Perception of limited-lifetime biological motion from different viewpoints

Simone Kuhlmann   Marc H. E. de Lussanet
Markus Lappe

Psychologisches Institut II, Westf. Wilhelms-Universität Münster, Germany
*To whom correspondence should be addressed; E-mail: kuhlmans@psy.uni-muenster.de.

Studies with time-limited point-lights suggested that biological motion does not require local motion detection. These studies used walkers seen from the side, but biological motion perception excels also when walkers are oriented towards the observer, or in intermediate, half-profile views. In perspective projection, the local motion of points on the body provides a cue to the 3D structure of the walker. Thus, local point motion that was irrelevant for walkers in profile view may become important for biological motion perception in perspective projection. We compared performance on forward/backward walking discrimination of walkers in orthographic and perspective projection when view orientations and with point lifetime was varied. We found no difference between orthographic and perspective projections. Walkers with point lifetime 1 allowed forward-backward discrimination reliably in non-profile views, suggesting that local image motion is not required. Discrimination performance became extremely difficult in the frontal view, however. Follow-up experiments that tested lifetime, view orientation, and specific information from the feet indicated that this dependence on viewing angle can be explained by the reliance of the forward/backward discrimination on information about the movement of the lower legs, which is difficult to ascertain in the frontal view.

1   Introduction

Our visual system is highly sensitive to the movement patterns of other living creatures. This ability is so well developed that we obtain an immediate, vivid percept of a walking human, already from seeing just a few points attached to the joints of a moving body (Johansson, 1973). Point-light displays contain both form and motion information. Each point at each time provides position information about a single spot on the body. Integration of the positions of multiple points, either per frame or over time, yields form information about the configuration of the body. At the same time, the temporal evolution of the positions of each single point provides local motion, acceleration, and trajectory information for that point.

The limited-lifetime technique can be used to investigate the contributions of motion, acceleration, and trajectory of individual points, while leaving global form intact. In limited-lifetime stimuli each single point is shown only for a limited number of successive image frames, after which it is extinguished. The number of frames that a point lives determines whether this point offers motion, acceleration, or trajectory information to the viewer. If the lifetime is limited to only a single frame the point cannot offer motion information because it is not moving with the limb between frames. The minimum lifetime for motion is two frames because then apparent motion sensors can be activated. A higher lifetime may improve the local motion sensing by spatio-temporal integration. If the point moves in a straight line the motion measurement will become more robust. If the point moves along a curved trajectory, on the other hand, simple spatio-temporal integration would introduce errors since the motion direction is changing between each pair of frames. Lifetimes longer than two frames also offer acceleration information, i.e, how the local motion changes over time. Lastly, the longer the lifetime the more information about the trajectory of the point is available. The trajectory is the curve in space that the point traverses over time and is independent of direction or speed of the motion of the

2

point. The trajectory cannot be calculated at any moment in time but is a shape that must be estimated from observing the positions of a point over time.

The limited-lifetime technique was first applied to biological motion by Neri et al. (1998) who used a lifetime of two frames in a walker with only six points placed randomly on the main joints of the body. Beintema & Lappe (2002) examined the role of local motion and global form with limited-lifetime walkers in which the individual points appear at random locations on the limbs of the body. Local motion of these points was manipulated by limiting the lifetime of the points, i.e., the number of frames that a point moves along with a single spot on the body. When lifetime was reduced to one frame only, no local motion information was present because the points did not follow the movement of the body. Nevertheless, naive observers spontaneously recognized these animations as human walkers (Beintema & Lappe, 2002) and could reliably judge the facing direction and the coherency of a walker, as well as discriminate between forward and backward walking (Beintema et al., 2006). Thus, local motion was not necessary for these tasks.

Lange et al. (Lange & Lappe, 2006; Lange et al., 2006) have suggested that a template matching analysis of the body configuration may underlie biological motion recognition. In this model, the positions of points in each stimulus frame are matched to templates of the human body in different postures. Local image motion from individual points is not used. The motion of the body is derived from analyzing the evolution of the best-matching body postures over time.

Thus, from experimental observations and computational considerations local image motion does not appear necessary for biological motion analysis. However, experiments that used the limited-lifetime technique have so far only used profile views of walking in orthographic projection (Figure 1 A). It is important to test the usage of local image motion in other view orientations and in perspective projection because the combination of profile view and orthographic projection is a special case for two reasons.

The first reason is the difference between orthographic and perspective projection. In or-

thographic projection, a point $P = (X, Y, Z)^T$ on the body is projected onto a point $p_{orth} = (x, y)^T$ in the image so that

$$p_{orth} = \begin{pmatrix} X \\ Y \end{pmatrix}. \tag{1}$$

Here, the projection is without loss of generality assumed to be along the Z-axis. Image coordinates (x, y) directly correspond to world coordinates (X, Y), and the depth coordinate Z is lost in the projection. The image motion v_{orth} of image point p_{orth} is

$$v_{orth} = \frac{d}{dt} p_{orth} = \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix}. \tag{2}$$

Therefore, the image motion is independent of the motion-in depth, $\dot{Z}$, of point P along the Z-axis. Any information about the motion-in depth component of the point on the body therefore has to be gleaned from the motion along the X- and Y-axes. This requires knowledge of the structure of the human body, as, for instance, provided by a template of the body. The visual information in image point positions and image point motions is mathematically insufficient to estimate body posture and movement (Ullman, 1984) and perceptual recognition can only be achieved when additional assumptions about the structure or movement of the body are introduced. This can be done either by assuming explicit body models (Rashid, 1980; Marr, 1982; Chen & Lee, 1992; Aggarwal & Cai, 1999) or biomechanical constraints on the body motions (Webb & Aggarwal, 1982; Hoffman & Flinchbaugh, 1982).

The mathematical insufficiency of the visual position and local image motion signals for biological motion recognition also holds for perspective projection. However, unlike in orthographic projection, the position and image motion signals in perspective projection contain information about the Z (depth) component of the walker. In perspective projection, point P is projected onto p_{persp} so that

$$p_{persp} = f \frac{1}{Z} \begin{pmatrix} X \\ Y \end{pmatrix}, \tag{3}$$

where f is the focal length of the projection. The image motion v_{persp} of image point p_{persp} is

$$v_{persp} = \frac{d}{dt} p_{persp} = f \frac{1}{Z^2} \begin{pmatrix} \dot{X}Z - X\dot{Z} \\ \dot{Y}Z - Y\dot{Z} \end{pmatrix} = f \frac{1}{Z} \begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} + f \frac{\dot{Z}}{Z^2} \begin{pmatrix} X \\ Y \end{pmatrix}. \tag{4}$$

4

Therefore, the image motion in perspective projection consist of a part that is specified by the motion $(\dot{X}, \dot{Y})^T$ of P in X and Y directions and a part that is specified by the motion-in-depth, $\dot{Z}$.

The comparison of the two projections shows that in orthographic projection all information in both the positions p_{orth} of image points and the motions v_{orth} of image points is related only to the X and Y coordinates of the body. Information about the Z component of the body structure and its motion is missing from the stimulus and can only be reconstructed by using external knowledge of the body structure. In perspective projection, on the other hand, both the positions p_{persp} of image points and the motions v_{orth} of image points carry information about the depth Z. Most importantly for our investigation, p_{persp} and v_{persp} carry independent information about depth, because p_{persp} depends on Z, i.e., the position in depth of point P, and v_{persp} depends on Z and also on $\dot{Z}$, i.e., the motion of P in depth. Therefore, in perspective projection the local image motion of a point may convey information over and above the information conveyed by the point positions. Hence we must ask, whether local image motion, which has previously been shown to not contribute to perception in the orthographic projection, will contribute in the case of perspective projection.

The second reason why the combination of profile view and orthographic projection is a special case has to do with the shape and limb movement of the walker. In profile view, the movement of the limbs is almost exclusively in parallel to the image plane. Since there is little motion along the depth axis, the lack of information about Z-axis motion in the orthographic projection is of no influence. In fact, in orthographic projection in the profile view, the depth distribution of the light points of the stimulus is entirely ambiguous and the stimulus is mathematically indistinguishable from
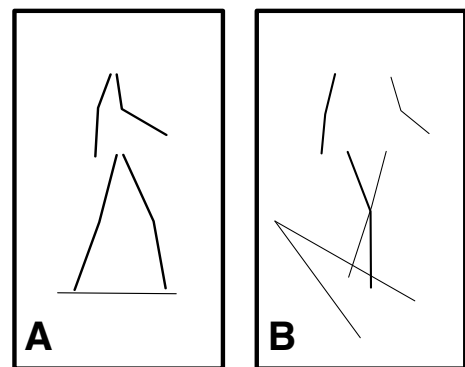
Figure 1: A: 2D orthographic projection, profile view and B: 3D perspective projection, half-profile view

a flat arrangement of light points in a single depth plane. For a template matching recognition procedure it would be sufficient to match the stimulus frames to two-dimensional templates. The true three-dimensional structure of the body becomes visually more apparent when the walker is shown in other view orientations and in perspective projection. For instance in the half-profile view (Figure 1 B), the movement of the limbs is directed in depth, and, because of the perspective projection, the visual speed of the limb movement gets smaller when the limb is further away then when it is closer to the observer. Thus, in these stimuli, visual speed is an independent cue to distance and hence to the three-dimensional structure of the stimulus.

In perspective projection, visual speed is also informative about the depth structure of the walker in profile view. Consider, for example, the movement of the shoulders. The shoulder nearer to the observer will move faster than the shoulder further from the observer. Thus, in perspective projection the visual motion of points on the body provides a cue to the 3D structure of the walker. In orthographic projection, the speed of point movement is independent of the distance to the observer.

Limited-lifetime experiments with walkers in profile view in orthographic projection showed no influence of local point motion on biological motion perception. However, in perspective projection, and in view orientations other than the profile view, local point motion carries information about the 3D structure of the walker. Thus, local point motion that is irrelevant in orthographic projection may become important for biological motion perception in perspective projection. We wanted to test whether this is the case.

From Johanssons demonstrations and a number of further studies (Verfaillie, 1993; Mather & Murdoch, 1994; Troje et al., 2005) it is known that observers not only readily recognize profile views of point-light walkers, but also point-light walkers seen in other view orientations. In this case, point-light actions convey a strong impression of depth even if static low-level depth cues are missing (Vanrie et al., 2004). The depth percept conveyed by a point-light walker even dominates over conflicting disparity depth cues (Bülthoff et al., 1998). It is possible that local motion information, which is not necessary in the profile view, aids the depth perception

6

process in other views by exploiting the relationship between speed and depth in the light point motion (Ullman, 1984). On the other hand, depth perception of 3D walkers could also be achieved by template-matching without exploiting local motion signals. Such template matching could either use 2D templates for particular viewpoints or full 3D representations of the walker.

In the present study, we used 3D limited-lifetime walkers to investigate the role of local motion in the perception of biological motion for the case of differently oriented 3D walkers. We asked observers to discriminate between a display of a forward walking figure and the same display in reversed order (similar to backward walking). In profile view this task is easy even with lifetime 1, so that it does not require local image motion (Beintema et al., 2006). We were interested whether this also holds true for other viewing angles. Specifically, as described above, image motion signals might convey information about the motion-in-depth of a point. If this is indeed the case, we would expect a difference in performance for non-profile views between orthographic and perspective projection. Moreover, if local point motion is important for biological motion perception in non-orthographic views, we would expect to find an advantage for lifetime 2 over lifetime 1 in perspective projection.

2 General methods

2.1 Subjects

Seven subjects (24-35 years, 3 females) participated in the experiments. All of them were experienced with psychophysical experiments involving biological motion stimuli. Apart from authors SK and MdL, the participants were naive to the objective of the experiments.

2.2 Stimuli

Stimuli displayed walking human figures which consisted of white points (0.15 x 0.15 deg) on a black background. Width and height of the stimulus subtended approximately 5 x 9 degree visual angle. The stimuli were based on the 3D joint positions of nine walking humans (5 male

and 4 female) recorded using MotionStar Wireless (Ascension Technology Corp., Burlington, USA). The forward translation was subtracted giving the impression of walking on a treadmill. Walking speed was normalized so that a complete walking cycle, consisting of two steps, took about 1.4 seconds. The stimulus sequence was either presented in normal (forward walking) or reversed (backward walking) frame order. The walker started from a random phase in the step-cycle and was shown for one complete walking cycle of 1.4 s. All walkers were presented in perspective and orthographic projection.

For the limited-lifetime walkers the points on the walker were assigned a random position on one of 8 limb segments (upper and lower parts of the arms and legs). The possible positions were distributed uniformly across the segments, each segment defined by the line connecting joints. The lifetime of a point, defined as the number of frames before the point was relocated to another location on the body, could be varied. Relocating the points to a new random location on the limbs after a limited number of frames disturbs the continuous motion and removes the local motion information (motion vector and trajectory information) carried by each point, without altering the temporal sampling of the sequence. The points were relocated in an asynchronous fashion.

The total number of points per trial is an important parameter for the performance (Beintema et al., 2006). It is calculated by multiplying the number of points per frame with the number of frames seen in the trial. This calculation is independent of whether the points stay on the same limb position over successive trials or not, since in both cases each frame provides a certain number of points that signal the current posture. For example, a stimulus with 4 points per frame provides over 8 frames a total of 32 points no matter if the lifetime is 1 or 8. In the former case, new point positions on the limbs are chosen in each frame. In the latter case, the same point locations on the limbs are used in each frame but because the body posture changes over those 8 frames each point provides new body posture information over the last frame. Thus, the total amount of body posture signals is the same in both cases but the latter condition, in addition, provides local motion and trajectory signals of each point. We

used conditions with either 128, 512, or 384 points per trial.

Depending on the experiment, different combinations of the following stimulus conditions were used. Limited-lifetime walkers had either two, four or twelve points per frame. For the two points per frame condition the total number of points per trial amounts to either 128 points with a frame duration of 22.2 ms or 512 points with a frame duration of 5.56 ms (i.e. 4 or 1 multiples of the 180 Hz at which the monitor displayed). For the four points per frame condition the total number of points per trial amounts to either 128 points with a frame duration of 44.4 ms or 512 points with a frame duration of 11.1 ms (i.e. 8 or 2 multiples of the 180 Hz at which the monitor displayed). For the twelve points per frame condition the total number of points per trial amounts to 384 points with a frame duration of 44.4 ms (i.e. 8 multiples of the 180 Hz at which the monitor displayed).

It is important to note that because of visible persistence the apparent number of simultaneously present points on the screen was higher than the number of points presented in each frame. Visible persistence describes the apparent duration of a point that is briefly flashed. It has been shown that brief flashes of light, such as the points that were presented for durations between 5.56 ms and 44.4 ms, remain visible for longer temporal intervals, up to 100 or 200 ms (Brown et al., 1974; Coltheart, 1980). Therefore, the stimuli appeared to consist of more than the 2 or 4 points which they physically contained. It is not known at what level of the visual pathway visible persistence is created, or whether it contributes to form recognition. For our analysis we focus on the number of points that are physically provided in each frame since this is the source of information present in the stimulus.

As a further stimulus condition, classic Johansson walkers were used which consisted of 12 light-points, displaying the joints of shoulders, elbows, wrists, hips, knees and ankles. In the last experiment a modification of the classic walker was used, where the foot-point could be positioned at different locations on the lower limb (a more detailed description can be found in the method section of experiment 4).

9

2.3 Procedure/ Experimental set up

Stimuli were displayed on an Iiyama Vision Master CRT Monitor (40 x 30 cm, 800 x 600 pixel) at a vertical refresh-rate of 180 Hz. The subjects were seated in a darkened room with their eyes about 70 cm in front of the monitor. They were asked to fixate a red fixation point in the middle of the screen. The walkers were presented in the center of the screen. Walkers measured 5 x 9 degrees of visual angles. When the walker disappeared the subjects had to press a response key. Thereafter a new trial started and a new walker appeared after 200 ms. The subjects task was to detect the walking direction (forward/backward) of the walker; pressing the 'up' (forward) and 'down' (backward) arrow keys of the keyboard.

2.4 Data analysis

The proportion of correct responses was assessed. T-tests or repeated measures analysis of variance (significance level = 0.05) on the d' values were conducted for statistical testing. The Scheffé-test was used as a posteriori procedure. For all post hoc tests an alpha significance level of 0.05 was used. Error bars in the figures give the standard error of the mean.

3 Experiment 1

In the first experiment we asked whether local motion information can improve performance on a forward/backward discrimination in perspective projection, when the walkers are presented in different view orientations.

3.1 Methods

Stimuli were limited-lifetime walkers with points lifetime one (no local motion signals) or two (with local motion signals) frames. We also varied the number of points per frame and the total number of points per trial, since these are parameters that are known to influence the performance in the profile view (Beintema et al., 2006). The limited-lifetime walkers had either two or four points per frame. A further stimulus was the classic Johansson walker, which

consisted of 12 light-points, displaying the joints of shoulders, elbows, wrists, hips, knees and ankles.

All walkers, the limited-lifetime walkers as well as the classic walkers were presented in perspective projection as well as in orthographic projection. Both walker types were also presented in three orientations, the profile view (0°), the half profile view (45°), and the frontal view (90°).

Each experiment session had 96 different conditions for the limited-lifetime walkers (2 lifetime x 2 points per frame x 2 points per trial x 2 play-directions x 2 projections x 3 orientations) and 12 for the classic walker (2 play-directions x 2 projections x 3 orientations) with 9 repetitions for each condition. One experiment session consisted therefore of 972 trials and took about 20 minutes. In each session all limited-lifetime walker and classic walker conditions were presented in randomized order. Each subject conducted three experiment session. The task was to detect whether the walkers walk forwards or backwards. The answers were giving by the 'up' and 'down' arrow keys of the keyboard.

3.2   Results
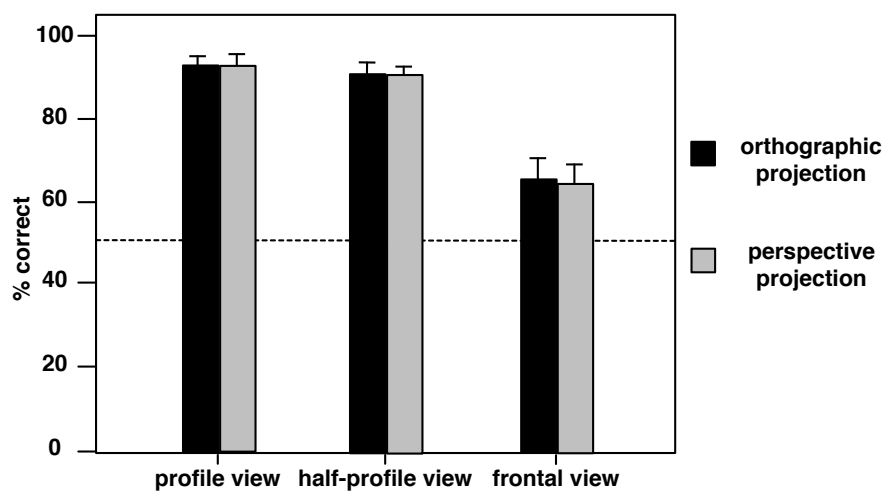
Figure 2: Experiment 1: Comparison of discrimination performance in orthographic and perspective projection.

Whereas orthographic projections do not contain direct information about the structure

11

and motion in depth, the perspective projection does. We compared the performances for the orthographic and perspective projections for the different views (Figure 2). For both cases we found equally good performance for profile and half-profile view and poor performance for the frontal view. To test for statistical differences we conducted a 2-way ANOVA on the factors projection and viewing angle (2 x 2 design with repetition on subjects). The main effect of viewing angle was statistically significant ($F_{(1,6)} = 10.3$, $p< 0.01$). Importantly, there was no significant difference between the two projection types ($F_{(1,6)} = 663.8$, $p= 0.27$). This indicates that the participants had no advantage of the perspective projection. Theoretically, the perspective projection also yields information about the motion-in-depth of the individual points, as explained in the Introduction. To test this possibility directly, we performed a more detailed analysis of the lifetime conditions.

The results for perspective projection in profile view are illustrated in Figure 3 A. In all conditions the performance with a lifetime of one frame was as good as with two frames. Performance was generally better with 512 points per trial than with 128 points per trial. With 512 points per trial the performance for the limited-lifetime walkers was as good as the performance for the classic walker independent of the number of points per frame. The results for the half-profile view were similar to the results for the profile view (Figure 3 B). A higher lifetime had no positive effect on the performance, and performance for 512 points per trial approached that of the classic walker.

To test for statistical differences between profile and half-profile view within the limited-lifetime conditions we conducted a 3-way repeated measures ANOVA with the factors lifetime, points per frame and viewing angle (2 x 2 x 2 design with repetition on subjects). To exclude ceiling effects from the testing, we only included the 128 points per trial conditions for the profile and half-profile view. There were significant main effects of viewing angle ($F_{(1,6)} = 10.3$, $p= 0.02$), and points per frame ($F_{(1,6)} = 22.8$, $p= 0.003$), but not for lifetime ($F_{(1,6)} = 1.9$, $p= 0.2$). There was a significant interaction between points per frame and lifetime ($F_{(1,6)} = 16.5$, $p= 0.007$). No other interactions were significant.
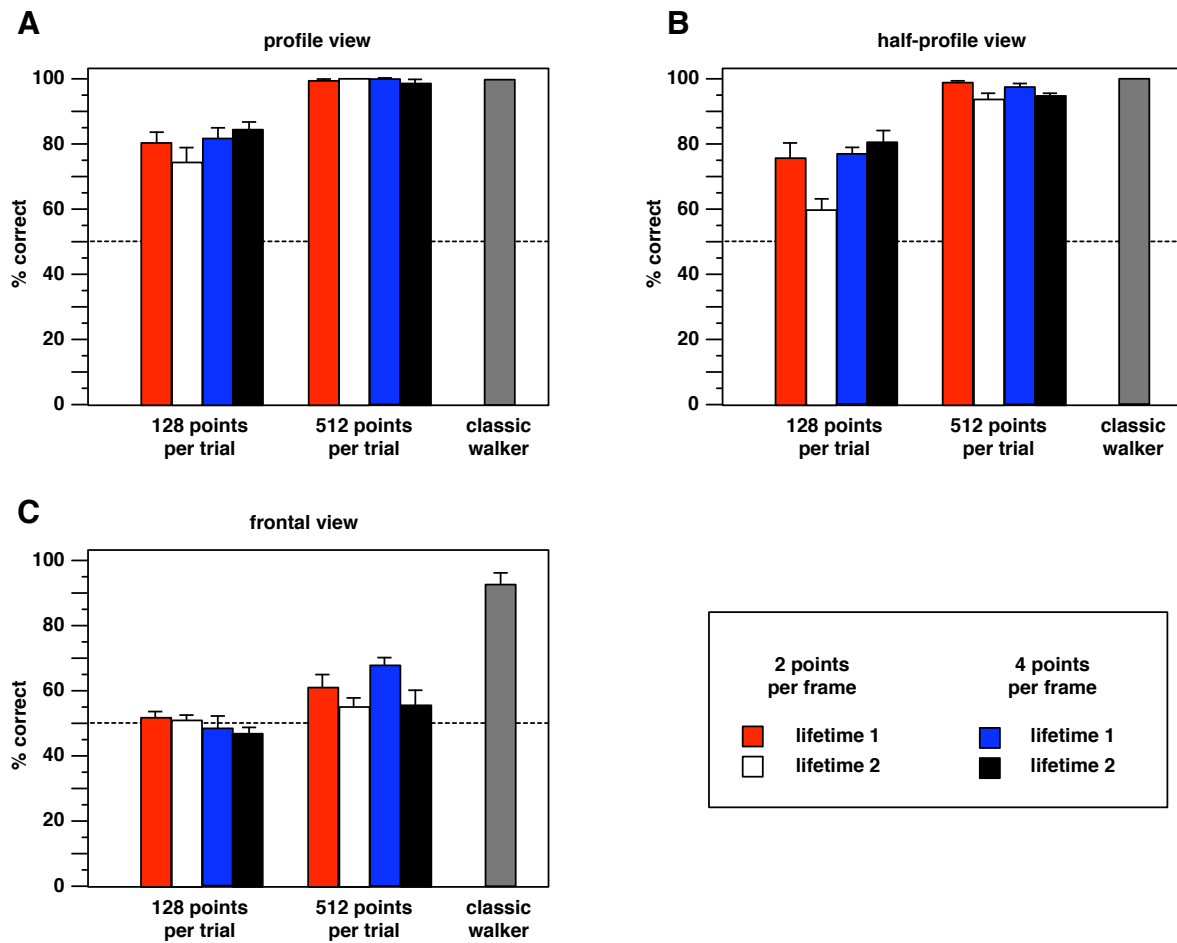
Figure 3: Experiment 1: Averaged percentage of correct responses for the profile, half-profile and frontal view in perspective projection. Data of the limited-lifetime walkers are split by the factors points per trial, points per frame and lifetime. Results of the classic walker are added for comparison. Error bars represent the standard error over subjects.

For the frontal view (see Figure 3 C) the performance for limited-lifetime walkers with 128 points per trial was not different from chance level (t-test; $p > 0.05$). In the two conditions with 512 points per trial and lifetime 1 the performance was significantly above chance level (t-test; for 2 points per frame $p = 0.01$ and for 4 points per frame $p = 0.004$) but still worse than the performance for the classic walker.

For comparison, the results for orthographic projection are shown in Figure 4. Results were very similar to those of the perspective projection, consistent with the overall analysis provided in Figure 2. For the profile and half-profile views, the performance with a lifetime of two
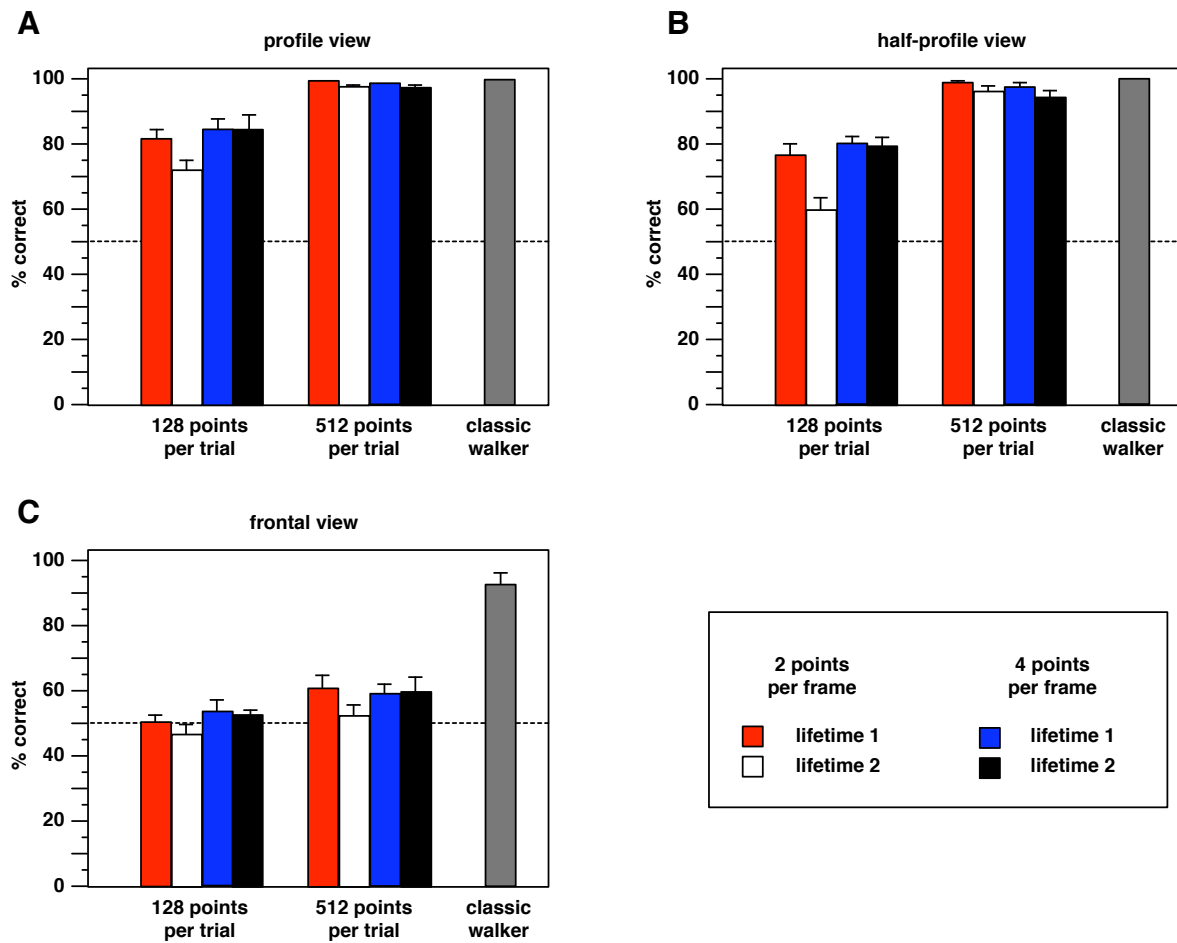
Figure 4: Experiment 1: Averaged percentage of correct responses for the profile, half-profile and frontal view in orthographic projection. Data of the limited-lifetime walkers are split by the factors points per trial, points per frame and lifetime. Results of the classic walker are added for comparison. Error bars represent the standard error over subjects.

frames was in no condition better than that with one frame. Performance was generally better with 512 than with 128 points per trial and approached that of the classic walker, independent of the number of points per frame. A 3-way repeated measures ANOVA with the factors lifetime, points per frame and viewing angle (2 x 2 x 2 design with repetition on subjects) on the 128 points per trial condition showed significant main effects of viewing angle ($F_{(1,6)} = 15.5$, $p = 0.008$), points per frame ($F_{(1,6)} = 19.7$, $p = 0.004$) and lifetime (lower performance with lifetime two, $F_{(1,6)} = 12.5$, $p = 0.01$) and no significant interactions. For the frontal view (Figure 4 C) the performance walkers with 128 points per trial was not different from chance

level (t-test; p > 0.05). The performance for the 512 points per trial conditions was lower than the performance for the classic walker, but significantly above chance level (t-test; p < 0.01), except for the condition with 2 points per frame and lifetime 2.
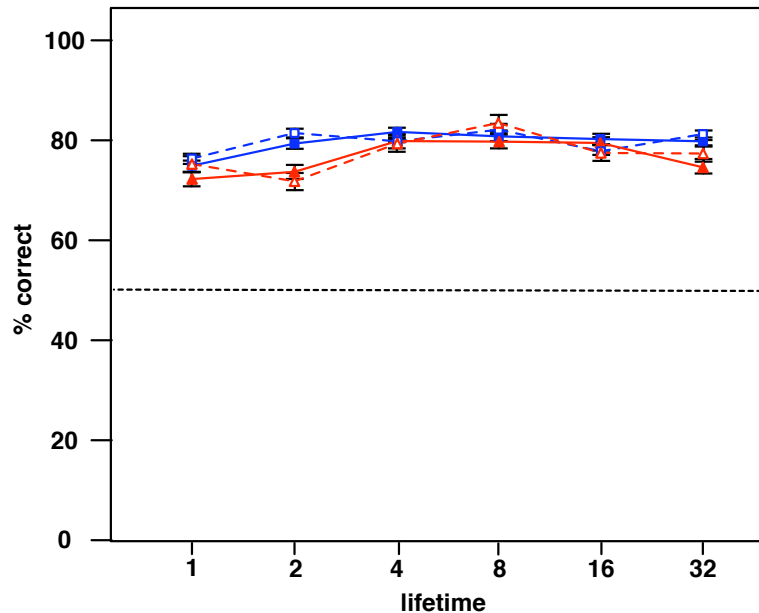
Figure 5: Experiment 1B: Averaged percentage of correct responses as a function of point lifetime. Data are split by profile (blue) and half-profile view (red) and by orthographic (dotted lines) and perspective projection (continuous lines). Error bars represent the standard error over subjects.

3.3   Experiment 1B

The results indicate that performance did not benefit from local motion information. However, in the previous experiment local motion was present only for 2 consecutive frames. That is, in the condition without local motion, points were relocated to a new position in every frame, while in the condition with local motion, points were relocated every two frames. To confirm that there is no benefit from local motion information we performed an additional experiment to test a wider range of point lifetimes. We varied lifetime between 1 and 32 frames. We concentrated this experiment on the condition with 128 points per trial and 4 points per frame since performance with 512 points per trial was already almost saturated in the lifetime 1 and 2 conditions, and the 2 points per frame condition with 128 points per trial in most cases indicated

a decline in performance for lifetime 2 over lifetime 1. We therefore thought that chances to see any benefit from higher lifetime would be highest in the 128 points per trial and 4 points per frame conditions.

Seven subjects took part in this experiment, four of them also participated in Experiment 1, three were new subjects. The seven subjects included two of the authors of the study. Each of the seven subjects conducted three experimental sessions. Each session had 48 different conditions (6 lifetime x 2 play-directions x 2 projections x 2 orientations) with 9 repetitions, and consisted therefore of 432 trials. Stimuli had 128 points per trial and 4 points per frame and were shown in profile and half-profile view. The frontal view will be tested separately in Experiment 3.

The results are displayed in Figure 5. Performance remained around 80 percent correct in all conditions. There was no effect of lifetime on performance in any condition. A 3-way repeated measures ANOVA with the factors lifetime, perspective and viewing angle (6 x 2 x 2 design) gave no significant effects or interactions.

3.4   Discussion

As explained in the Introduction, perspective projection of the movements of the light points might help to recognize biological motion if the walker is presented in non-profile views. Experiment 1 revealed no differences between the perspective and orthographic projections for any view, indicating that such motion-in-depth cues do not improve the recognition.

Overall performance differed between views. With respect to the half profile view, a possible advantage of perspective motion (lifetime 2 and higher) over non-motion (lifetime 1) might have been expected if local motion contributes to the perception of the 3D structure of the walker. However, increasing the lifetime did not have a positive effect on the performance. This showed that local motion information is not necessary for performing the task, and even does not give any advantage. In contrast, with 128 points per trial there even was a significant decrease from the profile view to the half-profile view, which was independent of all other fac-

tors. Discrimination performance for the limited-lifetime walkers otherwise was similar for the profile and the half profile view. Performance for 512 points per trial was as high as for classic walkers in profile and half-profile view.

There was a strong decrease in performance in the frontal view compared to the profile or half-profile view. Even with 512 points per trial the performance was a lot poorer than for the classic walker. For the classic walker there was also a small decrease in performance for the frontal view (t-test; p = 0.04). Results from both walker types suggest that the discrimination of walking direction in the frontal view is more difficult than in other views.
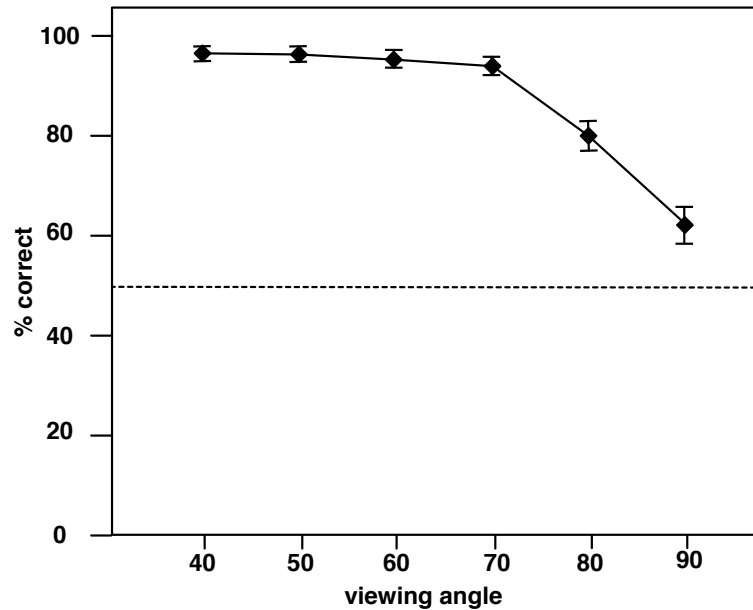
4 Experiment 2

The first experiment answered our main question in that the results showed no influence of local motion information on the perception of biological motion in perspective projection. The results of the first experiment suggested further that the frontal view is a special case for the forward/backward detection task. Therefore, in the following experiments the problem of the frontal view is further examined. In Experiment 2 we ask whether the poor performance for limited-lifetime walkers in the frontal view is limited to just the frontal view or whether performance gradually decreases between 40° and 90°. In Experiment 3 we investigate benefits from longer point lifetime the frontal view. In Experiment 4 we investigated the role of the foot-points for the frontal view.

4.1 Methods

Limited-lifetime walkers in different viewing angles were presented randomly. The orientations of the walkers varied in ten-degree steps from 40 to 90 degrees. The walkers had 4 points per frame and 512 points with a frame duration of 11.1 ms. We only used limited-lifetime walkers with lifetime 1 as there was no benefit for higher lifetimes in Experiment 1. All other methods were identical to those of Experiment 1.

4.2   Results and Discussion

Figure 6: Experiment 2: Averaged correct responses as function of the viewing angle. The stimulus was a limited-lifetime walker with lifetime 1, 8 points per frame, and 512 points per trial. The error bars represent the standard error over the individual subjects.

The percentages of correctly recognized walking directions are displayed in Figure 6. Performance was about equally good between 40° and 70° and with approximately 96 percent similar to the performance for the profile view in Experiment 1. From 70° to 80° the performance dropped to 80 percent and to 60 percent for 90°. However, performance remained above chance level even in the frontal view (t-test; frontal view $p = 0.02$; all other views $p > 0.0001$).

A repeated measures ANOVA showed a significant influence of viewing angle ($F(5,30) = 65.7$, $p < 0.0001$). Paired comparisons by Scheffé tests revealed no significant differences between 40° and 60° and neither between 60° and 70°. All other comparisons were significant ($p > 0.0003$). We thus concluded that a forward backward discrimination is more difficult in the frontal view than in other views.
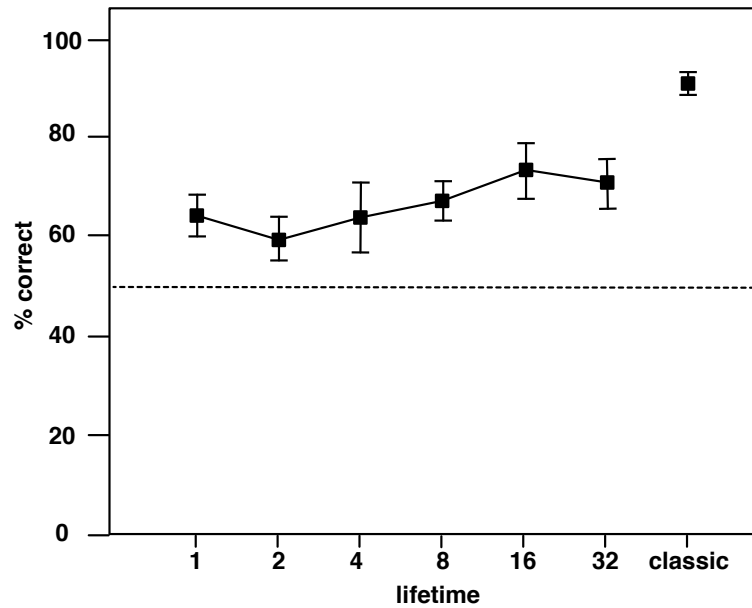
5 Experiment 3

The two above experiments showed that identifying the walking direction of a limited-lifetime walker was especially difficult in the frontal view. Although there was no difference between lifetime 1 and 2 in Experiment 1, it may be that information that becomes available only at longer point lifetimes is important. We thus asked whether an increase of lifetime beyond two frames could lead to an enhancement in performance in the frontal view. In Experiment 3 we gradually increased the lifetime until the lifetime of the limited-lifetime walker was similar to the lifetime of a classic walker. We compared the limited-lifetime walkers with a classic walker as control condition.

5.1 Methods

We used limited-lifetime walkers with a lifetime of 1, 4, 8, 16, or 32 frames. For the lifetime 32 condition each of the points changed its position on the body only once per trial. The limited-lifetime walkers had 12 points per frame, identical to the classic walker, and 384 points per trial with a frame duration of 44.4 ms (i.e. 8 multiples of the 180 Hz at which the monitor displayed). In addition to the limited-lifetime walkers we presented also a classic walker which consisted of 12 light-points, displaying the joints of shoulders, elbows, wrists, hips, knees and ankles.

5.2 Results

The results are displayed in Figure 7. Discrimination performance was high for the classic walker. For the limited-lifetime walker, overall performance was between 60 and 75 percent and never reached the performance level of the classic walker. A one-way repeated measures ANOVA revealed a significant main effect over the conditions ($F_{(6,25)} = 10.4$, $p < 0.0001$). The Scheffé tests revealed significant differences between the classic walker and all limited-lifetime walker conditions ($p < 0.04$), but no differences between the limited-lifetime walker conditions.

Figure 7: Experiment 3: Averaged correct responses in the frontal view as function of point lifetime. The error bars represent the standard error over the individual subjects.

5.3 Discussion

Increasing the point lifetime yielded to no significant improvement of performance, so there was still a clear difference between lifetime 32 and the classic walker. Thus, the difference between classic and limited-lifetime walker cannot be explained by a difference in lifetime, or local motion information, respectively.

The limited-lifetime walker with a lifetime of 32 frames resembled the classic walker in many aspects. Both possessed similar amounts of structural and local motion information, as the points of the lifetime 32 condition were relocated only once during the trial. However, the trajectories of the light points of the two walkers differ strongly because the points occupied different positions on the limbs.

For the classic walker the knee-, elbow-, hip-, shoulder-, and wrist-points move with approximately sinusoidal velocity profiles. Therefore their trajectories are symmetric in shape and carry practically no information about the walking direction. In contrast, the trajectories of the feet are asymmetric. They show a long backward movement with a slow rise when the

20

foot is lifted, and a quick forward movement with a sharp drop when the foot is put down. This asymmetry of the foot trajectory carries information about the walking. For instance, when walkers facing to the left have to be discriminated from walkers facing to the right, subjects rely very much on the visibility of the feet (Mather et al., 1992). Moreover, even when the other joints of the body are dislocated, the movement of the feet alone supports the perception of walking to the left or walking to the right (Troje & Westhoff, 2006). The information used is the differences in acceleration of the feet in the lift and drop phases (Chang & Troje, 2009). For the limited-lifetime walker, information from the feet is not directly available. Due to the random locations of the points there is not always a point located in the vicinity of a foot. Moreover, due to the relocation of the points, the amount of trajectory information from a single point depended on the lifetime of the point and was less than a walking cycle. Therefore, a comparison of the drop and the lift phases was not always possible.

Although the role of the feet for the discrimination between forward and backward walking has not been studied so far, and other views than the profile view have not been used, it is very plausible that the feet carry most of the information in those cases as well. Experiment 4 was designed to specifically test the role of the feet for walking direction discrimination in the frontal view.
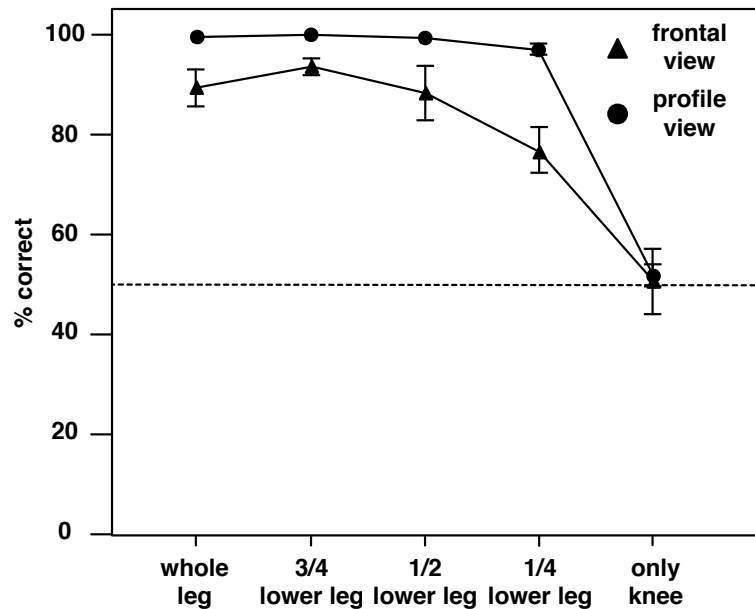
6   Experiment 4

To investigate the influence of the information from the foot-points we generated a classic walker on which the foot-points could be shifted toward the knees on the lower leg and thereby shortened the lower leg. The closer the foot-point was shifted to the knee point the less distinct was the asymmetry of the foot trajectory.

6.1   Methods

We generated walkers in which the lowest point was either positioned directly on the ankle, at three quarter of the lower leg, at the half of the lower leg, or at one quarter of the lower leg, and

a walker in which the foot-points were entirely omitted. In all other respects the walkers were identical to the classic walker with 12 points. All walkers were presented in the profile view and in the frontal view in blocked conditions.

6.2 Results

Figure 8: Experiment 4: Percentage of correct responses for walkers in which the lowest visible point was on different locations of the lower leg. The error bars give the standard error over subjects.

Figure 8 displays the results. Performance was overall lower in the frontal view than in the profile view and decreased with decreasing distance from the lowest point to the knee. For the profile view the performance was high for all inter-joint conditions. Performance was at chance-level when the foot-points were omitted for both profile and frontal view.

A two-way repeated measures ANOVA showed a highly significant main effect of the condition foot-point ($F_{(4,6)} = 89.6$, $p < 0.0001$) and a main effect of viewing angle ($F_{(1,6)} = 9.1$, $p = 0.02$). Paired comparisons by Scheffé tests revealed that the significant differences occurred between the condition where the foot-points were presented at one fourth of the leg and the condition where the foot-points were presented at three fourth of the leg ($p < 0.002$) as well as between the condition were only the knee points were shown and every other condition ($p <$

0.0001). Both conditions with missing foot-point did not differ from chance-level (t-test).

6.3   Discussion

For both views we found good performance for most of the inter-joint conditions, except for the condition where the lowest points were presented near to the knee-points, at one fourth of the leg. In this condition the subjects performed significantly poorer. For the condition where only knee points were presented the performance was at chance-level.

We conclude that information about the movement of the lower leg is necessary to discriminate walking direction. A similar importance of the foot-points has also been noted for example by Mather et al. (1992), who found that visibility of the trajectory of the feet and wrists are necessary for the discrimination of coherency and facing direction. Simulations of the template-model (Lange & Lappe, 2006) supported Mather's conclusion and suggested this is because the configuration of the extremities carries the most of the information about facing direction. Troje & Westhoff (2006) examined the inversion effect of biological motion and found that inverting only the feet of point-light displays has a much stronger detrimental effect than inverting all points except the feet. Chang & Troje (2009) showed that the comparison of the acceleration of the feet in the lift and drop phases is important.

However, in our experiments the discrimination performance was above chance level even when the point was presented at different locations between foot and knee. This suggests that it is not the foot itself that is important but rather information about the movement or the configuration of the lower leg. Moreover, in Experiment 2 discrimination was possible for the limited-lifetime walkers in view angles between 0 and 70 even though the trajectory of the foot was not directly available in those stimuli. In this case, information about the movement of the lower leg may be derived from a template analysis of the body configuration similar to that proposed for walkers in profile view (Lange et al., 2006; Lange & Lappe, 2006). The same can be expected for classic walkers with the lowest point placed somewhere on the lower leg since the movement of the leg can be inferred from the motion of a point on the leg in relation to the

knee. However, if the lowest point is very close to the knee estimation of the leg configuration is more difficult and errors become more likely.

In the geometric projection of the frontal view, however, the lower legs move almost exclusively vertically. The movement of the foot-point in the classic Johansson walker accurately describes the movement of the leg. The movement of a point somewhere on the lower leg also describes the movement of the lower leg sufficiently enough. In the limited-lifetime walker, however, the position of the point on the leg changes unpredictably. In other words, if the point on the lower leg is high in one frame and low in the next this must not indicate that the leg is lifted but could rather have resulted form the relocation of the point. Therefore, in the limited-lifetime walker the movement of the lower leg cannot be calculated from the positions of the light point on the leg over time. However, this is true only for the frontal view, because in all other orientations the perspective projection of the leg movement has a sideways component. In this case, a sequence of point positions on the leg allows to trace the orientation and movement of the leg over time. Thus, information about the leg configuration is available and supports estimation of the walking direction. Therefore, the failure to discriminate the walking direction of a limited-lifetime walker in the frontal view results from the particular projection properties of the walker in that orientation.

7   General discussion

We studied the use of local motion signals in discriminating biological motion of 3D oriented point-light walkers in profile, half-profile, and frontal views. Since the frontal view turned out to be a special case we will begin by discussing the profile and half-profile views and return to the frontal view thereafter. Our main question was whether speed and depth information derived from the local motion of the light points can aid biological motion recognition of 3D oriented walkers in perspective projection. To answer this question we measured discrimination performance on walkers with limited lifetimes of the point lights. We compared performance in perspective and orthographic projections between point lifetimes of 1 or more frames. If the

participants would use local motion to perceive walking direction one would expect a higher performance for higher lifetimes. Experiment 1 revealed no increase in performance between lifetime 1, 2, or higher.

Thus, we found no evidence of the necessity of local motion information for performing the walking direction detection task in either the profile view or the half-profile view.

Our stimuli either presented 2 or 4 points per frame. Because of visible persistence the apparent number of simultaneously visible points was higher than the number of physically present points (see methods). An increase in lifetime reduces the apparent number of simultaneously visible points, because the position change of a moving point between two frames is too small to give rise to two different apparent positions. One might argue, therefore, that the shorter the lifetime, the more configural information will be present in the stimulus. One might argue further that a constant level of performance for stimuli with longer lifetime might rely on local motion signals to compensate for the drop in configural information. Two lines of evidence argue against this. First, the performance in various discrimination tasks over a wide a range of lifetimes and frame durations has been shown to depend essentially only on the total number of points that were physically displayed in the stimulus (Beintema et al., 2006). Second, the template model (Lange et al., 2006; Lange & Lappe, 2006) could reproduce the performance data of Beintema & Lappe (2002) on lifetime variation, while simulating visible persistence but without using local motion signals.

As we expected from other studies with varying viewing angle, the performance for the classic walker was high in every view angle with only a slight drop in the frontal view. Vanrie et al. (2004) for example used a facing direction task and classic walkers in perspective projection (their Experiment 3). They found no difference in performance between profile and half-profile view, but in the frontal view the performance was only about 85 percent correct. Gender classification tasks (Mather & Murdoch, 1994) and person identification tasks (Troje et al., 2005), on the other hand, indicated an advantage for the frontal view with respect to profile and half-profile view.

The results for the limited-lifetime walker depended stronger on viewing angle. For profile and half-profile views, discrimination performance was similar for the classic walker and for the limited-lifetime walker with 512 points per trial, but for the frontal view performance with the limited-lifetime walker was much lower. The different behavior for the frontal view was investigated in three experiments. The first one (Experiment 2) showed that the difficulty for direction discrimination was restricted to a small range around the frontal view. Experiment 3 showed that longer lifetimes did not change performance, ruling out a direct role for local motion. Experiment 4 revealed that the missing information to perform the task was in the foot-points. Classic walkers with missing foot-points resulted in poor performance in frontal and profile view. Varying the position of the lowest point on the lower leg influenced performance on the limited-lifetime walkers but also on the classic walker. In the limited-lifetime walker, the lowest light point occupies varying positions on the lower leg and may therefore have led to an overall lower performance.

To reconcile our two main results, i.e. the lack of influence of local motion and the importance of the movement of the feet for the walking discrimination, we need to discuss how the movement of the feet (and the rest of the body) can be estimated from the stimuli that were presented. The results of experiment 1 show that local motion detectors are not used. This is consistent with current theoretical frameworks for biological motion recognition such as the template approach (Lange et al., 2006; Lange & Lappe, 2006) or the interactive encoding model (Dittrich, 1999). The interactive encoding model of Dittrich (1999) assumes three different routes by which the trajectory information of the points of an biological motion stimuli is processed further. These routes are connected by so called motion integrators. The first route is strictly based on the analysis of the structural components of human motion to reconstruct 3D body information out of the 2D trajectory information. The second route is linked to the memory system and allows to apply cognitive constraints relating to the human body and its motion trajectories for the 3D reconstruction. The third route relies on visual semantic stored in respect to action categories. Here, in contrast to the template matching, cognitive processes

aid the perception process. If low-level motion is lacking, recognition can be enhanced because input signals can be amplified by stored information. Thus, our findings support the interactive encoding model's proposition that the level of processing is highly variable depending on the type of information available to the viewer.

A slightly different approach is provided by template matching (Lange et al., 2006; Lange & Lappe, 2006). Template matching may estimate leg configuration if suitable templates are available. These templates may exists either as multiple 2D view templates with varying orientation or as 3D templates. A computation directly from the trajectory of the foot-point, which would be possible for the classic walker, is not possible for the limited-lifetime walker because the foot-point is unlikely to be visible for an extended amount of time due to point relocation. A computation of the leg trajectory from local motion signals of points on the leg is also unlikely because results in the lifetime 1 condition and the higher lifetime conditions were identical even though local motion is not available in the lifetime 1 condition. We therefore suggest that the trajectory of the leg is derived from template matching analysis. Whether this involves 2D, view-dependent templates or 3D templates is a matter or further research. Template matching processes as proposed by Lange & Lappe (2006) are simple automatic processes, which can take place at an early stage of visual processing.

We conclude that for discrimination of the walking direction of 3D oriented walkers there is no need for a local motion mechanism. Instead, the use of a local motion independent high-order motion detector based on static templates seems more plausible. The results of the profile and half-profile view show that local motion information does not add essential information to the perception process. However, global mechanisms that are used in other views are useless in the frontal view. Here local mechanisms like the use of local motion information, trajectory information or relative motion of the foot-points with respect to the knee aid perception.

Acknowledgments

References

Aggarwal, J. K. & Cai, Q. (1999). Human motion analysis: a review. *Comp. Vis. Image Understand.*, *73*, 428–440.

Beintema, J. A.; Georg, K., & Lappe, M. (2006). Perception of biological motion from limited lifetime stimuli. *Percept. Psychophys.*, *68*(4), 613–24.

Beintema, J. A. & Lappe, M. (2002). Perception of biological motion without local image motion. *Proc. Natl. Acad. Sci. USA*, *99*, 5661–5663.

Brown, R. W.; Pola, J., & Matin, L. (1974). Visual persistence: effects of flash luminance, duration, and energy. *Vis. Res.*, *14*(4), 295–303.

Bülthoff, I.; Bülthoff, H. H., & Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nat. Neurosci.*, *1*, 254–257.

Chang, D. H. F. & Troje, N. F. (2009). Acceleration carries the local inversion effect in biological motion perception. *J. Vis.*, *9*(1), 1–17.

Chen, Z. & Lee, H.-J. (1992). Knowledge-guided visual perception of 3-D human gait from a single image sequence. *IEEE-SMC*, *22*(2), 336–342. (doi:10.1109/21.148408).

Coltheart, M. (1980). Iconic memory and visible persistence. *Percep.Psychophys.*, *27*(3), 183–228.

Dittrich, W. H. (1999). Seeing biological motion - Is there a role for cognitive strategies? In Braffort, A.; Gherbi, R.; Gibet, S.; Richardson, J., & Teil, D., Eds., *Proceedings of the*

International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction, volume 1739 of *Lecture Notes In Computer Science*, pages 3–22, London, UK. Springer-Verlag.

Hoffman, D. D. & Flinchbaugh, B. E. (1982). The interpretation of biological motion. *Biol.Cybern.*, pages 195–204.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Percep. Psychophys.*, *14*, 201–211.

Lange, J.; Georg, K., & Lappe, M. (2006). Visual perception of biological motion by form: a A template-matching analysis. *J. Vision*, *6*(8), 836–849.

Lange, J. & Lappe, M. (2006). A model of biological motion perception from configural form cues. *J. Neurosci.*, *26*(11), 2894–2906.

Marr, D. (1982). *Vision*. Freeman, San Francisco.

Mather, G. & Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proc. Royal. Soc. London B*, *258*, 273–279.

Mather, G.; Radford, K., & West, S. (1992). Low-level visual processing of biological motion. *Proc. Royal. Soc. London B*, *249*, 149–155.

Neri, P.; Morrone, M. C., & Burr, D. C. (1998). Seeing biological motion. *Nature*, *395*, 894–896.

Rashid, R. F. (1980). Towards a system for the interpretation of moving lights display. *IEEE Trans. Patt. Anal. Mach. Intell.*, *2*(6), 574–581.

Troje, N. F. & Westhoff, C. (2006). The inversion effect in biological motion perception: evidence for a "life detector"? *Curr. Biol.*, *16*, 821–824.

Troje, N. F.; Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: effects of structural and kinematic cues. *Percept. Psychophys.*, *67*(4), 667–675.

Ullman, S. (1984). Maximizing rigidity: the incremental recovery of 3-D structure from rigid and nonrigid motion. *Perception*, *13*, 255–274.

Vanrie, J.; Dekeyser, M., & Verfaillie, K. (2004). Bistability and biasing effects in the perception of ambiguous point-light walkers. *Perception*, *33*, 547–560.

Verfaillie, K. (1993). Orientation-dependent priming effects in the perception of biological motion. *J. Exp. Psychol.: Hum. Percept. Perform.*, *19*, 992–1013.

Webb, J. A. & Aggarwal, J. K. (1982). Structure from motion of rigid and jointed objects. *Artificial Intelligence*, *19*, 107–130.