

MOTION-SWARM WIDGETS FOR VIDEO INTERACTION

Jeffrey E. Boyd

Department of Computer Science
University of Calgary
Calgary, AB, Canada T2N 1N4

ABSTRACT

Computer vision systems for human-computer interaction have tended towards more precise forms of interface that require complex vision tasks such as segmentation, tracking, object recognition, pose estimation, and gesture recognition. We present an alternate approach that extrapolates a method for en masse audience interaction through video. The en masse interaction simulates a particle moving in the field of motion created by the audience, and the audience interacts by manipulating the particle position. In this paper, we show that by adding sets of constraints to the particle motion, one can build GUI-style widgets. We describe several of these widgets and the results of a small-sample pilot study to test them. The results are not conclusive, but are encouraging, suggesting possibilities for video games and interactive theatre.

1. INTRODUCTION

As the field of computer vision matures and processor speeds increase, the drive for vision-based human-computer interaction (HCI) proceeds apace [1]. Surveys of vision-based HCI [2, 3] reveal a common theme: the goal of computer vision in the interface is to provide model of the world that is sufficient to interpret the actions of people as instructions to an interface. Typically this employs a plethora of techniques that have emerged from computer vision such as segmentation, object finding (e.g., face finding [4]), tracking, and gesture recognition. The common requirement is that the vision system be able identify and track a person, a body part, or a prop, and interpret either static states or the motion as a whole. In the complex environments in which people usually operate, this can be a complex task. Even model-free methods (e.g., motion-energy and motion-history images [5]) require that a computer vision system identify a set of pixels that belongs to a single physical object or person.

However, some environments make it difficult or impossible for a computer vision system to identify individual physical entities. For example, consider the scenario where an audience at a public event interacts with a large video display via a video camera. The camera can view the audience, but the prospect of identifying individuals in the image, let alone

tracking them, measuring their pose, or recognizing their gestures is daunting. In a project to develop interactive art for view by a large audience, we encountered exactly this situation, and developed a method for interaction we call *motion swarms* to provide a solution [6]. A motion swarm system measures a *motion field* for the scene viewed by a camera (be it a body part, a person, or hundreds of people). The system then simulates the motion of physical particles that react to the motion field. When both the particle position and the audience's image are mirrored in the display, the audience can use their motion to control the position of the particle as the basis for a form of interaction. The name *motion swarm* stems from both the swarm like appearance that emerges from the simulation of several particles reacting to motion, and the fact that we used motion swarms to interact with *Swarm Art* [7].

Motion swarm interaction with audiences works for several reasons. First, the measurement of the motion field need not be precise; it is enough to know where motion is and is not occurring in the image. Second, the mirroring of particle positions and the audience in the display provides a common coordinate system for the interaction without camera calibration. Finally, interactive art is not a demanding application; the interaction can be imprecise because the goal is entertainment and aesthetic value.

This raises a question: can we implement a system that has the desirable properties of motion swarms, but facilitates more precise interaction? This paper answers that question affirmatively. In order for motion swarms to work for interaction, it is essential to constrain the motion of the particles. Careful use of these constraints allows us to build video interaction widgets in the style of common graphical user interfaces (GUI), thereby adding more precision to the interaction. Preliminary results from a pilot study presented here show that interaction is possible with this method.

2. MOTION SWARMS AND AUDIENCE INTERACTION

Figure 1 describes schematically how motion swarms operate [6]. A field of motion (there are several options here, but we use motion history images [5]) acquired by a video camera exerts simulated forces on simulated particle motion. When

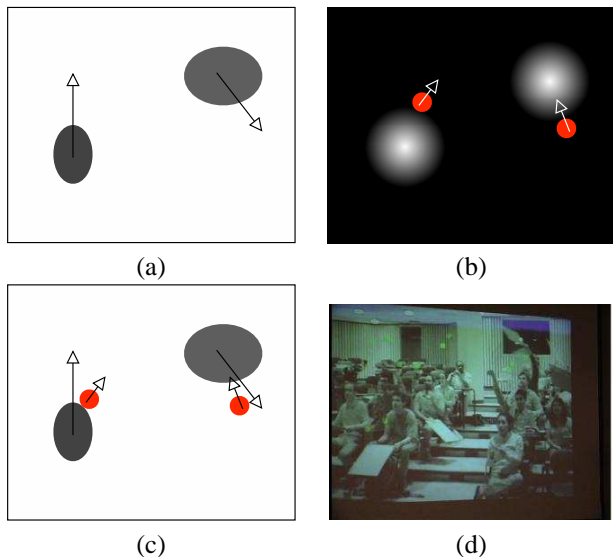


Fig. 1. Motion swarm interaction [6]. Two objects move in a video image (a), creating a motion field (b). Simulated particles, superimposed in red, react to the motion field. A kinetophobic particle on the left is repelled by the motion while a kinetophilic particle on the right is attracted to the motion. When the particles are viewed superimposed on the original image (c), the kinetophobic particle appears to be pushed by an object while the kinetophilic particle is drawn to an object. Audiences can interact with the particles (d), by *propelling* them around a display with their motion. In this image, the audience creates music by propelling particles to the top of the display to trigger sound events.

the simulated particles are superimposed on the video images, the objects in the image appear to propel the particle, creating a mechanism for interaction. Parameters of the simulation allow particles to be attracted to motion (kinetophilic) or repelled by motion (kinetophobic).

Motion swarms are particularly suited to interaction with *Swarm Art* [7] where a simulated motion swarm particle acts as the world centre for a flock of simulated boids [8]. People viewing the flock can modulate its motion by manipulating the motion swarm particle.

The particle motion must have some additional *structure* in the form of other constraining forces for interaction to be practical. The following describes a set of constraints that provide this structure.

Bounding Box: It is useful to constrain the particles to move only within a defined bounding box. At the very least, a bounding box the size of the images is necessary to keep the particles from going beyond the image boundaries. However, bounding boxes can also be useful to define smaller regions of interest.

Friction: Frictional forces act in opposition to particle velocity. Friction allows the particles to slow down in the

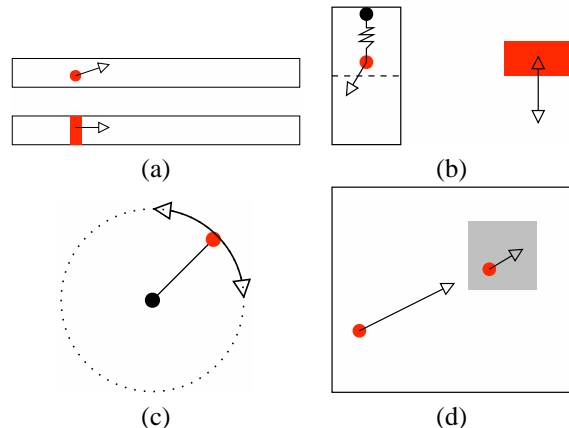


Fig. 2. Video widgets built from particle motion constraints.

absence of motion, and can prevent particles from *shooting past* regions of motion because their velocity is too high, confounding interaction.

Velocity Bounds: In addition to using friction to limit particle speed, we also can place bounds on the particle velocity.

Anchors: We can *anchor* a particle by defining a central position and adding a force that propels the particle toward that position. This is useful when we want to maintain a distribution of points throughout the image space. We model the anchoring force as a spring between the anchor position and the particle.

Pivots: In an elaboration on an anchor, we can force the particle to remain a fixed distance from an anchor point. This allows the particle to move in a circular orbit, responding to tangential forces generated by the motion field.

Each of these constraints has tunable parameters that affect the behavior of the particles tangibly. Part of creating a motion swarm system is tuning these parameters to get the right *feel* for the interaction.

3. FROM MOTION CONSTRAINTS TO VIDEO WIDGETS

This section describes how to extend motion-swarm interaction to achieve a more precise, GUI-style interaction. Figure 2 shows the configuration of four example widgets that can be constructed from motion swarm particles. The first, Figure 2(a) illustrates a horizontal slider widget built on a narrow bounding box. The particle is free to move within the two dimensional box (above), but due to the elongation of the box, the motion is mostly horizontal. Motion can push or pull the slider across the box. If only the horizontal position of the particle is displayed, the particle moves as a horizontal slider (below).

Figure 2(b) shows a button widget. A kinetophilic particle anchored at the top of a small box is free to move in

two dimensions (left). Motion near the particle pulls it down. When the particle crosses a horizontal threshold it activates a switch, rendered to move up and down only (right). In order to prevent multiple activations, the switch is disabled until the particle returns above the threshold.

Figure 2(c) shows a dial widget. A particle rotates about its pivot anchor. The angular position of the particle allows motion to set angular input.

In Figure 2(d), a particle is free to move about the display in two dimensions. Regions of the display that correspond to selections or actions have high coefficients of friction. The particle moves quickly until it is directed to a high-friction region where it slows, and in the absence of motion, comes to a stop, thereby *selecting/activating* the box.

Variations on these themes extend the possible interactions. Dividing a slider into discrete positions allows one to select menu items. Kinetophobic versus kinetophilic particles also change the nature of the interaction, but in general kinetophobic particles are more useful: kinetophilic particles *stick* to motion, preventing a person from *releasing* a particle.

4. PILOT STUDY

We developed a set of trial widgets to test on a small sample of subject in a pilot study. The goal of the study was to obtain a first impression of the usability of the widgets and to solicit feed back to guide development of more sophisticated interaction. We collect some demographic data, timed the subjects as they performed some simple tasks with the widgets, and then provided the subjects with an opportunity to provide feedback. Before timing, the subjects had an opportunity to familiarize themselves with the widget as they desired. The small sample (five subjects) consisted of students and faculty ranging in age from under 20 years to over 60 years. The study was approved the the Conjoint Faculties Research Ethics Board at the University of Calgary.

4.1. The Trial Widgets

Horizontal slider: The first trial widget was a horizontal slider as shown in Figure 2(a) using a kinetophobic particle. Figure 3(a) shows the slider in operation. Users were asked to move the slider from its left-hand extent to the right until it was aligned with a marker on the right-hand side. When the slider lined up, a disc-shaped indicator on the right became green.

Dial: Figure 3(b) shows a dial widget made from a kinetophobic particle rotating about a pivoting anchor (Figure 2(c)). The display showed the particle and a line through the anchor to the particle. Subjects were timed while moving the particle from the right-hand side of the circle, to the indicator on the upper left-hand side.

Horizontal slider menu: Figure 3(c) shows a horizontal slider widget as used in the first trial, but with a menu con-

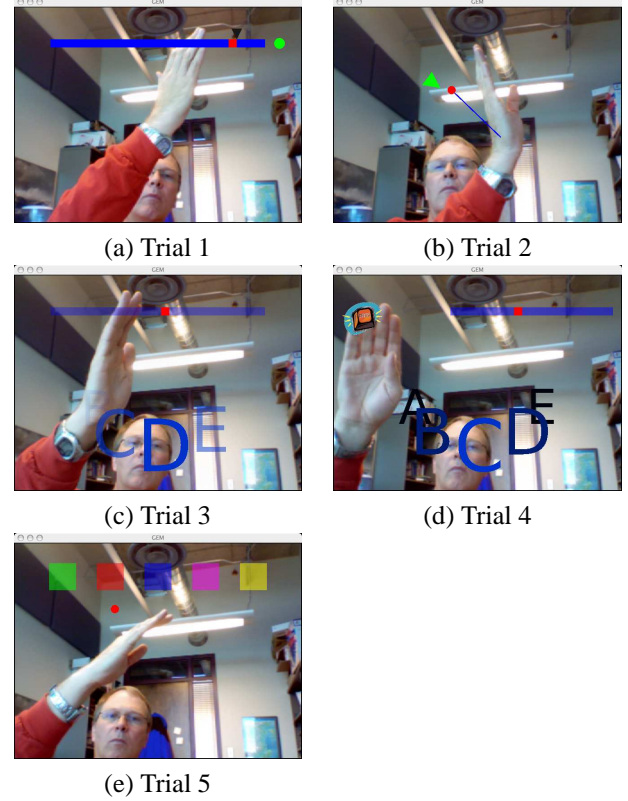


Fig. 3. Screen grabs of sample interaction trials: (a) a horizontal slider, (b) a rotary dial, (c) a horizontal slider operating a menu, (d) a horizontal slider and menu with activation button, and (e) a two-dimensional position mechanism.

sisting of the letters A through F displayed below. As the slider moves, the menu items appear to rotate through the display. The slider is divided into intervals within which the menu *sticks* to a single item. The effect is that within a range of slider motion, the menu item is stable. Users started with item A selected and were asked to select item E.

Horizontal slider menu with button: It is useful, if not essential, to indicate the selection of a menu item by activating a button. Figure 3(d) shows a variation on the previous menu that includes a button (Figure 2(b)) in the upper left-hand corner. Users were asked to select the items B-A-D with a combination of slider and button activations, starting from item A.

Two-dimensional box selection: Figure 3(e) shows the last trial widget, a particle with full freedom in two dimensions, and a set of high-friction boxes corresponding to item selections. Subjects were asked to maneuver the particle from the bottom centre of the display to select the red box.

4.2. Results

Table 1 summarizes the timing results for all five trials. Without a basis for comparison, it is difficult to say if these times

	Trial				
	1	2	3	4	5
minimum	5	3	7	14	5
maximum	35	21	16	52	19
mean	13.2	9.4	10.6	27.4	10.0
median	8	8	9	24	6
r_S	0.36	0.82	0.97	0.82	0.03

Table 1. Summary of times (seconds) for pilot study results. r_S is the Spearman rank correlation between trial time and subject’s age.

are fast or slow, but with the exception of a couple of trials for one of the subjects, all subjects quickly and competently completed the tasks. We also observed a strong correlation between the subject’s age and time to perform the tasks. This suggests a bias that favours younger users, but the sample is small, and the correlation was not there for all of the trials.

Positive comments about the interface suggested that for most subjects the interfaces were fun, and especially younger subjects were quick to suggest and interface to a game of some sort. Clearly, the idea of applications in games generated excitement. One subject was comfortable enough with the interface to use his head to steer the particle in the last trial, while still accomplishing a time near the median.

In trial four, most subjects chose to use a single hand, perhaps because that was sufficient for the first three trials. Those who use their left hand to operate the button while using their right for the slider found the interface easier to use.

Trial four was also the most difficult and highlighted one of the difficulties with this approach. When subjects used a single hand, they had a tendency to activate the button accidentally while adjusting the slider. The interface reacts to motion, and has no way to distinguish between deliberate and incidental motion.

One interesting observation in the last trial was that all but one subject manipulated the particle by separating degrees of freedom. That is, they would first move the particle horizontally (or vertically) to the correct position, then finish the task by moving the remaining degree of freedom. The subject who chose to follow a straight line had the fastest time, but only by a small margin.

5. DISCUSSION

Although just a pilot study with a small sample, the test results are encouraging. It is certainly possible to build interfaces that people can use. Motion-swarm widgets will not replace the ubiquitous keyboard and mouse GUI interfaces, but they open a realm of more complex interaction in complex environments.

The biggest difficulty with these widgets is the effect of incidental motion. Two things can be done to address this.

First, as selected items change, the interface and change properties to help the particle *stick* with the selected item. Second, we can make erroneous selections easier to correct. One limitation on the applications of motion-swarm interaction is the requirement to mirror the subjects in the display. No subjects mentioned this as a problem, but in the contrived trials, this would not have been as annoying as it might be in other applications. This is the price for avoiding camera calibration, segmentation, recognition, and tracking.

The potential for interacting with video games using this technology is the next step. Two-dimensional arcade-style games are the low-hanging fruit, but we anticipate that other, more elaborate game interfaces are possible. We are also speculating on taking a step backward with motion-swarm widgets, and testing them with an audience. Although operating a GUI en masse might be difficult, it suggests intriguing possibilities for interactive theatre.

6. REFERENCES

- [1] M. Kölsch, V. Pavlović, B. Kisačanin, and T. Huang, “Guest editorial: special issue on vision for human-computer interaction,” *Computer Vision and Image Understanding*, vol. 108, no. 1-2, pp. 1–3, October 2007.
- [2] M. Lew, E. M. Bakker, and N. Sebe, and T. S. Huang, *Human-computer intelligent interaction: a survey*, vol. 4796/2007 of *Lecture Notes in Computer Science*, pp. 1–5, Springer, Berlin/Heidelberg, 2007.
- [3] A. Jaimes and N. Sebe, “Multimodal human computer interaction: a survey,” *Computer Vision and Image Understanding*, vol. 108, no. 1-2, pp. 116–134, October 2007.
- [4] Viola and Jones, “Rapid object detection using a boosted cascade of simple features,” in *Computer Vision and Pattern Recognition*, Kauai, Hawaii, December 2001, vol. 1, pp. I–511–518.
- [5] J. W. Davis and A. F. Bobick, “The representation and recognition of human movement using temporal templates,” in *IEEE Computer Vision and Pattern Recognition*, June 1997, pp. 928–934.
- [6] Q. Nguyen, S. Novakowski, J. E. Boyd, C. Jacob, and G. Hushlak, “Motion swarms: video interaction for art in complex environments,” in *ACM Multimedia 06*, Santa Barbara, CA, October 2006.
- [7] C. Jacob, G. Hushlak, J. E. Boyd, P. Nuytten, M. Sayles, and M. Pilat, “Swarmart: interactive art from swarm intelligence,” *Leonardo*, vol. 40, no. 3, pp. 248–254, June 2007.
- [8] C. Reynolds, “Flocks, herds, and schools: A distributed behavioral model,” *Computer Graphics*, vol. 21, no. 4, July 1991.