EigenPhenotypes: Towards an Algorithmic Framework for Phenotype Discovery

Alexander Vaughan[1], Rahul Singh[2], Alan Shimoide[2], Ilmi Yoon[2], Megumi Fuse[1]
[1] Department of Biology, [2] Department of Computer Science
San Francisco State University, 1600 Holloway Avenue, San Francisco, CA 94132

Abstract

Studying the genetic control of molecular, anatomical and/or morphological phenotypes in model organisms is a powerful tool in the functional analysis of a gene. The goal of our research is to develop algorithms that discover phenotypes of behavior in model organisms, which may identify, categorize, and quantify these phenotypes under conditions of minimal a priori information. Starting from a non-invasive video monitoring of a model organism, we propose an eigen-decomposition of the organism's behavior captured in video. Traditional clustering techniques in space, time, and frequency can utilize this decomposition to characterize the categorical behaviors of an animal, and for an analysis of the behavioral repertoire. This supplies a quantified analysis of behavior with minimal assumptions, a crucial first step in the genetic analysis of behavior.

1. Introduction

The study of behavior is difficult, and particularly so in its methodology. Behavioral psychologists, ecologists, and neuroscientists have worked for many years to develop analyses that provide powerful, unbiased insight into the behavior of an animal - yet the relationship between genetic and phenotypic patterns in behavior remains obscure. We currently lack a general method for the identification of behavioral phenotypes and, more fundamentally, lack a means to quantify behaviors and behavioral repertoires too complex to be discerned by a human observer.

Traditional analyses of behavior have typically taken two approaches. Categorical analyses such as *ethograms* have first identified behaviors by category, working from initial observations. Quantitative records are made using these original definitions as a (relatively) impartial signpost, a formulation that minimizes post-hoc and ad-hoc errors of phenotype and pattern. In contrast, behaviorist psychologists have demanded novel behaviors of their subjects, in an effort to demonstrate the universality of underlying principles.

Our method proposes a third class of analysis alongside the categorical and behaviorist approaches, by the application of sensor technology, multivariate analysis and time-series methods [2]. If a video record of an animal's behavior is subject to a Singular-Value Decomposition (SVD), the resulting singular vectors capture the variance in its appearance over time. Moreover, they constitute a subspace of that appearance, and the organism's behavior defines its trajectory in that subspace. Behaviors may be identified either as dense loci or repetitive motifs in this trajectory, and behavioral phenotypes are identified by the pattern of specific behaviors. Because of their origin in the decomposition, we term these *EigenPhenotypes*.

As a proof-of-principle study, we present experiments showing that the molting of the moth *Manduca sexta* (an established phenotype) can be discovered without foreknowledge of the form or organization of that behavior.

In the molting behavior of the moth caterpillar *Manduca sexta*, a cascade of both steroid and peptide hormones coordinates the timing of cuticle loosening and shedding. Three behaviors have been described previously: two, termed pre-ecdysis I and II, appear to loosen the old cuticle before shedding; a third, called *ecdysis*, actually sheds the cuticle [3]. It is known that pre-ecdysis I and II overlap, but their interaction and modulation is unknown. Furthermore, he transition from pre-ecdysis to ecdysis is irreversible, but the control of its timing remains obscure. This ignorance stands despite the importance of this behavior: failure to shed the cuticle at ecdysis inevitably results in death.

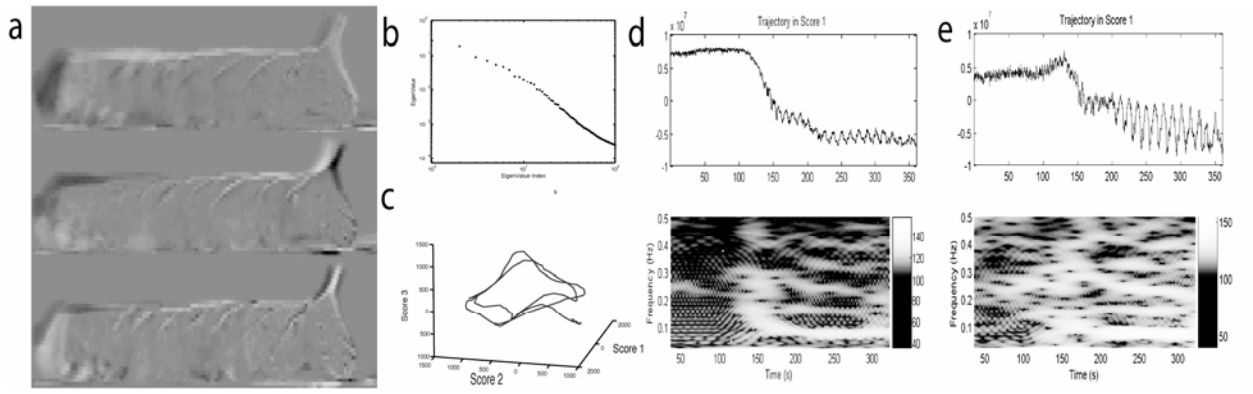

Figure 1. The decomposition of the behaviors of caterpillar ecdysis shows oscillatory behaviors. (a) The principal singular vectors for a short (~15s) video; major postural changes are dominant in the first several vectors, and detail emerges with the smaller singular values. (b) Corresponding singular values. (c) The trajectory of this video in the first three singular vectors; the oscillation reflects the underlying contractile movements of the animal. (d) The trajectory through the first singular vector for a 6 minute video shown directly and as a periodogram; the transition from pre-ecdysis to ecdysis is visible at ~100s. e) The same video, projected onto the subspace derived from the video of another animal's ecdysis, shows similar oscillations that reflect the underlying behavior.

2. Methods

We begin with a video of the behaviors, from which we segregate the animal from its background using a minimum bounding box. We then create a matrix *M* of dimensions $(f, m \times n)$, where f is the number of frames in the video and $m \times n$ denote the number of pixels. Our next step involves analyzing the video using the SVD. Unfortunately, the large data size precludes a direct application of the SVD technique. Instead, we utilize Monte-Carlo methods to reduce the computational load, while retaining an estimate of the resulting error [1]. The salient steps in the algorithm can be described as follows:

1. Sample s centered rows of M from a uniform distribution without replacement, and include each as a row of N
2. Compute $N*N^T$, and its k-rank SVD:

$$NN^T = \sum_{t=1}^{k} \lambda_t^2 w_t w_t^T$$

3. Compute $h_t = (N^T * w_t) / \left| N^T * w_t^T \right|$
4. Return ***H***, whose columns are the right singular vectors h_t; ***W***, whose columns are the $\boldsymbol{w}_t$, and $\sum$, a diagonal matrix containing the singular values $\lambda_1, \ldots, \lambda_k$.

This formulation decomposes ***N*** as ***N=W*∑*H*** in time $O(m \times n)$. The scores of all frames can be found as ***X = MH*∑,** each row of which is a trajectory scored against one right singular vector over time.

3. Discussion

Several points merit further elaboration. First, that the singular vectors do not parameterize the behaviors *per se*, but instead the animal's posture. However, the motifs that recur in the trajectory time series signify repetitions of a single behavior. Alternately, other vector spaces may provide better encoding for specific behavioral patterns. Periodic behaviors, for example, may be better decomposed in the frequency domain. Ultimately, the application of an appropriate decomposition and corresponding time-series analysis tools opens behavior to novel analysis, for which several tools will prove useful.

4. References

[1] P. Drineas, R. Kannan, and M.W. Mahoney, "Fast Monte Carlo algorithms for matrices III". Technical Report YALEU/DCS/TR-1271, Yale University, New Haven, 2004.

[2] S. K. Nayar, S. A. Nene, and H. Murase, "Subspace Methods for Robot Vision", TR CUCS-06-95 CS, Columbia, 1995.

[3] A. Novicki and J. Weeks, "The Initiation of Pre-Ecdysis and Ecdysis Behaviors in larval *Manduca Sexta*: The Roles of the Brain, Terminal Ganglion and Eclosion Hormone. J. Exp. Bio. **199**, 1757-1769 (1996).