

Paper:

Behavior Learning and Animation Synthesis of Falling Flat Objects

Kohta Aoki*, Osamu Hasegawa**,***, and Hiroshi Nagahashi**

*Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

**Imaging Science and Engineering Laboratory, Tokyo Institute of Technology
4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8503, Japan

***PRESTO, Japan Science and Technology Corp. (JST)

E-mail: {aoki, hasegawa, longb}@isl.titech.ac.jp

[Received August 11, 2003; accepted December 1, 2003]

In this paper, we describe an approach to learning patterns from sample data sequences and generating new data sequences through learned models. The target application of this work is the animation of natural phenomena, especially falling behavior of flat objects. The natural object or phenomenon to be animated is recorded using one camera, and its characteristic behavior is captured. Feature vectors are defined as the representation of behavior and are automatically extracted from captured videos. By learning the structure of a set of sample vector sequences, the learned model can generate a novel pattern through the underlying structure. These generated patterns could differ from every original vector sequence but preserve characteristics of subject behavior. We can use such patterns to synthesize natural-looking animation.

Keywords: pattern learning, computer animation, natural phenomena, self-organizing map

1. Introduction

The animation of natural phenomena is a challenging task in computer graphics, but a very useful one in many commercial and academic applications. Every natural phenomenon has some characteristic behavior. If such behavior is captured, it can be efficiently used in the animation of natural phenomena.

We propose an approach to the creation of image-based animation of natural objects or phenomena by learning patterns from various sample data that characterizes their behavior or motions. This paper seeks to animate a falling leaf or petal. Several researchers have reported physics- or dynamics-based simulations for the falling behavior of a flat object [?, ?]. Parameter optimization for these simulations and models is generally complicated. We focus on the synthesis of natural-looking animation rather than on the acquisition of such models.

One approach to acquiring motion data is a motion capture system, which tracks markers attached to an object and produces three-dimensional representation of its

movements. Many techniques have been proposed to synthesize novel motions and animation by editing obtained motion data [?, ?, ?]. However, motion capture is seldom applied to natural phenomena, because attached markers disturb natural motion, such as the fluttering of a butterfly's wings.

Our approach uses only behavior data of subject that can be obtained from video recordings. Some leaf- or petal-shaped image objects, or *sprites*, are composed onto appropriate background images through such sample data. Other sample-based techniques for animation synthesis have been developed, for example, speech animation by blending separated facial parts in sample images [?], and an approach to synthesizing various streams of images by rearranging original frames from source videos [?]. Ezzat *et al.* [?] presented a technique for synthesizing *videorealistic* speech animation, however there are some constraints on composing an image sequence of mouth movements into background images. Character animation by controlling sprites [?] can create effects similar to our approach. This approach is an extension to frame arrangement [?] and therefore should prepare many sprite frames for various animation, while we can synthesize much animation using behavior data, which is more easily obtained.

Machine learning methods that create animation and motions are described in [?, ?], where physics-based models are driven by learned controllers. In our approach, learning involves finding the structure of sample data. The learned model can generate a novel pattern that preserves the underlying structure. These processes of our approach correspond to two types of inference: *induction* and *deduction*. That is, induction is to learn the underlying structure of a behavior data set, and deduction is to generate various patterns through such structure. We exploit the self-organizing map (SOM) learning algorithm [?], which is a neural network model. A one-dimensional map may be regarded as a vector sequence or a pattern. In other words, we associate pattern generation with learning by the SOM algorithm. Natural animation can be created by moving sprites according to generated patterns that represent subject behavior.

Our system consists of three phases, as shown in **Fig.1**.

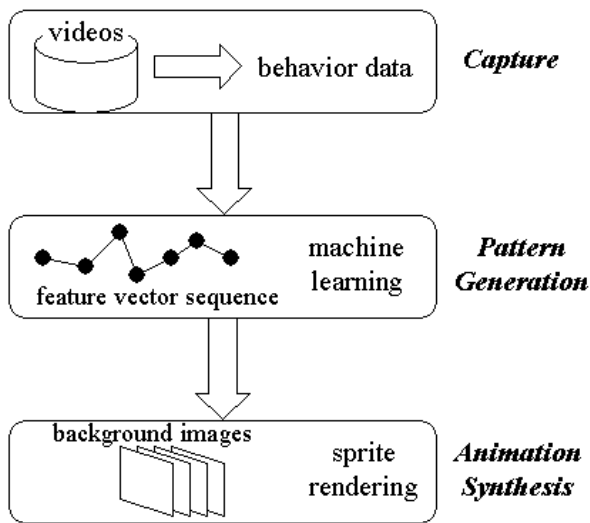


Fig. 1. System overview.

We first capture with one camera various scenes where a leaf is falling, and then track it in these videos. The characteristic behavior of a falling leaf is defined as a temporal sequence of feature vectors (Section 2). An SOM network partly approximates a set of sample vectors in the learning process. The learned network may be equivalent to a novel pattern (Section 3). We use these generated patterns to animate sprites of leaves (Section 4), and photorealistic videos that depict the subject are composed (Section 5).

2. Feature Extraction

First, we must record the subject as accurately as possible, so we used a high-speed camera capable of capturing 125 frames per second, and eventually captured 70 video streams. Each stream is composed of 50-90 frames with a resolution of 480×420 pixels. A target object, i.e., a leaf, is tracked in these videos. Then, the object region is simply segmented from each frame by background subtraction, because the recording was done in a monotone environment. These video recordings are only used to extract some features of subject behavior, so monochrome videos may be more convenient than color ones.

For a segmented object region in each frame, features are calculated to represent characteristic behavior of the subject. Through preliminary experiments, we determine the following appearance-based features:

- angle θ of the central principal axis,
- height L_1 and width L_2 of the minimum rectangle that includes the object region.

Motion-based features are defined as follows:

- dx and dy of movement per pixel between two consecutive frames.

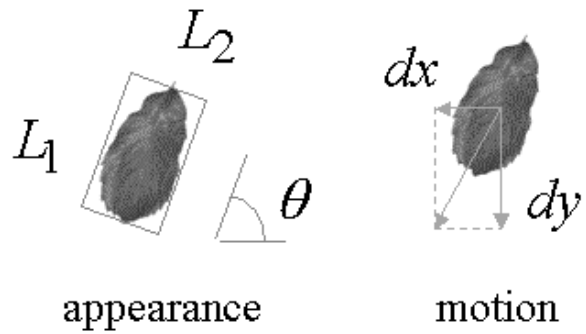


Fig. 2. Features expressing characteristic behavior of a falling leaf.

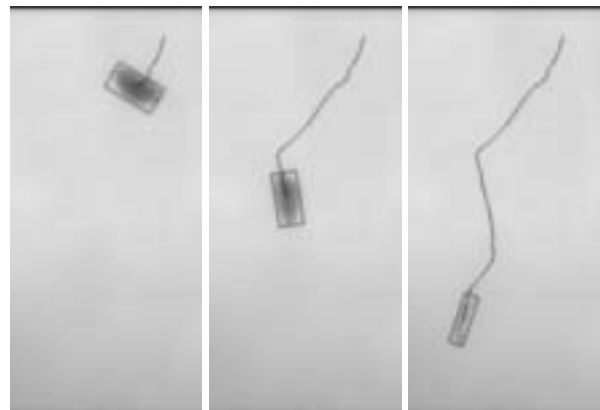


Fig. 3. Automatic tracking and extraction of a falling leaf from a given video stream captured by a high-speed camera.

These five features are simple but fundamental and effective at representing subject behavior, as shown in Fig.2. A sequence of feature vectors is automatically calculated from one video stream as follows:

$$[\theta(t) L_1(t) L_2(t) dx(t) dy(t)], \quad (t = 1, \dots, T) \quad (1)$$

where T is the length of a sequence.

Thus, our approach characterizes the falling behavior of a leaf using only the data obtained from two-dimensional image sequences. The obtained data, or feature vector sequences, is a primitive representation, but we can create sprite animation of a falling leaf. Figure 3 shows an example of automatic tracking and extraction of a falling leaf from the given video stream captured by the high-speed camera.

3. Learning and Pattern Generation

One common technique for learning sequential patterns is a simple recurrent network or *Elman Net* [?]. The network can develop internal representations for some input patterns, and produce the correct output for a given input.