Journal of Theoretical Biology 336 (2013) 185–199

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

Analysis and classification of collective behavior using generative modeling and nonlinear manifold learning

Sachit Butail [a], Erik M. Bollt [b], Maurizio Porfiri [a,*]

[a] Department of Mechanical and Aerospace Engineering, Polytechnic Institute of New York University, Brooklyn, NY 11201, USA
[b] Department of Mathematics and Computer Science, Clarkson University, Potsdam, NY 13699, USA

HIGHLIGHTS

- We present a data-driven approach to directly classify raw images of collective behavior.
- We rapidly create training images that match attainable far-field videos of real animal groups.
- We validate the setup on datasets of collective motion with increasing complexity.
- We demonstrate the proposed framework to classify raw videos of schooling zebrafish.

ARTICLE INFO

Article history:
Received 7 February 2013
Received in revised form
2 July 2013
Accepted 31 July 2013
Available online 2 August 2013

Keywords:
Classification
Collective motion
Fish schooling
Generative modeling
Isomap

ABSTRACT

In this paper, we build a framework for the analysis and classification of collective behavior using methods from generative modeling and nonlinear manifold learning. We represent an animal group with a set of finite-sized particles and vary known features of the group structure and motion via a class of generative models to position each particle on a two-dimensional plane. Particle positions are then mapped onto training images that are processed to emphasize the features of interest and match attainable far-field videos of real animal groups. The training images serve as templates of recognizable patterns of collective behavior and are compactly represented in a low-dimensional space called embedding manifold. Two mappings from the manifold are derived: the manifold-to-image mapping serves to reconstruct new and unseen images of the group and the manifold-to-feature mapping allows frame-by-frame classification of raw video. We validate the combined framework on datasets of growing level of complexity. Specifically, we classify artificial images from the generative model, interacting self-propelled particle model, and raw overhead videos of schooling fish obtained from the literature.

1. Introduction

The study of collective behavior is often complemented with the analysis of high-volume datasets available in the form of simulated trajectories (Vicsek and Zafeiris, 2012; Romanczuk et al., 2012; Frewen et al., 2011) and videos (Ballerini et al., 2008). Patterns in these datasets are recognizable to a trained observer, who can quickly determine whether a set of particles or a school of fish is moving together in coordination or in complete disorder. However, this standard of recognition is not available at a machine level, where we must classify the trajectory data by fitting it into activity models (Choi et al., 2009; Patterson et al., 2009; Li and Chellappa, 2010). The intermediate process of multi-target tracking is a computational overhead that scales with the number of animals observed (Delcourt et al., 2009; Parrish and Hammer, 1997). Although a naive comparison of images to an exhaustive database of training videos would preclude the need to track individuals, it would shift the computational burden from processing to storage. Instead, an enabling requirement for a fast, data-driven approach is to store recognizable patterns of collective behavior in a compact representation so that they can be archived and retrieved for quick comparison.

Assuming no coordination among individuals, we expect the trajectories of group members to be independent of each other, thus requiring a large number of degrees of freedom to describe the group motion; trajectories of coordinated individuals should instead be manifested through fewer degrees of freedom related to the movement of select group members. This realization forms the first step to reduce the data to a few important features that can faithfully represent the ongoing process. For example, images of animal groups may be classified on the basis of features of the

* Corresponding author. Tel.: +1 718 260 3681.
E-mail addresses: sb4304@nyu.edu (S. Butail), bolltem@clarkson.edu (E.M. Bollt), mporfiri@poly.edu (M. Porfiri).

0022-5193/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2013.07.029

spatial distribution, such as number of subgroups, population density, and group configuration, and features of the dynamics, such as change in group size, orientation, and speed. As a group of animals maneuver through space, temporal variation of these features can inform the nature of the interaction between them (Katz and Tunstrøm, 2011; Herbert-Read et al., 2011; Aureli et al., 2012).

Dimensionality reduction is the process of identifying low-dimensional representations that preserves the dissimilarities between points on a high-dimensional space (Cayton, 2005; Belkin et al., 2006). The high-dimensional data may be in the form of positions of multiple individuals or even raw images. Once generated, these representations, called embedding manifolds, can be used in a class of machine learning algorithms, called manifold learning, for analysis and classification. A similar approach is successfully implemented in human activity recognition (Elgammal and Lee, 2007; Blackburn and Ribeiro, 2007), face recognition (Yang, 2002), pose estimation (Elgammal and Lee, 2007; BenAbdelkader, 2010), exploration of video sequences (Pless, 2003), and handwriting recognition (Tenenbaum et al., 2000). If the difference between two such high-dimensional points is accurately represented by the Euclidean distance between them (Kirby, 2001), linear dimensionality reduction methods, such as principal components analysis and singular value decomposition can be used. These methods transform the dataset along directions of maximum variability, thereby preserving relative configuration. If the points on the input data cannot be faithfully differentiated by the Euclidean distance, nonlinear methods such as isometric mapping (Isomap) (Tenenbaum et al., 2000) and local linear embedding (LLE) (Roweis and Saul, 2000; Saul and Roweis, 2003; Wang et al., 2004) may be used. As an example, the distance between two cities on earth is correctly represented along the great circle (geodesic) and not a straight line (Cox and Cox, 1991). Similarly, images of human faces or handwritten letters are not usefully separated by a linear difference of intensity values (Tenenbaum et al., 2000). In earlier work, we have used Isomap to show that dimensionality of the low-dimensional embedding created from images of self-propelled particles is indicative of the degree of coordination (Abaid et al., 2012) as well as the number of subgroups (DeLellis et al., 2013). In contrast to the frame-by-frame classification method that is developed in this paper, inferences from the dimensionality of the embedding manifold in Abaid et al. (2012) and DeLellis et al. (2013) are made on the basis of long sequences, comprising a few thousand frames.

The application of manifold learning for analysis and classification begins with the collection of training data for sampling the input space. The success of manifold learning in image-based analysis depends on several factors, including the variability and number of training images (Law and Jain, 2006) and image representation (Pless, 2003; Souvenir and Pless, 2007). In face recognition, for example, a large number of centered images of the subject are used to exhaustively sample the expected embedding space (Tenenbaum et al., 2000; Roweis and Saul, 2000). Similarly, for pose estimation, multiple viewpoints and extended videos of human activity are used as training data (Elgammal and Lee, 2007). Whereas humans can be requested to move in a specific manner, achieving a similar task with real animals is impractical. Consequently, individual frames are tagged by experts to quantify specific behaviors followed by classification on the basis of goodness of fits (Kopman et al., 2013).

We investigate the possibility of using generative models to create training images of collective behavior that can then be used to classify real videos. A generative model is a probabilistic relation that maps features to observations; for example, the probability of an image of an animal group given the number of subgroups. Once a generative model is available, it may also be inverted so that features can be extracted from observations (Silva and Tenenbaum, 2003; Mann et al., 2011). To create training images, we generate particle positions that emphasize variations of group features and project them on a two-dimensional image plane. In this data-driven approach, we do not propose a new model of collective behavior, rather a method to generate images of group configurations and trajectories that efficiently sample an underlying feature space. We use the Isomap algorithm to embed these images on a low-dimensional manifold (Fig. 1a). The Isomap algorithm approximates the embedding manifold in the higher dimensional space by first constructing geodesics between pairs of points. It then uses classical multidimensional scaling to construct a low-dimensional representation between points to generate an isometric embedding. That is, distances within the manifold infinitesimally respect the Euclidean distances (Cox and Cox, 2000). (More details on the Isomap algorithm are given in Section 3.) Next, we learn the manifolds by deriving two invertible mappings: one from the manifold to the input images and the other from the

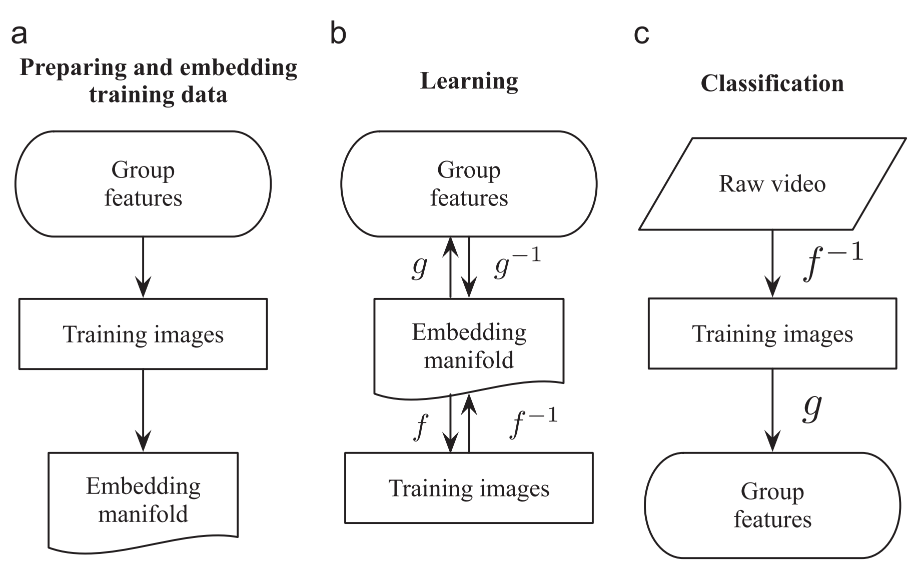

Fig. 1. (a) We use generative models to create synthetic training images that match a far-field view of animal groups. The training images are then compactly represented by the Isomap algorithm to a low-dimensional form called the embedding manifold. (b) We then derive invertible functions from the embedding manifold to the training images as well as to the group features. (c) The mappings are subsequently used to classify raw images of collective behavior.