Using eye tracking to investigate important cues for representative creature motion

Meredith McLendon * Ann McNamara† Tim McLaughlin‡ Ravindra Dwivedi§

Department of Visualization - Texas A&M University

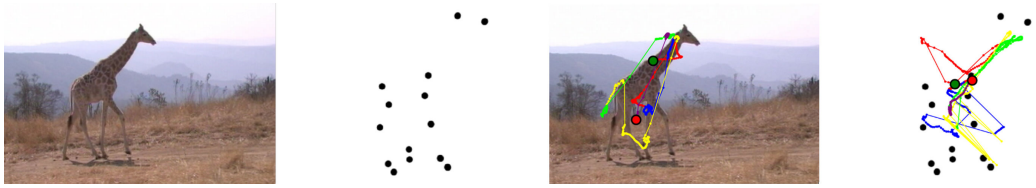

Figure 1: *We compared eye movements over video and point light display sequences. This image shows a single frame, from left to right, full resolution, PLD representation, eyetracked data over full and eyetracked data over PLD. Color changes indicate the passage of time. Results show that gaze patterns are similar over PLD and full resolution video.*

Keywords: animation, perception, eyetracking, point-light display

Abstract

We present an experiment designed to reveal some of the key features necessary for conveying creature motion. Humans can reliably identify animals shown in minimal form using Point Light Display (PLD) representations, but it is unclear what information they use when doing so. The *ultimate* goal for this research is to find recognizable traits that may be communicated to the viewer through motion, such as size and attitude and then to use that information to develop a new way of creating and managing animation and animation controls. The aim of this study was to investigate whether viewers use similar visual information when asked to identify or describe animal motion PLDs and full representations. Participants were shown 20 videos of 10 animals, first as PLD and then in full resolution. After each video, participants were asked to select descriptive traits and to identify the animal represented. Species identification results were better than chance for six of the 10 animals when shown PLD. Results from the eye tracking show that participants' gaze was consistently drawn to similar regions when viewing the PLD as the full representation.

1 Introduction

Artists representing the motion of creatures through moving images understand the importance of the coordinated motion of elements, both relative to one another and relative to the environment. Talented artists can impart identifiable human and animal motion characteristics to even highly abstracted, non-biological forms. Winsor McKay's presentation of a hand-drawn diplodocus in the 1914 film *Gertie the Dinosaur* initiated the modern era of animated creatures. Building upon McKay's work, animators at Walt Disney Productions established the Principles of Animation, defining methods for communicating a character's form, movements, and spirit with each element equally important [Johnston and Thomas 1981]. With the advent of computer animation, though the tools had changed, the Principles of Animation were retained by artists interested in representing human and animal motion [Lasseter 1987].

Computer graphics (CG) offers a variety of methods for defining motion including key-frame animation, data-driven action, rule-based and physically-based motion. Of these, data-driven, in the form of the Motion Capture (MoCap) of humans and animals, provides the most accurate representation of creature motion. However, MoCap technology is currently difficult to employ when the interest is animal motion, particularly wild animals. Creature animators often use video of animals in motion as visual reference for body posture and limb motion during locomotion. Even with expansive reference material, the animator's task of defining motion is compounded by the manner in which motion is defined in computer animation software.

In computer animation, the armature, or rig, provides the mechanism through which artists manipulate digital creatures. A rig is expressed as pivot locations in 3D-Euclidean space around which joints may rotate with varying degrees of freedom. Together with linked joint segments of different lengths, rigs create the appearance of points in space moving in a coordinated manner. When the joint configuration and its motion is inspired by biological motion and rendered surfaces are driven by these transforming points and the appearance of a creature in motion is presented.

Part of the challenge animators face when composing creature motion is that the CG animation tool is fundamentally based upon principles of robotics engineering. The level of abstraction required to transform biological motion into computer animation is a significant impediment. For example, animators must approach the problem of creating a tiger's walk by defining the position of each foot relative to the body and the other feet through the cycle of support (foot on ground), passing (foot off ground and behind opposite leg), high point (foot off ground and in front of opposite leg), and again to support. The ankle and knee positions can be either explicitly defined by the animator using forward kinematics or solved by the computer using inverse kinematics.

An animation rig that is biometrically accurate in structural behavior

*mgmiller@viz.tamu.edu
†ann@viz.tamu.edu
‡timm@viz.tamu.edu
§ravin@viz.tamu.edu

ETRA 2010, Austin, TX, March 22 – 24, 2010.
© 2010 ACM 978-1-60558-994-7/10/0003 $10.00

remains limited in its ability to facilitate believable creature locomotion because the rig is undiscriminating in regards to the relationship between how parts move and their importance to the perception of action. An experienced animator will likely recognize that when creating a walking motion defining foot placement relative to hip position is a priority over defining the relationship of the elbow to the shoulder. The animator's decision is based upon experience with not only observation of motion but viewer reaction to motion. To animate locomotion, an experienced animator will first define the relationship of the feet to the ground plane, then the feet to the hips. Only when those relationships are working well will the artist's attention turn to the movement of the upper torso. Thus intimating that there is a hierarchy in the value of moving parts as contributors to a character's motion.

To build an animation system for digital creatures that is informed by how we perceive motion, we must first understand how viewers interpret what they see and where they find this information within an image. Despite ever-increasing computing speed, a smart animation system must be computationally fast and intuitive to the artist. Therefore, before considering the engineering perspective of the problem, we must first determine what kinds of information can be communicated from the display of motion to the viewer from minimal visual data.

2 Previous Work

When approaching the problem of creating an animation system that is based upon the perception of biological motion, we are assisted by the fact that if only the joint pivot locations of a digital creature are rendered the resulting moving images are visually equivalent to the presentation of PLDs. By attaching small objects to joint pivot locations and creating high visual contrast relative to other elements of a scene Johansson [1973] developed a method for isolating motion from form as a collection of particles, now commonly known as PLDs. His methodology expanded early work on the coherence of motion patterns. Johansson's and subsequent studies demonstrated that PLDs of human actors in motion, though presenting significantly minimized visual information, carry all of the information necessary for the visual identification of human motion. Mather [1993] demonstrated that viewers are capable of identifying animal species in PLDs created Muybridge's photographic motion studies [1979] thus extending the range of perception of biological motion beyond the human figure.

Collections of moving points, such as PLDs, may exhibit coordinated or non-coordinated motion. Coordinated motion includes rigid and non-rigid relationships between a point and its neighbors. Manipulating this minimal information can even affect the perceived gender of PLD walkers. For example, exaggerating the movement of points representing the hips and shoulders can bias gender recognition [Cutting et al. 1978]. If the collections of points are based upon human or animal form but the presentation is inverted or some elements are time-phase shifted, then recognition is impeded. This suggests that perception of biological motion is dependent upon both the relationship of the points in motion to one another, and to the environment i.e. the relative velocities of leading and trailing points in a perceived structure and the reaction of the points' gravity [Shipley 2003].

As discussed earlier, an experienced animator recognizes the varying levels of importance of moving features on a digital creature. Visual perception and cognition studies correlate with this effect. By displacing portions of PLD of walking humans, cats, and pigeons, Troje [2006] showed that the local motion of the feet contained the key information leading to recognition of direction of travel. Thurman [2008] used "bubble" masks to randomly obscure portions of the PLDs of walking figures and showed that large scale relationships, such as configurations of points revealing body posture, are much less informative than localized actions such as the motion of the extremities (hands and feet).

Rather than using masking or displacement, we present an experiment that employs an eye-tracking to examine what regions of the video draw the viewer's gaze when viewing both standard video and PLDs of walking animals. We then compare this information to the viewer's success in identifying animal species and characteristics from the displays. Through this method, we aim to determine about the minimal information necessary to convey certain characteristics, and conversely, the characteristics that are capable of being communicated through minimal spatio-temporal information.

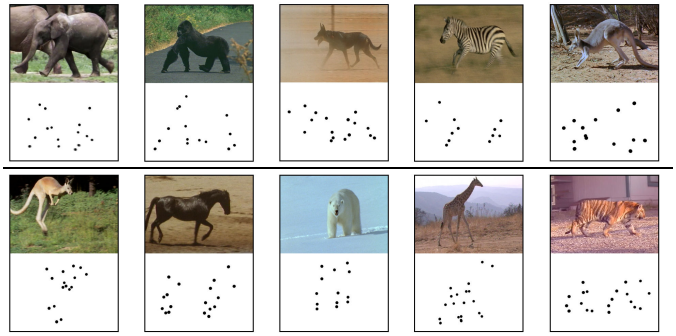

Figure 2: *This figure represents each of the stimuli used in the experiment. A still from the full resolution video is shown above while the PLD is shown below.*

3 Experiment

Our experimental design was inspired by Mather [1993]. He used PLDs which were rotoscoped using scanned photographs from Muybridge's study of animal locomotion. Like Mather, we used a list of animal species during the animal-identification task. Rather than generate PLDs from a sequence of photographs we used animal motion video as the basis for our stimuli and designed a more rigorous method for dot placement, as further explained in Section 3.1.

The stimuli for the experiment included video of 10 animals that were shown in both full video and a PLD representation. The resulting 20 short video clips, ranging in length from 2 to 10 seconds, contained at least one full gait cycle (see Figure 2). Participants first completed a short training session consisting of a single trial with meaningless data to familiarize participants with the experiment protocol. Each participant viewed the set of 10 PLD representations first, followed by the set of full videos. After viewing each video, participants were asked to select one characteristic from each pair shown in Table 1 that best described the video. They were then asked to identify the animal from the list of animals in Table 1. PLDs were presented first to eliminate any learning effects and bias that could have resulted from viewing full video first. A short break occurred between the sets of videos. Presentation within type was randomized to eliminate any learning effects. Each video clip measured 720 X 480 pixels. Video size was smaller than the viewing screen, and was therefore centered on a black background.

Ten participants took part in the study. All had normal or corrected-to-normal vision and were naive as to the purpose of the experiment. Participants were seated 75 cm in front of a 22 inch LCD monitor in a well lit room. Using a remote infrared camera-based

eye-tracking system[1], data pertaining to fixation position and saccades were recorded. A fixation is defined as any pause in gaze $\geq 150ms$. Participants were instructed to remain as still as possible during calibration. However, the equipment used to track the gaze is robust enough to quickly recalibrate the subject on reentering the view.

3.1 Preparing Stimuli

Sequences representing the motion of animals were gathered from a variety of sources. The primary source was Absolutely Wild Visuals [AbsolutelyWildVisuals 2009], a company specializing in stock footage of wild animals. Both stationery and moving cameras positions were included. The size of the animals and their PLD representations on screen varied from one-third to two-thirds screen height. All motion, apart from one sequence, was orthogonal to the screen.

For each sequence, major joint pivot locations of the animal were identified using skeletal reference material from [Feher, 1996] and on-line sources. The selection of important joints was driven by [Cutting et al. 1978] and included marking the head, spine, shoulder, hip, knee, ankle, wrist, and toe.

When specifying animal motion from a single 2D reference, a key challenge is to preserve joint lengths despite changes in spatial depth of the figure relative to the camera. Doing so successfully is crucial to the representation of rigid and non-rigid relationships between biological structures. To correctly handle this situation, we built a simple animation rig and relied on fixed joint lengths to preserve rigid connections (Figure 3). Distance from the hip to the knee and from the knee to the ankle, and articulating chains of joints to permit non-rigid relationships, such as the distance between the hip pivot and the shoulder pivot were preserved.

A flatly rendered sphere, the joint locator, was placed at each joint pivot point. Autodesk's Maya 2008 [Autodesk 2009] is an industry standard for 3D computer animation and was used to define the rigs and create the PLD animation. A 3D approach to point placement was preferred over 2D to allow us to account for depth and to render with 3D motion blur. Full video sequences were imported as an image plane and used for rotoscoping the motion. The rig was built on the image plane to maintain proportion with the footage. By working frame by frame, we determined the best size and fit for the rig over the entire video sequence.

For each frame, only those joint pivot points that were visible to the viewer in the original sequence were rendered in the PLDs. If the pivot point was occluded by the animal's body, such as a far leg passing behind a near leg, or occluded by objects in the environment such as grass, snow, or water, the pivots locators were not displayed. Visibility was managed on a frame-by-frame basis.

PLD sequences were presented in high contrast as black dots on a white background, maintaining the natural outdoor familiarity of darker figures against a lighter environment. Joint locator size relative to screen size was modulated by scaling the sphere based upon the figure's distance from camera. After the PLDs were completed for all 10 videos, we took the average dot size and scaled each figure to normalize the dot size across all videos. Since the physical size of the animals in the source footage varied widely this method maintained a consistency across the stimuli that was irrespective of the animal's size. We also animated the camera rendering the scene in order to stabilize the footage and remove any camera translation apparent in the original footage, allowing each clip to appear as if the animal is walking on a treadmill.

[1] faceLAB® by Seeing Machines, Inc.

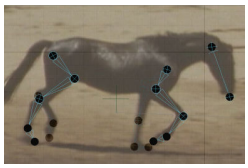

Figure 3: *A basic rig was constructed to preserve rigid joint relationships in the PLDs generated for use as stimuli.*

4 Results

4.1 Characteristics & Species Identification

Heavy	Light	Cat	Dog
Predator	Prey	Horse	Deer
Old	Young	Ape	Giraffe
Large	Small	Kangaroo	Rat
Furry	Hairless	Fox	Elephant
Tailed	Tailess	Zebra	Tiger
Strong	Weak	Raccoon	Bear

Table 1: *The list of characteristics (left) and animals (right) presented to participants after each video segment. Participants selected characteristics which best described the stimuli just viewed.*

In the full resolution video, as expected, participants correctly identified the animal 100% of the time. It is interesting to note that 25% of the time people could identify the animal correctly just from the PLD representation. However, some animals proved more recognizable in sparse form than others. The kangaroo, for example, was correctly identified by over half of the participants, whereas all 10 failed to recognize the bear (not side facing) or the elephant. While the orientation of the bear might cause this result, it is unclear why the elephant information does not convey the signature motion for the animal. It may be that the elephant shares enough structural similarities with other animals on the identification list that the motion patterns were dynamically similar. This effect can be seen in the results for the horse. Only three of the 10 participants correctly identified the horse, but if the other ungulate responses (deer, giraffe, and zebra) are included, the response improves to 70% correct.

In analyzing the trait responses, we treated the viewers' responses to the full videos as correct. Only traits with a consensus from the full view were analyzed. Some traits achieved a consensus across both video sets; for instance, the tiger was described as a strong predator in both the PLD and the full view. However, a consensus described the PLD polar bear as prey while also describing the full view polar bear as a predator.

4.2 Eye Tracking

Participant's eye movements were recorded as they performed the task in both the PLD and the full resolution video. In order to compare gaze patterns across the two stimuli, video were first segmented into individual frames. The centroid of each point in the PLD was chosen as the target region. Fixations were compared against each target region. The number of target regions varied across each animal, depending on whether the target region was visible in the full video. To reveal differences across both stimuli, fixations that occur (only) in the target regions of both sets of videos were accumulated. The percentage fixation time within target regions for each sequence is shown in Table 2. As can be seen from the table, the time spent fixating in the PLD videos is higher than in the full view videos. This is to be expected due to the high-contrast, sparse nature of the

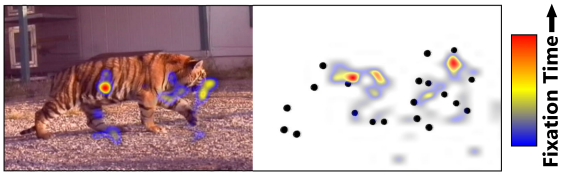

Figure 4: *Gaze Pattern across entire sequence for the Tiger sequence, PLD (left), Full(right). Even though the time spend fixating on each region is different, the pattern of gaze over each image is not significantly different. This holds true for the image sequences.*

PLD videos. The correlation between the time fixating in the 'dot regions' on the PLD when compared to the same regions in the full video is rather high at 92%. This means that despite the difference in total time fixating on the target regions, there is a definite pattern that emerges in both the PLD and full resolution videos. As can be seen in Figure 5, fixations are overlaid on the target regions (PLD centroids) for all frames of both video representations.

IMAGE	PLD	FULL	IMAGE	PLD	FULL
Bear	61%	12%	Dog	75%	56%
Elephant	40%	28%	Giraffe	94%	62%
Ape	17%	28%	Horse	50%	24%
KangaHop	21%	17%	KangaWalk	35%	37%
Tiger	49%	49%	Zebra	20%	34%

Table 2: *This table shows the percentage of fixations occurring within target regions. As expected, this number is higher for the sparse PLD representation. However, the corresponding areas in the full resolution images also show a higher number of fixations on target as opposed to non-target regions.*

Gaze is drawn toward similar regions in the full resolution video as in the PLD representation. Figure 5 shows overall correlation values of the percentage time fixating in target regions between each pair of videos. There is good agreement for most of the image sequences. A t-test performed on the percentage time fixating on target regions showed no significant difference between video sequence and PLD pairs, $t(9) = 1.8132, p < .05$, giving further confidence of similarity between gaze distribution over both sets of data.

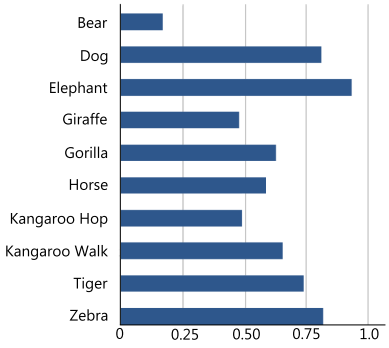

Figure 5: *This graph charts the correlation between fixations on the PLD and the corresponding fixations on the same locations in the full resolution video for each animal. While percentage times are higher in the PLD videos, correlations between the two sets of video reveal similarities in gaze patterns across presentation method.*

5 Conclusion & Future Work

To date, few studies have focused on the value of perception research as it applies to the generation of computer animation, particularly biological motion. While there is strong interest in the computer graphics community in creating the next generation of animation tools, there is no single consensus about how to best approach the subject. This project offers an initial approach that has the potential to satisfy the conditions of maximizing the visual result while minimizing the required input for a new system. Our investigation represents an initial step toward understanding what information is communicated by animal PLDs and how this minimal data can be transformed into useful information and applied for other uses. The results from this preliminary study are promising and several follow up experiments are imminent. In future, we plan to include stimuli displaying a wider range of motions and focus on the detection of traits across larger groups of animals as opposed to individual species.

References

ABSOLUTELYWILDVISUALS, 2009. http://www.absolutelywildvisuals.com.

AUTODESK, 2009. Maya.

CUTTING, J., PROFFITT, D., AND KOZLOWSKI, L. 1978. A biomechanical invariant for gait perception. *J. Experimental Psychology: Human Perception and Performance 4*, 3, 357–372.

E, M. 1979. *Muybridge's Complete Human and Animal Locomotion*. Dover.

FACELAB, 2009. facelab eyetracking system, http://www.seeingmachine.com.

FAHER, G., AND SZUNYOGHY, A. 1996. *Cyclopedia Anatomicae*. Black Dog and Leventhal Publishers.

JOHANSSON, G. 1973. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics 14*, 2, 201–211.

JOHNSTON, O., AND THOMAS, F. 1981. *The Illusion of Life: Disney Animation*. Hyperion.

LASSETER, J. 1987. Principles of traditional animation applied to 3d computer animation. In *SIGGRAPH '87: Proceedings of the 14th International Conference on Computer Graphics and Interactive Techniques*, ACM SIGGRAPH, 35–44.

MATHER, G. 1993. Recognition of animal locomotion from dynamic point light displays. *Perception 22*, 7, 759–766.

PYLES, J. A., GARCIA, J. O., HOFFMAN, D. D., AND GROSSMAN, E. D. 2007. Visual perception and neural correlates of novel 'biological motion'. *Vision Research 47*, 21 (Sept.), 2786–2797.

SHIPLEY, T. F. 2003. The effect of object and event orientation on perception of biological motion. *Psychological Science 14*, 377–380.

THURMAN, S. M., AND GROSSMAN, E. D. 2008. Temporal "bubbles" reveal key features for point-light biological motion perception. *Journal of Vision 8*, 3, 1–11.

TROJE, N. F., AND WESTHOFF, C. 2006. The inversion effect in biological motion perception: evidence for a life detector. *Current Biology 16*, 8 (Apr.), 821–824.