Computer Vision and Image Understanding 117 (2013) 1–11

Contents lists available at SciVerse ScienceDirect

Computer Vision and Image Understanding

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

Modeling dynamic swarms ☆

Bernard Ghanem [a,*], Narendra Ahuja [b]

[a] King Abdullah University of Science and Technology, Geometric Modeling and Scientific Visualization Center, Saudi Arabia
[b] University of Illinois at Urbana-Champaign, Electrical and Computer Engineering Department, USA

ARTICLE INFO

Article history:
Received 28 October 2011
Accepted 13 September 2012
Available online 26 September 2012

Keywords:
Swarms
Dynamic textures
Crowd behavior analysis
Spatiotemporal analysis
Computer vision

ABSTRACT

This paper proposes the problem of modeling video sequences of dynamic swarms (DSs). We define a DS as a large layout of stochastically repetitive spatial configurations of dynamic objects (swarm elements) whose motions exhibit local spatiotemporal interdependency and stationarity, i.e., the motions are similar in any small spatiotemporal neighborhood. Examples of DS abound in nature, e.g., herds of animals and flocks of birds. To capture the local spatiotemporal properties of the DS, we present a probabilistic model that learns both the spatial layout of swarm elements (based on low-level image segmentation) and their joint dynamics that are modeled as linear transformations. To this end, a spatiotemporal neighborhood is associated with each swarm element, in which local stationarity is enforced both spatially and temporally. We assume that the prior on the swarm dynamics is distributed according to an MRF in both space and time. Embedding this model in a MAP framework, we iterate between learning the spatial layout of the swarm and its dynamics. We learn the swarm transformations using ICM, which iterates between estimating these transformations and updating their distribution in the spatiotemporal neighborhoods. We demonstrate the validity of our method by conducting experiments on real and synthetic video sequences. Real sequences of birds, geese, robot swarms, and pedestrians evaluate the applicability of our model to real world data.

1. Introduction

This paper is about modeling video sequences of a dense collection of moving objects which we will call swarms. Examples of dynamic swarms (DSs) in nature abound: a colony of ants, a herd of animals, people in a crowd, a flock of birds, a school of fish, a swarm of honeybees, trees in a storm, and snowfall. In artificial settings, dynamic swarms are illustrated by: fireworks, a caravan of vehicles, sailboats on a lake, and robot swarms. A DS is characterized by the following properties. (1) All swarm elements belong to the same category. This means that the appearances (i.e. geometric and photometric properties) of the elements are similar although not identical. For example, each element may be a sample from the same underlying probability density function (pdf) of appearance parameters. (2) The swarm elements occur in a dense spatial configuration. Thus, their spatial placement, although not regular, is statistically uniform, e.g., determined by a certain pdf. (3) Element motions are statistically similar. (4) The motions of the swarm elements are globally independent. In other words, the motions of two elements that are sufficiently well separated are independent. However, this is not strictly true on a local scale because if they are located too close compared to the extents of their displacements, then their motions must be interdependent to preserve separation. Thus, the motion parameters of each element vs. the other elements can be considered as being chosen from a mutually conditional pdf. Occasional variations in these swarm properties are also possible, e.g. elements may belong to multiple categories such as different types of vehicles in traffic. Fig. 1 shows some examples of DS.

This definition of DS is reminiscent of dynamic textures (DT). Indeed, a DS is analogous to a DT of complex nonpoint objects. The introduction of complex nonpoint objects introduces significant complexity: (1) Extraction of nonpoint objects becomes necessary, whose added complexity is evident from, e.g., the algorithm of [1]. (2) Motion for nonpoint objects is richer than point objects, e.g., rotation and nonrigid transformations become feasible. Since most work on DTs has focused on textures formed of pixel or subpixel objects, DS is a relatively unexplored problem. Tools for DS analysis should be useful for general problems such as dynamic scene recognition, dynamic scene synthesis, and anomaly detection, as well as, specific problems such as the motion analysis of animal herds or flocks of birds. In this paper, we present an approach to derive the model of a DS from its video, and demonstrate its efficacy through example applications. Before we do this, we first review the work most related to DS, namely, that on DT.

☆ This paper has been recommended for acceptance by Y. Aloimonos.

* Corresponding author.
E-mail addresses: bernard.ghanem@kaust.edu.sa (B. Ghanem), n-ahuja@illinois.edu (N. Ahuja).

1077-3142/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cviu.2012.09.002

Fig. 1. Examples of swarms.

1.1. Related work

A DT sequence captures a random spatiotemporal phenomenon which may be the result of a variety of physical processes, e.g., involving objects that are small (smoke particles) or large (snowflakes), or rigid (flag) or nonrigid (cloud, fire), moving in 2D or 3D, etc. Even though the overall global motion of a DT may be perceived by humans as being simple and coherent, the underlying local motion is governed by a complex stochastic model. Irrespective of the nature of the physical phenomena, the objective of DT modeling in computer vision and graphics is to capture the nondeterministic, spatial and temporal variation in images.

As discussed earlier, although the basic notion of DTs allows that both spatial and temporal variations be complex, the limited work done on DT's has focused on moving objects (texels) that have little spatial complexity, even as they exhibit complex motion. The texels are of negligible size (e.g. smoke particles), whose movement appears as a continuous photometric variation in the image, rather than as a sparser arrangement of finite (nonzero) size texels. Consequently, the DT model must mainly capture the motion and less is needed to represent the spatial structure.

Statistical modeling of spatiotemporal interdependence among DT images serves as being closest to the work we present here. This work includes the spatiotemporal auto-regressive (STAR) model by Szummer and Picard [2] and multi-resolution analysis (MRA) trees by Bar-Joseph et al. [3]. The DT model of Soatto et al. [4] uses a stable linear dynamical system (LDS). LDS mixture models have been developed in [5] and applied to DT clustering and segmentation. A bag-of-LDS model is proposed in [6] to account for view-invariance in DT recognition. Furthermore, the basic LDS model is extended to represent the incidence of multiple co-occurring DTs in the same video sequence, thus, leading to a layered LDS model for video [7,8]. In [9], a mixture of globally coordinated PPCA models is employed to model a DT. Moreover, a DT can be represented as a distribution of responses to spatiotemporal filters encoding oriented structures, which are shown to be discriminative of different DT classes [10]. Recently, the spatiotemporal variations in a DT has been described using dynamic fractal analysis, which in turn has shown great success in DT classification [11].

Along with their merits, the previously proposed models also suffer from certain shortcomings. (i) These models make restrictive assumptions about the DT sequences. Most of them assume that there is a single DT covering each frame in the sequence, while the others that consider multiple DT's are usually limited to particle textures (e.g. water and smoke). Consequently, these models cannot be easily extended to dynamic swarms. Even if the texels were known beforehand, learning a separate model for each texel does not guarantee the underlying spatiotemporal stationarity of DS. (ii) They do not make a clear separation between the appearance and dynamical models of the DT. The approach proposed in [12] explicitly aims at this separation, but it is limited to fluid DT's only.

Another body of work that is related to our swarm motion models a DT as a set of dynamic *textons* (or *motons*) whose motion is governed by a Markov chain model [13,14]. This generative model is limited to sequences of particle objects (e.g. snowflakes) or objects imaged at large distances. The texton dynamics are constrained by the underlying assumptions of the model, which state that all textons have the same frame-to-frame transformation, that this transformation is constant over time, and that the dynamics of spatially neighboring textons are independent. While this work does involve moving objects containing more than one pixel per object as well as some interpixel spacing, its modeling power still does not match the needs of properties (1–4) of a DS given above.

In the rest of this paper, we refer to the objects forming a swarm as swarm *elements*. We propose a probabilistic model that learns both the spatial layout of the swarm elements and their joint dynamics, modeled as linear transformations, which allow for a clear separation between the appearance and dynamics of these elements. This joint representation takes into account the interdependence in the properties of elements that are neighbors in space and time. This is done by enforcing stationarity only within spatiotemporal neighborhoods. This local stationarity constraint allows us to model DS sequences that not only exhibit globally uniform dynamics (to which previous methods are limited), but also sequences whose element properties and dynamics gradually change, in space and time.

1.2. Overview of proposed model

Given a DS sequence in which swarm elements undergo locally stationary transformations, we iterate between learning the spatial layout of these elements (their binary alpha mattes and frame-to-frame correspondences) and their dynamics. We estimate swarm dynamics such that they follow a probabilistic model that enforces