Improving Shape Retrieval by Learning Graph Transduction

Xingwei Yang[1], Xiang Bai[2,3], Longin Jan Latecki[1], and Zhuowen Tu[3]

[1] Dept. of Computer and Information Sciences, Temple University, Philadelphia
{xingwei,latecki}@temple.edu
[2] Dept. of Electronics and Information Engineering, Huazhong University of Science and Technology, P.R. China
xiang.bai@gmail.com
[3] Lab of Neuro Imaging, University of California, Los Angeles
zhuowen.tu@loni.ucla.edu

Abstract. Shape retrieval/matching is a very important topic in computer vision. The recent progress in this domain has been mostly driven by designing smart features for providing better similarity measure between pairs of shapes. In this paper, we provide a new perspective to this problem by considering the existing shapes as a group, and study their similarity measures to the query shape in a graph structure. Our method is general and can be built on top of any existing shape matching algorithms. It learns a better metric through graph transduction by propagating the model through existing shapes, in a way similar to computing geodesics in shape manifold. However, the proposed method does not require learning the shape manifold explicitly and it does not require knowing any class labels of existing shapes. The presented experimental results demonstrate that the proposed approach yields significant improvements over the state-of-art shape matching algorithms. We obtained a retrieval rate of **91%** on the MPEG-7 data set, which is the highest ever reported in the literature.

1 Introduction

Shape matching/retrieval is a very critical problem in computer vision. There are many different kinds of shape matching methods, and the progress in increasing the matching rate has been substantial in recent years. However, all of these approaches are focused on the nature of shape similarity. It seems to be an obvious statement that the more similar two shapes are, the smaller is their difference, which is measured by some distance function. Yet, this statement ignores the fact that some differences are relevant while other differences are irrelevant for shape similarity. It is not yet clear how the biological vision systems perform shape matching; it is clear that shape matching involves the high-level understanding of shapes. In particular, shapes in the same class can differ significantly because of distortion or non-rigid transformation. In other words, even if two shapes belong to the same class, the distance between them may be very

D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part IV, LNCS 5305, pp. 788–801, 2008.

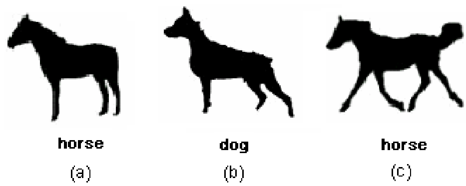

Fig. 1. Existing shape similarity methods incorrectly rank shape (b) as more similar to (a) than (c)

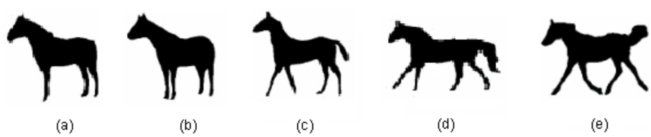

Fig. 2. A key idea of the proposed distance learning is to replace the original shape distance between (a) and (e) with a geodesic path in the manifold of know shapes, which is the path (a)-(e) in this figure

large if the distance measure cannot capture the intrinsic property of the shape. It appears to us that all published shape distance measures [1,2,3,4,5,6,7] are unable to address this issue. For example, based on the inner distance shape context (IDSC) [3], the shape in Fig. 1(a) is more similar to (b) than to (c), but it is obvious that shape (a) and (c) belong to the same class. This incorrect result is due to the fact that the inner distance is unaware that the missing tail and one front leg are irrelevant for this shape similarity judgment. On the other hand, much smaller shape details like the dog's ear and the shape of the head are of high relevance here. No matter how good a shape matching algorithm is, the problem of relevant and irrelevant shape differences must be addressed if we want to obtain human-like performance. This requires having a model to capture the essence of a shape class instead of viewing each shape as a set of points or a parameterized function.

In this paper, we propose to use a graph-based transductive learning algorithm to tackle this problem, and it has the following properties: (1) Instead of focusing on computing the distance (similarity) for a pair of shapes, we take advantage of the manifold formed by the existing shapes. (2) However, we do not explicitly learn the manifold nor compute the geodesics [8], which are time consuming to calculate. A better metric is learned by collectively propagating the similarity measures to the query shape and between the existing shapes through graph transduction. (3) Unlike the label propagation [9] approach, which is semi-supervised, we treat shape retrieval as an unsupervised problem and do not require knowing any shape labels. (4) We can build our algorithm on top of any existing shape matching algorithm and a significant gain in retrieval rates can be observed on well-known shape datasets.

Given a database of shapes, a query shape, and a shape distance function, which does not need to be a metric, we learn a new distance function that is