ANALYSIS OF COLLECTIVE BEHAVIOR AMONG SPECIES BY THE PRESENCE OF LOW-DIMENSIONAL EMBEDDING MANIFOLDS

Pietro DeLellis*
Department of Systems and Computer Engineering
University of Naples Federico II,
Via Claudio 21, Naples, 80125 Italy

Giovanni Polverino
Gozde Ustuner
Maurizio Porfiri†
Department of Mechanical and Aerospace Engineering
Polytechnic Institute of New York University,
6 MetroTech Center, Brooklyn, New York, 11201 USA

Nicole Abaid
Department of Engineering Science and Mechanics
Virginia Polytechnic Institute and State University,
Blacksburg, Virginia, 24061 USA

Erik M. Bollt
Department of Mathematics
Clarkson University,
Potsdam, NY, 13699 USA

ABSTRACT

Animal groups embody patterns that are recognizable to even untrained human observers. However, there is no universal definition of collective behavior within or between species, which makes comparison among different animal aggregations impossible. In this work, we use the ISOMAP, a machine-learning algorithm originally developed for machine vision, to quantitatively explore low dimensionality definitive of collective behavior in five different animal species over three experimental conditions: natural motion, attraction to a single source, and attraction to two sources. This algorithm embeds video data from each experiment onto a low-dimensional manifold and identifies the dimensionality of that manifold. We find that the method differentiates between the embedding manifold dimension for the five species and, within the three experiments for two of the species, it reports significantly different results depending on the experimental condition. This analytical method may enable identification of the underlying motivations of collective behavior, which are not completely understood in the biology community as of yet.

*Also at the Department of Mechanical and Aerospace Engineering, Polytechnic Institute of New York University, Brooklyn, NY 11201, USA.

†Address all correspondence to this author. Email: mporfiri@poly.edu

INTRODUCTION

Animals in almost every taxon exhibit collective behavior at some point in their life cycle [1]. These groups- including fish schools [2], bird flocks [3], insect swarms [4], and human crowds [5]- comprise a wide variety of interactions among their members which result in strikingly different behaviors. For example, a fish school may exhibit loose social shoaling while foraging for food, but highly aligned motion when responding to a predator [6]. While observable behaviors exhibited by animal groups are well studied in the biology literature, the mechanisms driving collective behavior are established only in part. Although drivers such as hydro- or aerodynamic advantage, increased ability to identify predators, and increased mating success may provide a foundation for collective behavior [7], the full range of benefits and detractors which result from living in a group has yet to be demonstrated.

Animal groups exhibiting collective behavior demonstrate patterns that are recognizable to even untrained human observers [8]. However, there is no previously-established universal definition of collective behavior within or between species, which makes comparison among different animal aggregations impossible. The motivation for this work is to enable the analysis of collective behavior by applying algorithms for pattern recognition, studied by the machine learning community, to data from animal groups.

Preprint submitted to 6th Annual Dynamic Systems and Control Conference.
Received March 7, 2013.

Within the last few decades, a large number of machine learning algorithms have been developed with the goal of extracting patterns from large and high-dimensional data sets [9]. These algorithms, including support vector machines [10], local linear embedding [11], hidden Markov models [12], and principal component analysis [13], find applications in an ample spectrum of science and engineering problems, from individual human recognition through biometric data [14] to identifying trends in climate and weather [15]. Many of these methods are based on embedding a data set on a manifold and then seeking patterns in the manifold, which correspond to patterns in the raw data set. Among these algorithms, the isometric mapping (ISOMAP), originally developed for machine vision [16], is unique in that it preserves geodesic distances in the raw data set and in the lower-dimensional coordinates it extracts. This property ensures that features of the low-dimensional manifold, on which the data set may be embedded, are present in the raw data itself. In its most fundamental interpretation, this property means that a high-dimensional data set which can be embedded on a low-dimensional manifold using the ISOMAP algorithm can be completely described by a small number of parameters.

In this work, we use ISOMAP to quantitatively explore collective behavior in five different animal species- namely humans, ants, frogs, chickens, and mosquitofish- in three experimental conditions: natural motion, attraction to a single source, and attraction to two sources. This algorithm has been previously used to define collective behavior as the presence of low-dimensional embedding manifolds in a large data set [17]. In this previous study by our group, such low-dimensional structures are identified in a simulated swarm exhibiting collective behavior. On the contrary, low-dimensional manifolds are absent when collective behavior is not observed. The same analysis is also conducted on video data from a fish school. In [18], group fragmentation in multi-agent systems is studied by analyzing the topological features of these low-dimensional manifolds. Here, using appropriately scaled physical and temporal dimensions, we compare the dimensionality of the embedding manifolds identified in each data set by the ISOMAP algorithm. We find that the species has a significant effect on the embedding manifold dimension when all experimental conditions are amalgamated, while the number of sources does not play a dominant role when all species are combined. However, when each species is evaluated independently, the number of sources have a significant effect in two cases. These results suggest that topological features of the ISOMAP embedding manifold offer viable tools to identify and classify collective behavior within and between social animal species.

MATERIALS AND METHODS

The experiment described in this work was approved by the Polytechnic Institute of New York University (NYU-Poly) Animal Welfare Oversight Committee AWOC-2012-102. Both the housing and the experimental procedure were designed to minimize stress in the animals.

Animals and apparatus

Mosquitofish (*Gambusia affinis*) and juvenile chickens (*Gallus domesticus*) were procured from an online aquarium source (LiveAquaria.com, Rhinelander, WI, 54501 USA) and an online poultry source (Meyer Hatchery, 626 Ohio, 89 Polk, OH, 44866 USA) respectively, while underwater frogs (*Xenopus laevis*) were obtained from a local aquarium store (Petland Discounts, Brooklyn, NY, 11201 USA). Humans (*Homo sapiens*) were recorded in proximity of NYU-Poly at Six MetroTech Center (Brooklyn, NY, 11201 USA) and ants (*Tetramorium caespitum*) in a public park in Brooklyn (Ave K and Ocean Pkwy, Brooklyn, NY, 11230 USA).

Captive populations of fish, frogs, and chickens were observed in controlled conditions. Fish and frogs were housed in the facility vivarium in the Department of Mechanical and Aerospace Engineering at NYU-Poly, while chickens were maintained in a private facility in proximity of NYU-Poly. Approximately 20 individuals for each species were acclimated for a minimum of 12 days prior to the experimental campaign. Fish and frogs were housed separately in holding tanks 50cm long, 25cm wide, and 30cm high, with a capacity of 36 liters each, while chickens were housed in an open-top cubic structure with side length 43cm. Water and air temperature were maintained constant at $26 \pm 2^{\circ}$C, and illumination was provided by diffused lights for ten hours each day in accordance with the circadian rhythm of these species. All populations were fed once a day after the conclusion of the daily experimental session. Fish were fed with commercial flake food (Hagen Corp., Nutrafin max, USA), frogs with frozen bloodworms, and chickens with commercial granulated food (Chick Starter-200lbs, P/U, Meyer Hatchery, USA.

For the three captive populations, the housing structures doubled as experimental apparatuses. For data acquisition of collective behavior in frogs and chickens, a digital video-camera (Canon, Vixia HG20, Japan) was suspended above test structure. For experiments with fish, the camera recorded a lateral view of the aquarium. In each case, the camera was placed at an appropriate distance to capture the entire domain accessible by the animals. The feeding sources for the fish were floating rings adhered to the aquarium side, into which floating flake food was placed as a stimulus. For frogs, the food source was negatively buoyant, thus not necessitating any physical constraint in the experimental apparatus. The food sources used for chickens were standard feed bowls.

Populations of humans and ants were observed in uncontrolled conditions and their collective behavior was recorded by suspending the camera to provide a bird's eye view of the unconstrained experimental area. The height of the camera was large enough to capture animals proximal to the two attractive sources, when they were present. For ants, the food source was placed on the ground close to the animals. For humans, the attractive sources were features already existing in the physical landscape: a breakfast kiosk and a metro station entrance. Figure 1 shows sample experimental video frames for each species.

Preprint submitted to 6th Annual Dynamic Systems and Control Conference.
Received March 7, 2013.

Experimental procedure

The selected animal taxa were chosen based on their availability, cost, and the possibility to be easily managed in a simple captive condition. Controlled laboratory experiments were favored for fish, frogs, and chickens to optimize managing experimental procedures.

The experiments were conducted as follows: the social behavior of each animal group was analyzed in three scenarios. Specifically, we analyzed the baseline of the social behavior for each animal group and their intra-specific interactions, as well when one or two contemporary sources of attraction were experimentally provided to them. The aim of the experiments was twofold:

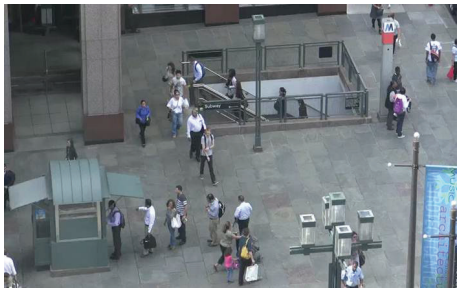
1. understanding the different behaviors across species and
2. investigating how the presence one or more sources of attraction distort the natural behavior in each species.

The sources of attraction were represented by sources of food in all the experiments, except for the case of humans, in which the sources of attraction were represented by the breakfast kiosk and by one of the entrances of the metro station in Jay Street, Brooklyn, NY, which are active hubs during the time when the data was acquired, see Fig. 1(a). The behavior of each animal group was studied and the experimental setup adapted accordingly to their average dimension and maximum speed. All the parameters were calculated before the start of the experiments to standardize the recordings. Specifically, each group had a different characteristic length and magnitude of velocity; therefore different frame samples and different recording time are selected for each species. The second parameter standardized was the distance between the feeding sources. This evaluation was based on the distance observed between the kiosk and the subway station (7m) used as attraction sources in the experiment with humans. The ratio between this distance and the mean body length was kept constant across the animal groups. The main experimental parameters are reported in Table 1.

Humans, ants, frogs, chicken, and fish were experimentally observed at the same time window, same perspective, and position everyday for ten consecutive days. Experimental videos were processed and transformed to snapshots (thirty frames per second) by using MTS converter and Avidemux programs.

Data analysis

As observed in [17], a formal definition of collective behavior is the existence of a low-dimensional stable invariant manifold in the space of the trajectories of the agents of the system. To detect the existence of such a manifold, we use the ISOMAP algorithm, which is a method for computing a quasi-isometric low-dimensional embedding of a set of high-dimensional data points [16]. Note that we apply the ISOMAP algorithm directly to video footage data sets, as low-dimensionality manifests itself even in the high-dimensional space of images in which collective dynamics is observed [17]. Now, we will first review the main



(a)

(b)

(c)

(d)

(e)

Figure 1. SNAPSHOT OF VIDEO DATA FROM EXPERIMENTS WITH (a) HUMANS, (b) ANTS, (c) FROGS, (d) CHICKEN, AND (e) FISH.

Preprint submitted to 6th Annual Dynamic Systems and Control Conference.
Received March 7, 2013.

steps of the algorithm and then describe its application on the collected video data.

ISOMAP algorithm Consider an array of n d-dimensional data points with the goal of embedding them on a possibly lower dimensional manifold. Namely, given a data set $\mathcal{Z} = \{z_i\}_{i=1}^n \subset \mathbb{R}^d$, we aim at building a corresponding data set $\mathcal{Y} = \{y_i\}_{i=1}^n \subset \mathbb{R}^{\bar{d}}$, embedded in an invariant manifold, and assessing if $\bar{d} \ll d$. The manifold can be represented through the following implicit parametrization

$$\mu : \mathcal{Y} \to \mathcal{Z}, \tag{1}$$

where

$$z_{ij} = \mu_j(y_{i_1}, \ldots, y_{i_{\bar{d}}}), \tag{2}$$

for $i = 1, \ldots, n$, and $j = 1, \ldots, d$. Here, the second subscript is used to identify vector components.

The ISOMAP algorithm is based on the classical multidimensional scaling method (MDS) [19], which is not applied to the ambient Euclidean space, but rather considers shortest paths along a discrete graph approximation of the manifold. The main steps of the algorithm can be summarized as follows:

1. **Construct a neighbors graph to approximate the embedding manifold**. We introduce the graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where the elements of the set of vertices $\mathcal{V} = \{v_i\}_{i=1}^n$ match the data points $\mathcal{Z} = \{z_i\}_{i=1}^n$, while the elements of the set of edges $\mathcal{E}$ are unordered pairs of vertices in the graph. We assign edges to connect vertices that are either ε-neighbors or ν-nearest neighbors. For instance, we can construct a ν-nearest neighbors graph, which consists of edges $\{v_i, v_j\}$ corresponding to the ν-closest data points z_j to z_i, for each $i = 1, \ldots, n$, with respect to the Euclidean distance in the ambient space, denoted by $d_Z(z_i, z_j)$. We define $M_n \in \mathbb{R}^{n \times n}$ as the matrix encoding the weighted graph of intrinsic manifold distances corresponding to the graph $\mathcal{G}$, whose ij-th entry is denoted by $M_n(i, j)$. For each edge $\{v_i, v_j\} \in \mathcal{E}$, we define the distances $M_n(i, j) \approx d_Z(z_i, z_j)$ and, for all $\{v_i, v_j\} \notin \mathcal{E}$, we set $M_n(i, j) = \infty$ to forbid jumps between branches of the underlying manifold.
2. **Compute the graph geodesic matrix to approximate the geodesic of the manifold**. This step can be performed using well-established methods to compute shortest paths, such as Floyd's algorithm [20] or Dijkstra's algorithms [21]. From M_n, we compute an approximate geodesic distance matrix $D_M \in \mathbb{R}^{n \times n}$, whose ij-th entry is the shortest weighted path length between each v_i to v_j, being an approximation of manifold geodesic distances.
3. **Approximate manifold distance by ν-nearest neighbor distance**. The distance matrix D_M from the previous step is used to approximate the geodesic distances of the manifold between z_i and z_j by the distance between v_i and v_j. The accuracy of the approximation increases with data density. If ν is too large or data density is too low, then some neighbors might be on separate manifold branches, resulting in a poor representation of the manifold.
4. **Perform an MDS on D_M**. The input to the MDS is the matrix D_M computed in step 2) from the input data $\mathcal{Z}$. The outputs are the projective variables $\mathcal{Y}$ in the intrinsic variables.

The outputs of the ISOMAP algorithm are the transformed data points on an embedding manifold for the input data set $\mathcal{Z}$ and the vector R of residual variances, which, in turn, quantifies the proportion of data points not lying on such manifold. From the norm of the residual variances, we determine the dimensionality of the embedding manifold that well approximates $\mathcal{Z}$. Specifically, to identify the dimensionality, we say that it corresponds to the minimum value $\bar{d}$ such that $R(\bar{d})$ is less than 0.05.

Video analysis Before starting the analysis, we need to appropriately sample the set of video frames, so that in each experiment the differences between one pixel and the next one is comparable. To this aim, the sampling period s has to be inversely proportional to the speed v_{pix} of each species on the screen. As reference, we consider the experiments in which the speed is higher, that is, the fish experiment with feeding sources, and we set the sampling period to 1 frame. All the other sampling periods are taken accordingly and are reported in Table 1. After extracting the video frames, the next step is the application of the ISOMAP algorithm to 960×640 pixel values in the range 1 to 256 for the gray scale image at each of the sampled frames, using $\nu = 11$.

	Fish	Humans	Ants	Frogs	Chickens
T	30	600	390	480	480
s	1	20	13	16	16
δ	16	700	1	16	35

Table 1. EXPERIMENTAL PARAMETERS: LENGTH T OF THE VIDEOS (SECONDS), SAMPLING PERIOD, AND DISTANCE δ BETWEEN THE TWO SOURCES (CENTIMETERS). IN THE FISH EXPERIMENTS, v_{pix} IS REDUCED BY A FACTOR OF TWENTY FOR CONDITIONS WHERE SOURCES ARE ABSENT. IN THAT CASE, WE SET $T = 600$ AND $s = 20$.

Statistics

For each species, we acquire ten replicate trials in each of the three conditions: zero, one, and two sources. These trials are

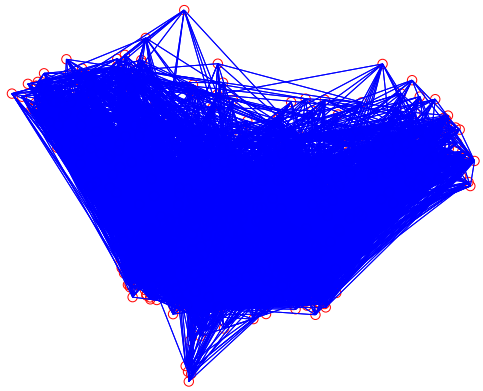


Figure 2. TWO-DIMENSIONAL EMBEDDING MANIFOLD GENERATED BY THE ISOMAP ALGORITHM FOR A REPRESENTATIVE TRIAL FROM FISH EXPERIMENTS WITHOUT SOURCES OF ATTRACTION.

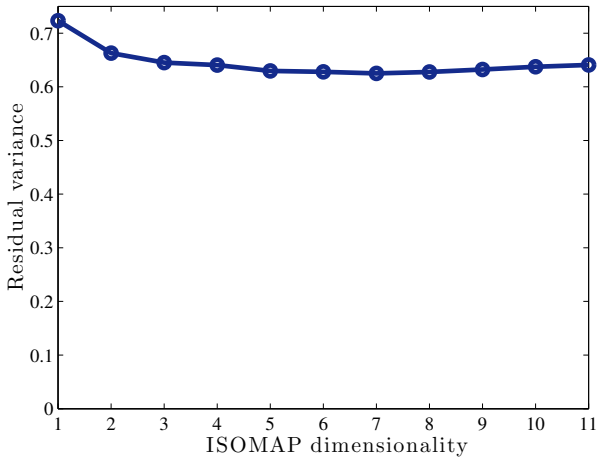


Figure 3. RESIDUAL VARIANCE FOR THE ISOMAP ALGORITHM PERFORMED ON A REPRESENTATIVE TRIAL FROM FISH EXPERIMENTS WITHOUT SOURCES OF ATTRACTION.

analyzed individually with a modified code from [22] to yield an isomap dimensionality. To analyze the effect of species and experimental condition on ISOMAP dimensionality, we perform a two-way analysis of variance (ANOVA) [23] with condition and species as independent variables and the ISOMAP dimensionality as dependent variable. A one-way ANOVA is also performed on each species' ISOMAP dimensionalities independently with condition as independent variable. For both ANOVAs, post-hoc analysis is achieved using Fisher's protected least squares differences (PLSD) [24] when a significant main effect is observed.

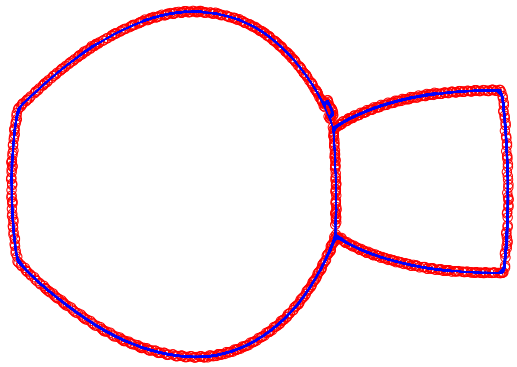


Figure 4. TWO-DIMENSIONAL EMBEDDING MANIFOLD GENERATED BY THE ISOMAP ALGORITHM FOR A REPRESENTATIVE TRIAL FROM FISH EXPERIMENTS WITH ONE SOURCE OF ATTRACTION.

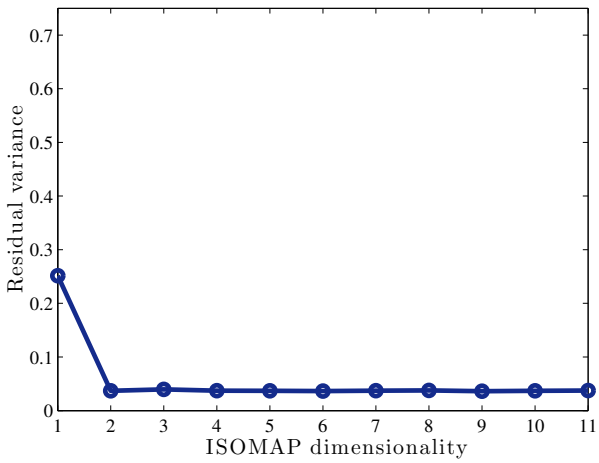


Figure 5. RESIDUAL VARIANCE FOR THE ISOMAP ALGORITHM PERFORMED ON A REPRESENTATIVE TRIAL FROM FISH EXPERIMENTS WITH ONE SOURCE OF ATTRACTION.

RESULTS

To provide an insight on the outputs of the ISOMAP algorithm, we focus on two representative trials selected from the fish experiments, with zero and one source of attraction, respectively. In Figs. 2-5, we report the two-dimensional embedding manifold and the residual variance plot for the selected trials. Figure 2 clearly illustrates that the dimensionality $\bar{d}$ is higher than 2, while from Fig. 3, we observe that $R(d) > 0.05$ for all $d = 1, \ldots, 11$, and therefore $\bar{d} > 11$. When a source is added, then the manifold is clearly two-dimensional, as illustrated in Figs. 4 and 5:

Preprint submitted to 6th Annual Dynamic Systems and Control Conference.
Received March 7, 2013.

dimension 2 is clearly detected, as $R(2) = 0.04$.

When we amalgamate all trials from each species, we find that species has a significant main effect on the dimensionality of the ISOMAP manifold (Two-way ANOVA, $F_{(4,133)} = 21.7$, $p < 0.01$). Results are reported in Fig. 6, including pairwise post-hoc tests. The ISOMAP dimensionality from the only insect species, ants, is significantly different from all other species. Two of the remaining species, frogs and fish, are significantly different from all but one other species and two, humans and chickens, are significantly different from all but two other species.

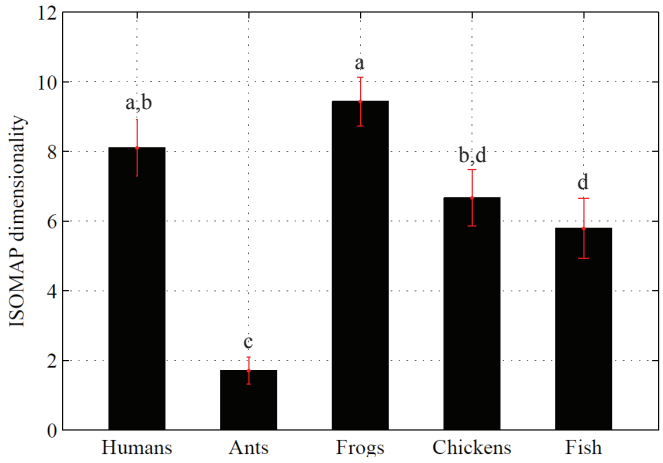


Figure 6. MEAN ISOMAP DIMENSIONALITY FOR THIRTY TRIALS OF EACH SPECIES, THAT IS, WITH ALL THREE EXPERIMENTAL CONDITIONS COMBINED. ERROR BARS SHOW ONE STANDARD ERROR. MEANS NOT SHARING A COMMON SUPERSCRIPT ARE SIGNIFICANTLY DIFFERENT IN POST-HOC TESTS (FISHER'S PLSD, $p < 0.05$).

Combining all the species for each experimental condition, the effect of condition fails to reach statistical significance (Two-way ANOVA, $F_{(2,133)} = 2.3$, $p = 0.11$), see Fig. 7. Nevertheless, the p-value from pairwise post-hoc tests show higher significance comparing zero and a non-zero source conditions. In particular, we find $p = 0.07$ comparing zero and one source and $p = 0.25$ comparing zero and two sources.

To delve further into the behavior of each species, we consider a one-way ANOVA with condition as independent variable for each species, see Fig. 8. We find no significant main effect of condition for humans, ants, and frogs (One-way ANOVA, $F_{(2,27)} = 2.1$, $p = 0.14$ for humans; $F_{(2,27)} = 0.7$, $p = 0.51$ for ants; $F_{(2,27)} = 1.2$, $p = 0.31$ for frogs). However, condition significantly influences ISOMAP dimensionality for chickens and fish (One-way ANOVA, $F_{(2,27)} = 3.6$, $p < 0.05$ for chickens; $F_{(2,25)} = 44.3$, $p < 0.01$ for fish). In both cases, we find more significant differences when zero and non-zero source conditions

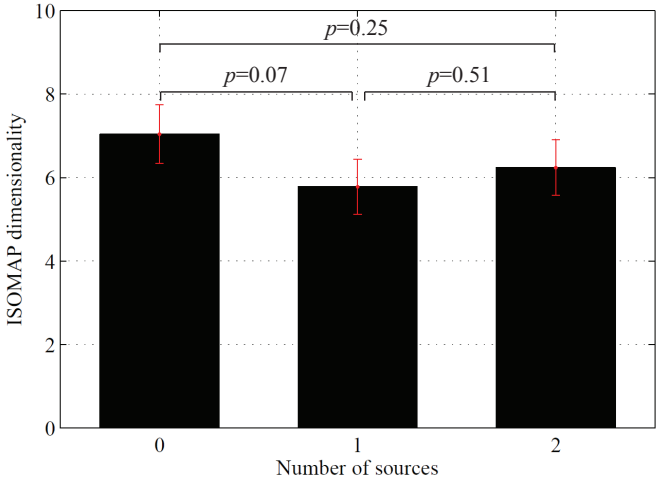


Figure 7. MEAN ISOMAP DIMENSIONALITY FOR FIFTY TRIALS OF EACH CONDITION, THAT IS, WITH ALL FIVE SPECIES COMBINED. ERROR BARS SHOW ONE STANDARD ERROR. SIGNIFICANCE FROM POST-HOC TESTS ARE INDICATED (FISHER'S PLSD).

are compared, rather than comparing the one and two source conditions.

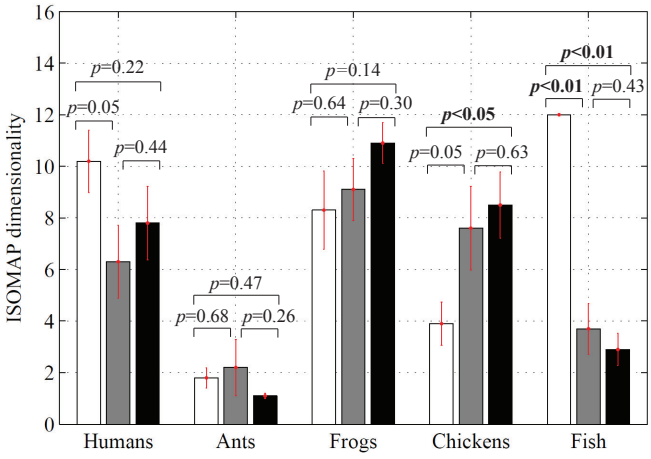


Figure 8. MEAN ISOMAP DIMENSIONALITY FOR TEN TRIALS OF EACH CONDITION FOR EACH OF THE FIVE SPECIES. ERROR BARS SHOW ONE STANDARD ERROR. SIGNIFICANCE FROM POST-HOC TESTS ARE INDICATED (FISHER'S PLSD) AND SIGNIFICANT DIFFERENCES ARE BOLDED.

CONCLUSIONS

In this paper, we analyzed collective behavior in animal groups. Specifically, we collected video data of five different

species over three different experimental conditions: natural motion, attraction to a single source, and attraction to two sources. For each species and experimental condition, video data were collected at the same time window for ten consecutive days. According to [17], we defined collective behavior as the existence of a low-dimensional stable invariant manifold in the space of the trajectories of the agents of the system. The ISOMAP algorithm, originally developed for machine vision, was employed to embed the collected video data on low-dimensional manifolds and thus detect the presence of low-dimensionality in terms of small residual error of fit. The dimensionality of the manifold that well approximates the data was detected by analyzing the norm of the residual variances. The results of the ISOMAP analysis were then amalgamated over all trials from each species, and a statistical analysis was performed to test how significant are the effects of species and experimental conditions.

We found that the algorithm differentiates between the embedding manifold dimension for the five species and, within the three experiments for two of the species, reports significantly different results depending on experimental condition. Stimulated by the results presented in this paper, current work is devoted to the formalization of a method for automated classification of collective behavior in animal groups from row video data, that may provide valuable insight to identify the underlying of collective motion, and to characterize similarities and peculiarities across the species.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation under Grant No. CMMI-1129820 and CAREER Grant No. CMMI-0745753. Additional support has been provided in part by the Honors Center of Italian Universities (H2CU) through a scholarship to G. Polverino. At Clarkson University, this work was also supported by the National Science Foundation under Grant No. CMMI-1129859.

REFERENCES

[1] Krause, J., and Ruxton, G. D., 2002. *Living in Groups*. Oxford University Press, Oxford, U.K.

[2] Aoki, I., 1980. "An analysis of the schooling behaviour of fish: Internal organization and communication process". *Bulletin of the Ocean Research Institute of the University of Tokyo,* **12**, pp. 1–65.

[3] Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., Procaccini, A., Viale, M., and Zdravkovic, V., 2008. "Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study.". *Proceedings of the National Academy of Sciences,* **105**(4), pp. 1232–1237.

[4] Yates, C. A., Erban, R., Escudero, C., Couzin, I. D., Buhl, J., Kevrekidis, I. G., Maini, P. K., and Sumpter, D. J. T., 2009. "Inherent noise can facilitate coherence in collective swarm motion". *Proceedings of the National Academy of Sciences,* **106**(14), pp. 5464–5469.

[5] Ma, J., Song, W.-G., Zhang, J., Lo, S.-M., and Liao, G.-X., 2010. "k-nearest-neighbor interaction induced self-organized pedestrian counter flow". *Physica A: Statistical Mechanics and its Applications,* **389**(10), pp. 2101–2117.

[6] Partridge, B. L., 1982. "Structure and function of fish schools". *Scientific American,* **246**(6), pp. 114–123.

[7] Rubenstein, D., and Kealey, J., 2012. "Cooperation, conflict, and the evolution of complex animal societies". *Nature Education Knowledge,* **3**(10), p. 78.

[8] Project PigeonWatch, [Online]. http://www.birds.cornell.edu/ pigeonwatch.

[9] Gong, Y., and Xu, W., 2007. *Machine Learning for Multimedia Content Analysis*. Springer-Verlag, New York, NY.

[10] Tong, S., and Koller, D., 2002. "Support vector machine active learning with applications to text classification". *Journal of Machine Learning Research,* **2**, pp. 45–66.

[11] Roweis, S., and Saul, L. K., 2000. "Nonlinear dimensionality reduction by locally linear embedding". *Science,* **290**(5500), pp. 2323–2326.

[12] Rabiner, L. R., 1989. "A tutorial on hidden markov models and selected applications in speech recognition". *Proceedings of the IEEE,* **77**(2), pp. 257–286.

[13] Jolliffe, I., 2002. *Principal Component Analysis*. Springer-Verlag.

[14] Damousis, I. G., and Argyropoulos, S., 2012. "Four machine learning algorithms for biometrics fusion: A comparative study". *Applied Computational Intelligence and Soft Computing,* **2012**, p. 242401.

[15] Hsieh, W. W., 2008. *Machine Learning Methods in the Environmental Sciences*. Cambridge University Press.

[16] Tenenbaum, J. B., de Silva, V., and Langford, J. C., 2000. "A global geometric framework for nonlinear dimensionality reduction". *Science,* **290**(5500), pp. 2319–2323.

[17] Abaid, N., Bollt, E., and Porfiri, M., 2012. "Topological analysis of complexity in multiagent systems". *Physical Review E,* **85**(4), p. 041907.

[18] DeLellis, P., Porfiri, M., and Bollt, E. M., 2013. "Topological analysis of group fragmentation in multi-agent systems". *Physical Review E,* **to appear**.

[19] Cox, T. F., and Cox, M. A., 1994. *Multidimensional scaling*. Chapman and Hall, London.

[20] Floyd, R. W., 1962. "Algorithm 97: Shortest path". *Communications of the ACM,* **48**(6), p. 345.

[21] Dijkstra, E. W., 1959. "A note on two problems in connexion with graphs". *Numerische Mathematik,* **1**(1), pp. 269–271.

[22] A Global Geometric Framework for Nonlinear Dimensionality Reduction. http://web.mit.edu/cocosci/isomap/isomap.html.

[23] Mickey, R. M., Dunn, O. J., and Clark, V. A., 2004. *Applied Statistics: Analysis of Variance and Regression*. John Wiley

& Sons, Hoboken, NJ.

[24] Echevarria, D. J., Hammack, C. M., Pratt, D. W., and Hosemann, J. D., 2008. "A novel behavioral test battery to assess global drug effects using the zebrafish". *International Journal of Comparative Psychology,* **21**(1), pp. 19–34.