors [5,30]. In this manner, Alhashim et al. [1] and Kleiman and van Kaick [28] use the cost of deformation or editing operations between corresponding parts to quantify dissimilarity, and Averkiou et al. [2] use the parameters of oriented bounding boxes to represent parts. Recently, following the success of deep learning methods applied on various problems, high-level shape descriptors that are learned using deep networks on collections of shapes [23,35,38,39], have been shown to provide good performance for classification and retrieval, suggesting that these learned features possess some notion of shape semantics. However, it is not uncommon for shapes belonging to the same fine-grained semantic class to agree on a subset of their attributes while, often significantly, disagreeing on others. Thus, in cases such as these, we conclude that global metrics learned at the category level cannot easily be used to distinguish between fine-grained classes.

To address this issue, some works opt to analyze distinctive regions of shapes [33,34], which are common among shapes of the same class but uncommon for others, such that the intra-dissimilarity values are small when the focus is on the distinctive regions. Kim et al. [27] compare shapes based on regions of interest selected by users and Gao et al. [18] refine the distinctive regions iteratively by requesting users to select similar and dissimilar shapes. Huang et al. [24] extract the distinctive regions for multiple fine-grained classes separately with a semi-supervised metric learning method. These metrics can compare shapes based on certain distinctive regions, but each metric can only handle one class. As different fine-grained classes have different distinctive attributes, these methods fail to separate the classes using a single metric. As demonstrated by Frome et al. [14] and Hong et al. [22], an effective way to separate classes is by fusing multiple local finer-scale metrics. Some works fuse multiple metrics by simply concatenating the anchors [15] and others by summing up the weighted differences of each metric [40]. Since the weights are fixed for all shapes, they are utilizing the information present in the category in its entirety to define one global metric to compare all pairs of shapes. We observe that certain metrics may not necessarily apply to all shapes, thus using them collectively often carries little meaning.

Considering the highly non-linear nature of the local features of fine-grained classes, employing deep learning methods to extract the significant aspects of the features is a natural choice. Data-driven methods often require large amounts of labeled data to support good generalization over the learned space, such that meaningful deep features can be extracted for the training data [4]. When the focus is put on the division of a class into sub-categories, acquiring a large labeled dataset is not straight-forward, since the fine-grained classes may overlap with each other, and they are not well defined. For the purpose of handling small labeled datasets, the Siamese network [11,21], has been proposed, where the set is organized into positive and negative pairs. Naturally, the training quality of the Siamese network highly depends on the input pairs and the distinction between positive and negative, a requirement which cannot necessarily be fulfilled when fine-grained classes overlap.

In this paper, we develop a dissimilarity metric to separate multiple fine-grained classes by fusing together several distinctive metrics. Differently from existing methods, our metric uses a different fused basis for each pair of shapes based on their relevant classes, rendering our metric sensitive to fine-grained classes. The embedding generated naturally by our fused dissimilarity metric shows its ability to recognize fine-grained classes, and separate them physically. We show that our dissimilarity metric demonstrates improved performance on shape retrieval and classification applications.

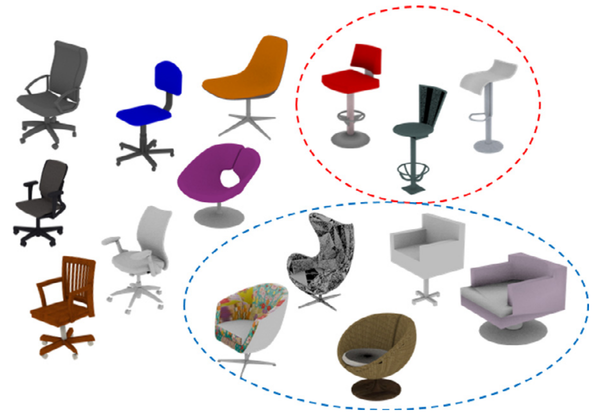

Fig. 1. A reference set of swivel chairs. The shapes in this class exhibit large variation in the backrest, seat and handles. The shapes circled in red belong to both the swivel chair and bar chair classes. The shapes circled in blue belong to both the swivel chair and sofa classes. However, the swivel sofas may not be selected by a user to appear in the sofa reference set. Thus, we expect our input, consisting of small labeled reference sets, to be far from complete. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3. Method

3.1. Overview

To develop a shape dissimilarity metric in order to distinguish between fine-grained classes, we ought to first form a notion of the classes we are working with. A fine-grained class is a flexible and loose concept, and is often a matter of semantics. This leads to a highly diverse space of classes that is hard to define globally. For example, a class may exhibit a common overall style, e.g., sofas, while another may favor common local regions, e.g., swivel chairs. Thus, forming an idea of the nature of our fine-grained classes from the input, is an essential first step. Accordingly, the input to our problem consists of multiple small reference sets representing fine-grained classes, such that a user is only required to select a few representative shapes per class. Note that we do not assume a complete labeling of the reference sets at all, for example, a swivel sofa may appear in the swivel chair reference set, but not be selected to appear in the sofa set, as shown in Fig. 1.

Given several small reference sets as input, we first learn a consistent shape representation to capture the characteristics of our shapes. Next, we leverage metric learning to develop a distinct metric for each class, so that inter and intra class distinctions can be separated. Lastly, we fuse the various learned metrics together by considering the metrics for relevant classes for each pair of shapes. These steps are described in detail below.

3.2. Shape representation

Finding the commonalities and differences between shapes necessitates a conveniently comparable representation. A plethora of shape features have been designed and utilized by many for various comparison-based tasks. Our approach is oblivious to the choice of features, and only requires consistency among them.

Recently, deep autoencoders have been shown to be a powerful tool for representation learning. An autoencoder is composed of an encoder component followed by a decoder component, such that the purpose of the encoder is to extract the most crucial information from the input data and encode it to a representation that can be successfully reconstructed back to the input, by the decoder component.

For our experiments, we opt to train a simple deep autoencoder for