Shape of Motion and the Perception of Human Gaits

Jeffrey E. Boyd
Department of Electrical and Computer Engineering
University of California, San Diego, La Jolla CA 92093-0407

James J. Little
Department of Computer Science, University of British Columbia
Vancouver, B.C., Canada V6T 1Z4

Abstract

Researchers in computer vision have recently demonstrated several systems that interpret motion and optical flow without using a model of kinematic structure. These non-structural methods usually integrate a field of motion into a more compact representation. In this paper, we present the design of an experiment to investigate the relationship between the non-structural shape-of-motion algorithm and human perception of gait. We take the features used by the algorithm, and use them to synthesize gait-like optical flow. A group of subjects then views the flow stimuli and records their perceptions. The motion stimuli are designed to differ in structure but have similar shape-of-motion. We wish to show that we can vary the structure of a stimulus without altering its perception, so long as we maintain the shape-of-motion. Pilot study results illustrate the experiment design. By performing this experiment to relate gait perception with gait synthesis, we are able to probe the computer vision algorithm.

1: Introduction

Recent research in computer vision shows an interest in methods for perception of human locomotion and other activities. The methods fit into two broad categories: structural and non-structural. Structural methods use a model of human kinematic structure and possibly dynamics. In contrast, non-structural methods (sometimes referred to as *appearance-based* or *model-free*) avoid using such models. For example, Little and Boyd demonstrate non-structural gait recognition using *shape-of-motion* features [13, 6, 14] (described in Section 2). Polana and Nelson [15, 16, 17] look at global spatial distributions of motion for a figure engaged in some activity. They are able to recognize different activities by comparing motion statistics computed over a coarse mesh. Baumberg and Hogg [3] give a method to describe the shape of a walking human body as a function of time. In later work [4], they describe the variation in the shape over time as the changing shape of a vibrating plate. Bobick and Davis [10] describe another non-structural approach that analyzes the shape of a motion-energy image (MEI), a summation of optical flow over a sequences of images. Features that describe shapes in the MEI are used to recognize activities.

A common theme in these non-structural methods is the integration of a field of motion into a more compact representation. There is evidence in the psychophysics literature to suggest that spatial integration of motion is also important in human perception of motion. For example, Williams and Sekuler [20] show that perception of a field of randomly moving points is related to motion of the field as whole, i.e., spatial integral of the motion. If the motion of the points is randomly distributed over all directions, then there is no perceived large-scale motion. However, if they are distributed only over a smaller range of angles, then there is indeed a perception of the points moving *en masse*. Boyd and Little [6] sought to explain other psychophysical observations [2, 5, 11, 12, 18], based on moving light displays (MLDs), in terms of *shape-of-motion* features. The explanation was based on a comparison of psychophysical results and the properties of the *shape-of-motion* algorithm, indicating that there are many consistencies. The comparison leads to some conjecture about what one might observe in human perception if it is indeed related to *shape-of-motion*.

Our goal is to explore the relationship between non-structural properties of motion and optical flow. We form the hypothesis that motion stimuli may be treated as a field of optical flow that can be effectively characterized by a set of global *shape-of-motion* features. This paper presents the design of an experiment that tests this hypothesis, focusing on human locomotion.

The experiment uses the relationship that arises between perception and synthesis. We create a set of motion stimuli, all but one of which have non-structural *shape-of-motion* features like those for a gait, but varying in their underlying kinematic structure. If the hypothesis is true then we should see that perception is independent from the selection of stimulus. On the other hand, if the hypothesis is false and structure is critical, then the stimuli should all be perceived differently.

Although the experiment is primarily psychophysical in nature, it demonstrates a useful method to test a computer vision algorithm. We start with a computer vision algorithm that we know can perceive differences in human gaits under a set of controlled conditions. We then turn around and use that algorithm to synthesize human gaits. If the synthesis is successful, then we may assume that the algorithm is sensitive to the right things, at least for a gait. If not, then we know that there is more to the gait than what the algorithm measures, and perhaps we get a clue as to what else we should look for in the motion. The quality of the synthesis provides a means to evaluate the algorithm.

2: Background

2.1: Shape-of-Motion

Figure 1 illustrates the flow of data through the *shape-of-motion* system to create non-structural features that are used for recognition of individual gaits [14]. The system begins with an image sequence of $n+1$ images featuring a single pedestrian walking in front of a static background, and then derives n dense optical flow images. For each of these optical flow images, the system computes m characteristics that describe the shape of the motion (i.e., the spatial distribution of the flow), for example, the centroid of the moving points, and various moments of the flow distribution. Some of these are pixel coordinates, but all are treated as time-varying scalar values. Table 1 summarizes the scalar values used. Rearranging the scalar values forms a time series for each scalar. A walking pedestrian undergoes periodic motion, returning to a standard position after a certain time period that depends on the frequency of the gait. The system analyzes the periodic structure of these time series and determines the fundamental frequency of the variation of each scalar.

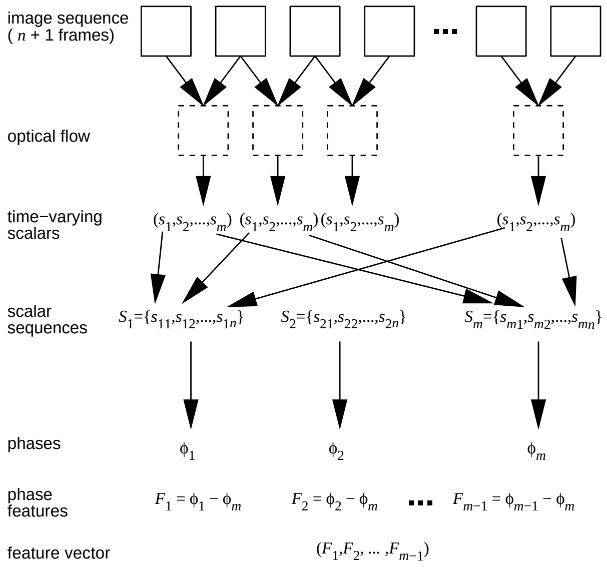


image sequence (n + 1 frames)
optical flow
time−varying scalars
$(s_1,s_2,...,s_m)$ $(s_1,s_2,...,s_m)$ $(s_1,s_2,...,s_m)$ $(s_1,s_2,...,s_m)$
scalar sequences
$S_1=\{s_{11},s_{12},...,s_{1n}\}$ $S_2=\{s_{21},s_{22},...,s_{2n}\}$ $S_m=\{s_{m1},s_{m2},...,s_{mn}\}$
phases
ϕ_1 ϕ_2 ϕ_m
phase features
$F_1 = \phi_1 - \phi_m$ $F_2 = \phi_2 - \phi_m$ $F_{m-1} = \phi_{m-1} - \phi_m$
feature vector
$(F_1,F_2, \cdots ,F_{m-1})$

Figure 1. The data flow for *shape-of-motion* gait analysis. The spatial distribution of the optical flow computed for each frame of a video sequence is characterized by a set of scalars. Over the duration of the sequence, each scalar forms a time series that is cyclic. The phase relationships among the time series give features that are useful for gait recognition.

The set of time series for a view shares the same frequency, or simple multiples of the fundamental, but their phases vary. To make different sequences comparable, the system subtracts a reference phase, ϕ_m, derived from one of the scalars. Each image sequence is characterized by a vector, $F = (F_1, \ldots, F_{m-1})$, of $m - 1$ relative phase features. Little and Boyd showed that some of the features are sensitive to individual gaits, and that the sensitivity can be used for recognition.

At no point in *shape-of-motion* recognition is the structure of the walking subject recovered. The process operates entirely without a model of human kinematics.

2.2: Human Perception of Gait

Much of the psychophysics literature pertaining to motion perception refers to moving light displays (MLDs). MLDs are useful because they conceal the underlying structure of an image so that it cannot be perceived from a static image, only a moving sequence [11, 12]. We have suggested that gait perception from MLDs may not require identification of kinematic structure [6]. This is based on two observations:

1. the *shape-of-motion* algorithm applies equally well to both gray-scale images and MLDs, and
2. various observations concerning human perception of moving light displays can be explained in terms of the algorithm.

The first observation is based on a comparison of *shape-of-motion* features derived from gray-scale sequences and features derived from the equivalent MLD. The features are similar for the two types images. The second observation is based on the following evidence reported

Description	Label	Formula
x coordinate of centroid,	x_c	$\sum xT / \sum T$
y coordinate of centroid,	y_c	$\sum yT / \sum T$
x coordinate, centroid of $\|(u,v)\|$ distribution	x_{wc}	$\sum x\|(u,v)\|T / \sum \|(u,v)\|T$
y coordinate, centroid of $\|(u,v)\|$ distribution	y_{wc}	$\sum y\|(u,v)\|T / \sum \|(u,v)\|T$
x coordinate of difference of centroids	x_d	$x_{wc} - x_c$
y coordinate of difference of centroids	y_d	$y_{wc} - y_c$
aspect ratio (or elongation) – ratio of length of major axis to minor axis of an ellipse	a_c	$\lambda_{max}/\lambda_{min}$, where λs are eigenvalues of second moment matrix for motion distribution
elongation of weighted ellipse	a_{wc}	as in a_c, but for weighted distribution
difference of elongations,	a_d	$a_c - a_{wc}$
x coordinate, centroid of $\|u\|$ distribution	x_{uwc}	$\sum x\|u\|T / \sum \|u\|T$
y coordinate, centroid of $\|u\|$ distribution	y_{uwc}	$\sum y\|u\|T / \sum \|u\|T$
x coordinate, centroid of $\|v\|$ distribution	x_{vwc}	$\sum x\|v\|T / \sum \|v\|T$
y coordinate, centroid of $\|v\|$ distribution	y_{vwc}	$\sum y\|v\|T / \sum \|v\|T$

Table 1. Summary of scalar *shape-of-motion* descriptors. Summations are over the entire image. u and v are the x- and y-direction optical flow values respectively. The function T segments the image. $T = 1$ for pixels that are moving and $T = 0$ for stationary pixels.

in the psychophysics literature.

Sumi [18] looked at the effect of inverting the display of an MLD on recognition. His observations showed that while most people were often able to recognize a gait from an inverted MLD, they reported that the gait looked odd and failed to recognize that it was, in fact, a normal gait that was inverted (a perception that is contrary to the kinematic structure). Inverting the MLD reverses the phase of some, but not all of the *shape-of-motion* features. Furthermore, the relative phase of the features is preserved. Thus, the *shape-of-motion* is not expected to change much for the inverted image, making the observation of human perception consistent with the algorithm. Sumi's subjects often reported seeing gaits but thought that the gait was odd. Perhaps the oddness represents a failure to map human structure onto the stimulus once the subject believes that the stimulus represents a human.

Barclay *et al.* [2] showed that a detectable amount of gender recognition is possible from MLDs, although the recognition was only slightly better than chance. This evidence does not indicate the mechanism used for the recognition, but the success of the *shape-of-motion* algorithm suggests that kinematic structure is not necessary. In fact, Little and Boyd [14] suggest that although *shape-of-motion* features are not structurally based, they may be sensitive to variations in the build of the observed person, explaining the algorithm's ability to recognize individuals. The MLD evidence does not preclude a non-structural mechanism.

Bertenthal and Pinto identify the importance of phase in gait perception [5]. Their experiments use a stimuli that consists of an MLD for a walker, masked by dots moving in the background. Some of the stimuli are perturbed by altering the phase of oscillation of a limb. They observed that perception of gait from the unperturbed stimulus was better than for the perturbed one, but both were significantly better than chance. Bertenthal and Pinto assumed that kinematic structure was important and perturbed only the phase of the motion. Clearly, phase is important for perception of motion in humans, just as it is for the *shape-of-motion* algorithm.

2.3: Integrated Features of Cyclic Motion

Boyd and Little [6] offered the following conjecture to explain the psychophysical observations, based upon the *shape-of-motion* algorithm. Consider the image of a pedestrian to be a collection of moving points. As a simple approximation, the motion for each point in the pedestrian can be expressed as the sum of a linear motion and an oscillatory motion. For example, let the x-coordinate of an arbitrary point i be

$$x_i(t) = x_{i0} + v_x t + A_i \cos(\omega t + \phi_i), \tag{1}$$

where x_{i0} is a constant, v_x is the mean velocity of the person, and A_i, ω and ϕ_i are the amplitude, frequency and phase of the oscillation. $x_{i0} + v_x t$ is the linear part of the motion. All points share the same frequency, ω, but vary in A_i and ϕ_i depending on where they are in the body. The x-coordinate of the centroid is

$$\bar{x}(t) = \frac{1}{n}\sum_i x_i(t) = \frac{1}{n}\sum_i \{x_{i0} + v_x t + A_i \cos(\omega t + \phi_i)\}, \tag{2}$$

where n is the number of image points in the pedestrian. As part of extracting a phase feature we discard the linear portion of the motion leaving

$$\bar{x}(t) = \frac{1}{n}\sum_i A_i \cos(\omega t + \phi_i). \tag{3}$$

The summation in Equation (3) is the sum of a set of phase vectors, or phasors. Phase vectors are commonly used to perform computations with rotating vectors that share a common frequency, such as in electrical power systems. In short, the summation can be treated as the sum of a set of vectors, each vector having magnitude A_i and direction ϕ_i.

Several conclusions may be drawn from this conjecture:

1. Recognition of objects from MLDs does not necessarily happen because they capture the structure of objects in a scene, but because they adequately sample the motion of points in the entire object.
2. If it is only necessary for the dots in an MLD to sample the motion, then perception of gait should not require that the dots be at joints, since that is purely a structural concern.
3. Any MLD that has *shape-of-motion* features corresponding to gait should be perceived as gait.

The remainder of this paper describes the experiment we designed to to test the conjecture of Boyd and Little.

3: Experiment Design and Motion Stimuli

The conjecture in the previous section indicates how one should synthesize an optical flow field so that is has the same characteristics as the field for a human gait. We test the perception of flow synthesized this way, to see if it is indeed perceived as a human gait. The use of an algorithm for perception as a tool for synthesis can point to the specific part of the motion that the algorithm is sensitive to.

We use a factorial experiment to test the perception. Ideally, we should use the *posttest-only control group design* described by Campbell and Stanley [7], but the number of unordered combinations of factors is too numerous and forces us to use some repeated measures. Section 4.1 describes the actual method we used in our pilot study. We discuss a final design in Section 5.

In the experiment, we present subjects with a motion stimulus and ask them to record, in not more than two or three words, what they perceive in the motion. We design the stimuli to have specific properties that are controlled by a set of factors. The factors, our independent variables, are:

1. **the motion:** how the motion was created for the stimulus and what its *shape-of-motion* features are,
2. **encoding:** how the motion was encoded into images in video sequence, and
3. **duration:** how long the subjects view the stimulus.

The following subsections describe these factors in detail.

3.1: Motion

The motion in the stimuli is a critical part of the experiment since we want to relate our results to the *shape-of-motion* algorithm. Our intention is to generate synthetic fields of optical flow with the *shape-of-motion* features of a human gait. Unfortunately, it is not a simple matter to create images to such a specification. We must also consider that it may not be necessary to reproduce all the *shape-of-motion* features in the stimulus. The reason for this stems from the difference between recognizing some motion as a gait, and recognizing and individual gait. In recognition tasks, *shape-of-motion* features that are consistent for individual gaits, but vary greatly over the population of gaits are desired. To simply recognize a gait requires a feature that remains consistent over all gaits, but varies greatly over the population of all motions. Sorting out what features are important for what perception is will not be a simple task. We make a start here by focusing on x_c and y_c.

In lieu of an algorithm to create synthetic flow fields to our specification, our strategy is to start with data for a real pedestrian, and then to distort it in a manner that maintains the *shape-of-motion* but may or may not be consistent with the underlying kinematic structure. Specifically, we generate the five motion patterns as follows:

On Joints (*on-joints*) The first stimulus is a simple MLD of a walking person. For this we use the synthetic canonical walker described by Cutting [9], which has the lights placed on major joints. We relied on a translation of Cutting's original Fortran program into C that was used by Bertenthal and Pinto [5].

Off Joints (*off-joints*) We perturb the *on-joints* stimulus by moving the points of light away from the joints and along the limbs. The displacement is random, sampled from a uniform distribution ranging from -0.75 to 0.75, and scaled by the length of the appropriate limb. For example, the point at the knee can end up anywhere from three quarters of the way up the thigh to three quarters of the way down the shin. Points for head, wrists and ankles can only move towards the shoulder, elbows and knees respectively, thus ensuring that all points are somewhere on the body, but not necessarily on the specific joint. Cutting's algorithm considers occlusions of points in an ad hoc manner. We ignore these occlusions to simplify the synthesis.

Off Body (*off-body*) The *off-body* stimulus is also a perturbation of *on-joints*, but rather than keep all the points on the body, we move the points by a constant random displacement in a random direction. We select the magnitude of the displacement from a normal distribution, mean zero and standard deviation of 20 pixels (the torso of the walker is about 35 pixels). The direction of the displacement is selected from a uniform distribution ranging from $-180°$ to $180°$. We ignore the occlusions.

Oscillating Off Body (*osc-off-body*) The *osc-off-body* stimulus is like *off-body*, with the exception that the displacement is not constant, but oscillates at the same frequency as the gait. The effect is one of the light orbiting about a fixed point on the joint of the walker. We ignore the occlusions here too.

Random Oscillations (*random-osc*) The *random-osc* stimulus is the *straw man* in the set of motion. It consists of a set of randomly spaced points, placed such that the aspect ratio is about that of *on-joints*. The phases of the oscillations are randomly distributed over $-180°$ to $180°$. The amplitudes are normal with standard deviation of one quarter of the height of the distribution.

Selected frames from the five motions are shown in Figure 2 in MLD form. The perturbations of points in *off-joints*, *off-body* and *osc-off-body* are chosen such that the coordinates of the centroids should be nearly unchanged. For *off-body* and *osc-off-body*, this happens because the magnitude and direction of the displacement are unbiased. There is a bias to the direction of displacement for *off-joints* because the points are restricted to lie on the limbs, and because points at the extremities can only be displaced towards the center of the body. It is difficult to predict what will happen to the other features under these perturbations.

We encountered difficulty computing the *shape-of-motion* features from the synthetic stimuli. The first problem was with the motion of the walker with respect to the background. Previous work on *shape-of-motion* used sequences of pedestrians walking across a static background, allowing us to rely on motion to segment the moving figure. The Cutting synthetic walker data yields points for a static walker before a static background, giving the impression that the person is walking on a treadmill. This makes the torso nearly stationary. We were not able to resolve the tiny motions in the torso with the optical flow algorithm so they could not contribute to the distributions, and gave results that we could not interpret. We then resorted to computing an ideal optical flow given that we know the positions of the points and the frame-to-frame correspondences. Still, we could not see well-defined peaks in the spectra of the features. In the end we added a constant x velocity to all the points to simulate a walker moving across a static background. This failure of the synthetic data to behave like the real data that we had used often in the past caught us off guard and is a weakness in this experiment that we have yet to resolve. Clearly, the subjects will not see treadmill motion and we have yet to relate *shape-of-motion* to treadmill gaits.

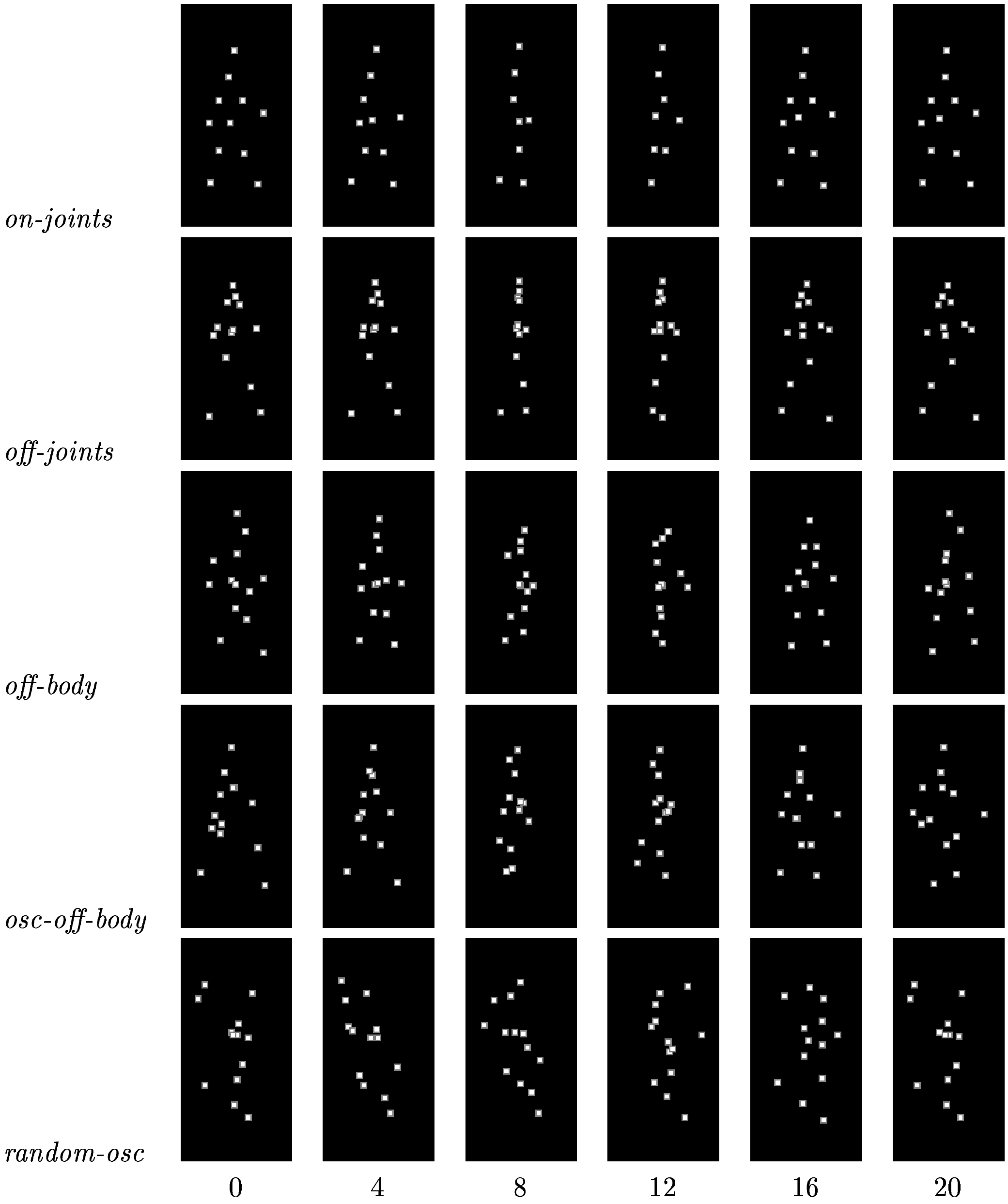


Figure 2. Sample images from motion stimuli: (*on-joints*) the Cutting canonical walker MLD sequence, (*off-joints*) walker with dots on the body, but moved away from joints, (*off-body*) walker with dots randomly displaced from joints, (*osc-off-body*) walker with dots displaced from joints by oscillating distance of random amplitude and phase, and (*random-osc*) a set of random dots moving cyclically with random phase and amplitude, same frequency as *on-joints*.

Table 2 summarizes the *shape-of-motion* features for the five stimuli, computed using point correspondences and a constant x velocity added. The right column of the table indicates the between-person standard deviations for the feature that was measured by Little and Boyd [14]. y_c is the reference phase and is, therefore, always zero. The table

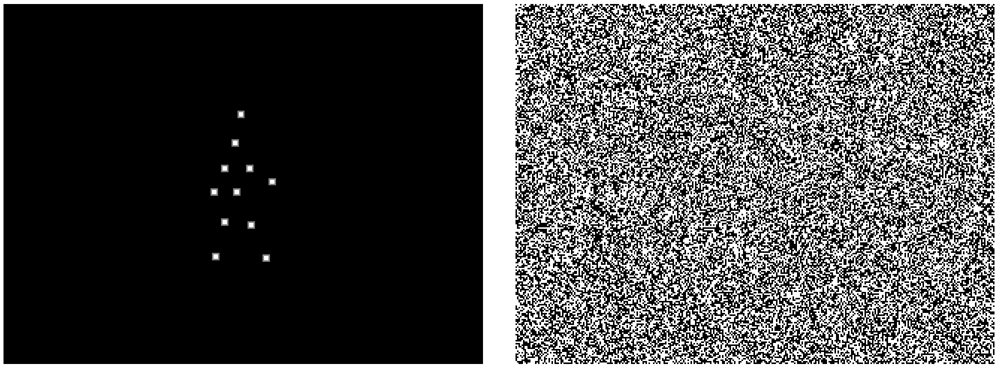
			Stimulus			
Feature	*on-joints*	*off-joints*	*off-body*	*osc-off-body*	*random-osc*	σ
x_c	0.11	0.16(0.8)	0.14(0.6)	0.19(1.3)	0.30(2.8)	0.065
y_c	0.00	0.00	0.00	0.00	0.00	-
x_{wc}	0.05	0.18(2.9)	0.19(3.1)	0.18(2.8)	0.27(4.7)	0.045
y_{wc}	0.00	0.26(3.1)	0.23(2.7)	0.18(2.2)	-0.07(-0.9)	0.083
x_d	0.47	-0.31(6.9)	-0.25(8.8)	-0.34(5.9)	-0.37(5.0)	0.032
y_d	-0.28	-0.18(1.3)	-0.22(0.8)	-0.26(0.3)	0.41(-3.9)	0.077
a_c	-0.04	-0.06(-0.2)	-0.02(0.3)	-0.06(-0.3)	0.19(2.9)	0.077
a_{wc}	-0.04	-0.05(-0.2)	-0.04(0.0)	-0.07(-0.5)	0.27(5.1)	0.063
a_d	0.39	-0.14(-2.9)	0.12(-1.5)	0.01(-2.1)	-0.09(-2.6)	0.18
x_{uwc}	0.04	0.18(3.3)	0.22(4.1)	0.17(3.0)	0.26(5.1)	0.045
y_{uwc}	0.00	0.26(2.7)	0.23(2.5)	0.19(2.0)	-0.08(-0.9)	0.094
x_{vwc}	0.08	0.37(2.2)	0.09(0.0)	0.18(0.7)	0.27(1.4)	0.13
y_{vwc}	0.02	0.42(3.7)	-0.07(-0.8)	-0.09(-1.0)	0.16(1.3)	0.11

Table 2. Summary of $shape\text{-}of\text{-}motion$ features for the five motion stimuli. Values are phases scaled to lie in the range $[-0.5, 0.5]$. Values in the parenthesis are deviations from the feature value for *on-joints* scaled by σ (in the right column). σs are the between-person standard deviation for the feature computed for earlier recognition experiments. These are used as a rough guide only for comparing features.

verifies that x_c and a_c values for *off-joints*, *off-body* and *osc-off-body* are similar to that for *on-joints*. As expected, the weighted distribution and difference features vary from the *on-joints* value the most. *random-osc* is significantly different from *on-joints*.

3.2: Encoding

When creating the motion stimuli, we produce a sets of point coordinates (as described in Section 3.1; one set per time slice in the motion sequence. We encode those points into a motion stimulus in two ways. The first is to produce the well-know MLD. Each point coordinate gives an image coordinate for a point of light on a black background in the frame of a stimulus. The second type of encoding results from an effort to further hide the structure behind the motion. We create a static binary background image where each pixel is selected randomly from $\{black, white\}$, and $P(black) = P(white) = 0.5$. Then rather than represent the pixel coordinates with points of light, we use a small patch of binary pixels (generated using the same criteria as the background). Although we know from Johansson [11] that without *a priori* information regarding the contents of an image, a single frame of an MLD is difficult to interpret, if one knows that a frame represents a human, one could connect the dots and guess at the kinematic structure. With our random background and patch stimulus, a static frame cannot convey any information at all. All of the pixels are randomly generated. The sequence must be moving in order to see anything beyond uncorrelated pixels. MLDs have been adequate in the past for evaluating the perception of motion, and we have no particular reason to believe they would not suffice for our purposes. We derived this encoding scheme while trying to generate fields of optical flow, and this experiment offers an opportunity to test it. The encoding variable can have values M and R for MLDs and random dot patterns respectively. Figure 3 shows a sample



(a) (b)

Figure 3. Examples of images in a motion stimulus illustrating the two encoding methods: (a) MLD and (b) random dot encoding. Nothing can be seen in the random dot image without any motion since there is no correlation between pixels.

frame for each of the two encoding methods.

3.3: Duration

In humans, the length of time a stimulus is presented can affect the perception. Presumably, a longer period of time allows slower mechanisms to come into play. Johansson [11], Barclay *et al.* [2], and Sumi [18] all cite variations in perception with the duration of the stimulus. Their observations suggest that a duration from about $1s$ up to about $4s$ should suffice. For the pilot study, we use four durations, $1s$, $2s$, $4s$, and $8s$.

4: Pilot Study

We collected data for a pilot study [8] to debug the design of the experiment. This section presents the experimental method and a sample analysis of the results.

4.1: Method

The five motion stimuli (*on-joints*, *off-joints*, *off-body*, *osc-off-body*, and *random-osc*), the two motion encodings (M and R) and four durations ($1s$, $2s$, $4s$, and $8s$) give a total of 40 separate motion sequences in the factorial experiment. We created digital movies at a resolution of $240 pixels$ by $320 pixels$, for each of these sequences using a Silicon Graphics workstation. We then transferred the sequences onto an SVHS video tape, in a random order, scaling the images to just fill the entire screen. We selected a group of subjects (students in a computer vision class) and showed every subject every stimulus in the same order. Before viewing the stimuli, we gave the subjects a sheet of paper with instructions saying that:

- they are about to see a series of 40 video clips,
- the clips vary in duration from about one second up to eight, and
- they should indicate what the believe is the cause of the motion in each clip.

They received no other information about the experiment. The paper had 40 numbered blanks with space for at most 3 or 4 words in which the subjects could record their responses. We review the responses and categorize them. This is a subjective process, but leaves the subjects free of expectations about the stimuli.

Once the responses are categorized, the raw data consists of a tuples containing the levels of the three independent variables, and the dependent variable, the categorized responses. There are 40 tuples for each subject, barring missing data. We form the data into a 4-way contingency table and test for dependencies among the variables [19, 8, 1]. Our analysis is described further in Section 4.2.

We recognize that by showing all subjects identical stimuli in the same order confounds validity due to the effects of multiple measures. With 40 stimuli it was simply convenient to show a single group all the stimuli in one shot. While the data in the pilot study may not lead to valid conclusions, it is sufficient for the purposes of the pilot study. In Section 5 we describe a better method for collecting data that will reduce the problems with multiple measures in the ultimate experiment.

4.2: Analysis

The first step in analysis of the data was to categorize the subject responses. After scanning the data, we decided that the responses fell naturally into three categories:

1. a person walking (W),
2. a person walking but in an unusual manner (WS), and
3. a some other motion not related to humans or human locomotion(X).

It was obvious when a subject perceived a normal human gait. Responses that we categorized as walking, but strange included

- the specific response that it was an odd human gait,
- a human gait, but the person is carrying something,
- a human gait, but the person is stepping back and forth, and
- a human running.

Explanations for the randomly oscillating points included:

- points swirling in the wind,
- rotation of some sort,
- random movement,
- a vortex,
- cellular motion, and
- rotating DNA helix.

Clearly the stimulus gave the impression of rotation. Although the process of categorizing the data is subjective, we can give the categories some validity by getting one or two other people to categorize the responses independently.

To analyze the categorized responses, we form a contingency table and look for dependencies among the variables. Tables 3 and 4 combined show the complete contingency table for the pilot study data.

encoding	*duration*	*motion*	*response*		
			W	WS	X
M	$1s$	*on-joints*	4	2	0
		off-joints	4	1	1
		off-body	0	3	3
		osc-off-body	0	3	2
		random-osc	0	0	5
	$2s$	*on-joints*	5	1	0
		off-joints	3	3	0
		off-body	0	4	2
		osc-off-body	0	4	2
		random-osc	0	0	6
	$4s$	*on-joints*	5	1	0
		off-joints	3	1	1
		off-body	1	4	1
		osc-off-body	1	4	1
		random-osc	0	0	6
	$8s$	*on-joints*	5	1	0
		off-joints	2	3	1
		off-body	0	4	2
		osc-off-body	1	4	1
		random-osc	0	0	6

Table 3. Pilot-study contingency table for MLD motion encoding only, showing frequencies split by all variables. Categories for responses are: (W) person walking, (WS) person walking, but strangely, and (X) explanations other than human locomotion. Duration values are: $1s$, $2s$, $4s$, and $8s$. Descriptions for motions are in Section 3.1.

The first thing we look at in the analysis is to determine if there is any dependency of the subject responses to the factors. For this we use a two-way contingency table analysis, comparing the responses against the joint distribution of encoding, duration and motion, i.e, the concatenation of Tables 3 and 4. We use the χ^2 statistic to test the hypothesis that the variables are independent. The G statistic [8] is a popular alternative, but for the pilot study data at least, there are several zero frequencies for which G is undefined. The result of the test for independence gives $\chi^2 = 202.6$ with degrees of freedom $df = 78$. The probability that such values occur if the variables are independent is $p < 0.0001$. We can safely conclude that responses depend jointly on encoding, duration and motion. This is not surprising since a quick scan of the data shows that subjects never mistook *random-osc* for a human and rarely thought that *on-joints* looked like anything other than a human walking.

To find the source of the dependency, we start by examining the conditional dependence of subject response on encoding. There are 20 two-way response versus encoding tables conditioned on the combinations of duration and motion. We compute χ^2 for each table and the sum over all tables[1]. This yields $\chi^2 = 32.6$ ($df = 29$) and $p = 0.29$. Thus

[1] A few of the tables had marginal sums of zero making it impossible to compute χ^2. In those cases we dropped the all-zero columns and compute χ^2 for the remaining data, adjusting df accordingly.

encoding	*duration*	*motion*	*response*		
			W	WS	X
R	$1s$	*on-joints*	4	2	0
		off-joints	4	1	1
		off-body	4	0	1
		osc-off-body	3	2	1
		random-osc	0	0	4
	$2s$	*on-joints*	4	1	0
		off-joints	5	1	0
		off-body	4	1	1
		osc-off-body	3	2	1
		random-osc	0	0	6
	$4s$	*on-joints*	2	3	1
		off-joints	2	1	3
		off-body	1	4	1
		osc-off-body	0	5	1
		random-osc	0	0	6
	$8s$	*on-joints*	4	1	1
		off-joints	4	2	0
		off-body	1	4	1
		osc-off-body	2	3	1
		random-osc	0	0	6

Table 4. Pilot study contingency table for random-dot motion encoding only, showing frequencies split by all variables. Categories for responses are: (W) person walking, (WS) person walking, but strangely, and (X) explanations other than human locomotion. Duration values are: $1s$, $2s$, $4s$, and $8s$. Descriptions for motions are in Section 3.1.

the encoding methods appear to be equivalent. However, an inspection of the individual conditioned tables suggest that, for the *off-body* motion at durations of only $1s$ or $2s$, there is a dependency. A look at the data shows that subjects tended to interpret the *off-body* motion as a normal gait with random-dot encoding, but as a strange gait, or non-human, with MLD encoding. Given the methodological flaws in our data collection, we should still consider the random dot encoding for the ultimate experiment.

In a similar manner, we test for dependency of subject response on duration. There are 10 two-way response versus duration tables conditioned on the combinations of encoding and motion. The result is $\chi^2 = 37.7$ ($df = 45$) and $p = 0.77$, indicating that the responses were independent of duration. This comes as a surprise given the results of others, but can be attributed to the interference of repeated measures (see Section 5).

Finally, we want to know if there is a dependency between subject responses and the motion stimulus. This will indicate whether or not perception is tied to *shape-of-motion*. Again we produce a set of two-way response versus motion tables, one for every combination of encoding and duration. The result is $\chi^2 = 188.0$ ($df = 64$) and $p < 0.0001$, indicating a strong dependency. This is the major source of dependency in the data and as stated previously, because of the *random-osc* stimulus, that is to be expected. However, what we really want to know is whether or not maintaining *shape-of-motion* features while altering

	compare to			
motion	*on-joints*		*random-osc*	
off-joints	12.23(14)	$p = 0.59$	70.42(16)	$p < 0.0001$
off-body	38.19(16)	$p = 0.0014$	55.48(12)	$p < 0.0001$
osc-off-body	32.60(16)	$p = 0.0083$	59.81(13)	$p < 0.0001$

Table 5. Comparison of perception of *off-joints*, *off-body* **and** *osc-off-body* **to** *on-joints* **and** *random-osc*. χ^2 **values and the degrees of freedom in parentheses are the sums** χ^2**s for the two-way comparisons of subject response versus motion conditions on the joint distribution of encoding and duration.**

structure still leads to the same perception. To answer this we repeat the analysis, but for only two motions at a time. For example, when we compare *off-joints* with *on-joints* we get $\chi^2 = 12.23$ ($df = 14$) and $p = 0.59$. In other words, the perception does not appear to depend of whether or not the dots are on the joints. For completeness, we compare *off-joints* to *random-osc* and get $\chi^2 = 70.42$ ($df = 16$) and $p < 0.0001$. Comparisons of *off-joints*, *off-body* and *osc-off-body* to *on-joints* and *random-osc* are summarized in Table 5.

Table 5 shows that for the *on-joints* and *off-joints* stimulus, subject response is independent of the motion. This is the result we are looking for. However, we do not see the same independence for *on-joints* compared to *off-body* and *osc-off-body*.

5: Discussion

The purpose of the pilot study was to debug the experiment design. Therefore, we refrain from drawing conclusions from the data and analysis, and focus on what needs to be changed before collecting data in the final version.

Our data collection method is prone to multiple-treatment interference affecting the external validity (the ability to generalize the results) of the experiment [7]. For the full factorial experiment, each subject views 40 motion sequences, but there are only five different motions in all. A look at the randomized order in which the subjects view the stimuli shows that by the 12th sequence the subjects had viewed all five motions for at least two seconds. By the 19th sequence that rises to four seconds. It was clear from the responses that the subjects had identified the five motions and could quickly relate them to the previous sequences. This confounds the effect of duration in particular. Once a subject has seen the stimulus for four seconds or more, a brief exposure of one or two seconds is likely to produce the same response. In this light, it is not surprising to see that observations did not depend much on duration. Therefore, in the final experiment, we will reduce the number of levels of duration to three ($1s$, $2s$ and $4s$), then split the stimuli by duration, and show any individual subject only sequences of a single duration. This gives us the control we need over duration, but leaves interference within the motion factor. After subjects have seen what is obviously a human, it affects their subsequent responses. To address this we will present the stimuli to subjects in varying order, i.e., a counterbalanced design [7]. So long as there is not a complex interaction among the various motions, this method should be adequate.

Although we do not want to put too much emphasis on the pilot study data, it did suggest that the random dot stimulus was nearly equivalent to the moving light display.

For the final experiment we will use the random dot stimulus only. It is truer to our original intention of generating fields of optical flow, and there was a hint in the pilot data that the ability of the random dot stimulus to conceal structure may yet prove significant. In the worst case, it will make no difference at all. Using only the single encoding scheme also simplifies the analysis by removing a factor.

Another concern is that the *straw-man* stimulus, *random-osc*, is too feeble. In the small sample, no subject ever mistook it for a human. All descriptions suggested a perception of some sort of rotation. In the final experiment, we will replace the random elliptical oscillations with random pendular oscillations. Pendular oscillations should give us a *straw-man* stimulus that is not drastically different from the other stimuli. This will also give more credibility to one of our desired conclusions, that the failure to perceive the random pattern as human is because of *shape-of-motion*, and not for some other reason.

Finally, our goal is to learn about the *shape-of-motion* algorithm, and improve it by understanding its relationship with human perception, if there is one. To achieve this we need a better way to relate the stimuli to *shape-of-motion*. We have taken two steps to accomplish this:

1. perturb a known gait stimulus in such a way that we can predict the effect on at least some of the *shape-of-motion* features, and
2. compute (as best we could) the features of the the resulting randomly perturbed stimuli.

We did see that the x_c and a_c features for *on-joints*, *off-joints*, *off-body* and *osc-off-body* matched well, and did not match *random-osc*. However, we had no control over any of the weighted-distribution and difference features. Several questions remain to be answered about *shape-of-motion* features in this experiment, including:

1. how do we compute *shape-of-motion* not only for a walker moving across the field of view, but also when the field of view tracks the walker, and when the walker is on a treadmill,
2. why does the Cutting synthetic stimulus exhibit optical flow that is so differently from that observed for real stimuli, and
3. which *shape-of-motion* features are important?

This last item is particularly important since the list of *shape-of-motion* features is based purely on intuition about what is important for recognizing motion. Not everything on the list is likely to important, particularly for human observers, and almost certainly there are omissions form the list.

The experiment described tests a computer vision algorithm for recognition of gaits, by seeing whether or not it can be turned around and used to synthesize gait-like optical flow. It is unlikely that we will find, as a result of this experiment, that we can synthesize a completely realistic gait just from *shape-of-motion* features. It will be necessary to follow-up by refining the representation of the flow field used for synthesis, to get more gait-like perception. We can then proceed to repeat the cycle, producing a new algorithm that recognizes the features that were good for synthesis, and so on.

6: Conclusions

We present the design of an experiment that evaluates the quality of a computer vision algorithm by using the algorithm as a guideline for the synthesis of a stimulus, and then evaluating the quality of the resulting stimulus. In a factorial experiment, a group of subjects views a set of synthetic motion stimuli and records their perceptions. The factors that we control are actual motion, the length of time a subject views a stimulus, and how the motion is encoded into images. Four of the motions have the same *shape-of-motion* features as a walking person, but differ in structure, while the fifth contains random motion. We wish to show that we can vary the structure of a synthetic stimulus without altering its perception, so long as we maintain the *shape-of-motion*.

Pilot study results showed flaws in our method of data collection affecting the external validity of the experiment. To correct the flaws we propose the following changes for the final experiment:

- show each subject only sequences of a single duration, and
- show the stimuli to different subjects in different orders.

These changes address the interference of repeated measures that confounds our ability to generalize the results. We also encountered difficulty relating the stimuli to *shape-of-motion* because the algorithm behaved differently when applied to synthetic data than it did for real data.

References

[1] Abacus Concepts, Inc., Berkeley, CA. *StatView*, 1994.

[2] C. D. Barclay, J. E. Cutting, and Lynn T. Kozlowski. Temporal and spatial factors in gait perception that influence gender recognition. *Perception and Psychophysics*, 23(2):145–152, 1978.

[3] A. M. Baumberg and D. C. Hogg. Learning flexible models from image sequences. Technical Report 93.36, University of Leeds School of Computer Studies, October 1993.

[4] A. M. Baumberg and D. C. Hogg. Learning spatiotemporal models from training examples. Technical Report 95.9, University of Leeds School of Computer Studies, March 1995.

[5] B. I. Bertenthal and J. Pinto. Complementary processes in the perception and production of human movements. In L. B. Smith and E. Thelen, editors, *A Dynamic Systems Approach to Development: Applications*, pages 209–239. MIT Press, Cambridge, MA, 1993.

[6] J. E. Boyd and J. J. Little. Global versus structured interpreation of motion: moving light displays. In *Proceedings of IEEE Nonrigid and Articulated Motion Workshop*, pages 18–25, San Juan, Puerto Rico, June 1997.

[7] D. T. Campbell and J. C. Stanley. *Experimental and quasi-experimental designs for research.* Rand McNally, Chicago, 1963.

[8] P. R. Cohen. *Empirical methods for artificial intelligence.* The MIT Press, Cambridge, Massachusetts, 1995.

[9] J. E. Cutting. A program to generate synthetic walkers as dynamic point-light displays. *Behavior Research Methods and Instrumentation*, 10(1):91–94, 1978.

[10] J. W. Davis and A. F. Bobick. The representation and recognition of human movement using temporal templates. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 928–934, San Juan, Puerto Rico, June 1997.

[11] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14(2):201–211, 1973.

[12] G. Johansson. Visual motion perception. *Scientific American*, pages 76–88, June 1975.

[13] J. J. Little and J. E. Boyd. Describing motion for recognition. In *IEEE Symposium on Computer Vision*, pages 235–240, Coral Gables, Florida, November 1995.

[14] J. J. Little and J. E. Boyd. Recognizing people by their gait: the shape of motion. *Videre*, 98. to appear.

[15] R. Polana and R. Nelson. Detecting activities. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2–7, 1993.

[16] R. Polana and R. Nelson. Recognition of nonrigid motion. In *1994 DARPA Image Understanding Workshop*, pages 1219–1224, 1994.

[17] R. Polana and R. Nelson. Nonparametric recognition of nonrigid motion. Technical report, University of Rochester, 1995.

[18] S. Sumi. Upside-down presentation of the johansson moving light-spot pattern. *Perception*, 13:283–286, 1984.

[19] R. E. Walpole and R. H. Myers. *Probability and statistics for engineers.* MacMillan Publishing Co., Inc., New York, 1978.

[20] D. W. Williams and R. Sekuler. Coherent global motion percepts from stochastic local motions. *Vision Research*, 1984.