LightGuide: Projected Visualizations for Hand Movement Guidance

Rajinder Sodhi[1,2] **Hrvoje Benko**[1] **Andrew D. Wilson**[1]

[1]Microsoft Research
One Microsoft Way
Redmond, WA 98052
{benko, awilson}@microsoft.com

[2]Department of Computer Science
University of Illinois
201 North Goodwin Avenue, Urbana, IL 61801
rsodhi2@illinois.edu

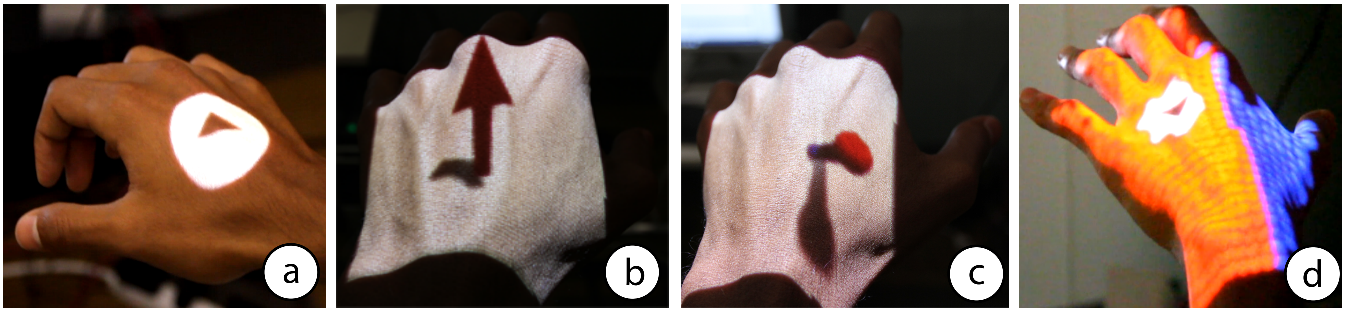

Figure 1. An overview of the range of 3D cues we created to help guide a user's movement. In (a), a user is shown a 2D arrow with a circle that moves in the horizontal plane, (b) shows a 3D arrow, (c) a 3D path where blue indicates the movement trajectory and (d) uses positive and negative spatial coloring with an arrow on the user's hand to indicate depth.

ABSTRACT

LightGuide is a system that explores a new approach to gesture guidance where we project guidance hints directly on a user's body. These projected hints guide the user in completing the desired motion with their body part which is particularly useful for performing movements that require accuracy and proper technique, such as during exercise or physical therapy. Our proof-of-concept implementation consists of a single low-cost depth camera and projector and we present four novel interaction techniques that are focused on guiding a user's hand in mid-air. Our visualizations are designed to incorporate both feedback and feedforward cues to help guide users through a range of movements. We quantify the performance of *LightGuide* in a user study comparing each of our on-body visualizations to hand animation videos on a computer display in both time and accuracy. Exceeding our expectations, participants performed movements with an average error of 21.6mm, nearly 85% more accurately than when guided by video.

Author Keywords

On-demand interfaces; on-body computing; appropriated surfaces; tracking; spatial augmented reality;

ACM Classification Keywords

H.5.2 [Information interfaces and presentation]: User Interfaces - Input devices & strategies;

CHI'12, May 5–10, 2012, Austin, Texas, USA.

INTRODUCTION

When performing gestures that are intricate or that require a great deal of technique, physical feedback from an instructor can often be useful for performing a movement. For example, when someone wants to perform the proper technique for a weight training exercise, an instructor often gives instantaneous feedback by gradually correcting the position of the user's body through physical touch.

While this exchange seems crucial, the availability of such a resource disappears when a user is no longer in the presence of an instructor. Instead, directing human movement is usually accomplished through video recordings, diagrams, animations, or textual descriptions. We rely on a bevy of online resources that include detailed graphical imagery or do-it-yourself videos (see Figure 2). However, without incremental and real-time feedback, interpreting and following a set of movements can still be a challenge.

In this paper, we explore an alternative approach to movement guidance where body movement can be directed using projected visual hints. Our system, *LightGuide*, provides users with real-time incremental feedback for movement guidance that is projected directly on their hand (see Figure 1). *LightGuide* provides a unique benefit to existing gesture guidance methods: users can focus their attention directly on the body-part rather than divide their attention between a video screen and the movement. Users can move their body-parts freely in space, releasing the user from always being orientated towards a video screen. All our system requires is a projector and depth-sensing camera. While our system does not require a user to be physically instrumented with a device, these hints can also be used in a body-worn [12,13] or on limited-screen space handheld devices, such as smartphones.

Thus, our work provides three primary contributions: First, we introduce a series of unique visualizations for movement guidance that incorporate feedback and feedforward cues. Second, we contribute a prototype system, *LightGuide*, which is comprised of a single overhead projector and depth-sensing camera to sense the user and their movements. Our proof-of-concept system facilitates the display of our visual hints on a user's body and allows us to replay pre-recorded or system-generated 3D paths at a user-driven pace or dynamically controllable speeds. Finally, we show results of our quantitative comparative evaluation, qualitative user feedback and discuss the pros and cons of our approach.

MOTIVATION

We can envision a number of practical applications that leverage on-body projected hints for guidance. For example, imagine an amateur athlete working on punching exercises during martial arts training. With projected hints, the system can direct the user towards the optimal reach of the arm to ensure that the shoulder is not overextended to cause injury. In another example, physical therapy patients recovering from an injury can be guided through practicing exercises at home. Novice musicians learning to play an instrument can by directed to the correct posture when their form begins to drift. We believe that all of these movements can be guided with correct spatially registered projections on a user's body.

RELATED WORK

Our on-body projection approach draws from a variety of fields, including computer-aided instruction, augmented reality, and projection-based guidance. Here we focus on the relevant related work from these areas to position and understand the contributions of the present work.

Computer-Aided Task Guidance

Receiving task guidance through computer-aided instruction has been a research focus for decades. Demonstration through animated-videos have shown that computer-based instruction can improve task performance, particularly in assembly based tasks [31]. Palmiter and Elkerton showed that well placed textual hints with animated videos can also give immediate benefits for task performance [23]. Others have explored adding graphical visual hints to video in post to help users explore a range of dance movements [29]. While these hints are statically placed in video, previous literature has also looked at using co-located real-time feedback and feedforword mechanisms to provide on-demand assistance to guide users through gestural tasks [3,11]. Such systems lead users in-situ, while a user is in the process of performing the gesture. Our work draws upon this prior research where we explore how co-located projection-based on-body hints can help show similar improvements for movement tasks.

Figure 2. Examples of how people currently follow instruction for movement (e.g. Kinect virtual avatar, stretching, rhythmic dance notation [4]**).**

Task Guidance in Augmented Reality

The field of Augmented Reality (AR) has shown a number of methods to provide guidance by using head-mounted displays or mobile devices to convey instructions that are superimposed on virtual or real world video feeds [1,8]. Feiner et al. explored using AR to guide users through repairing laser printers. More recently [21,22], AR has been demonstrated for a variety of tasks, such as playing the guitar or manufacturing. In tangible AR, White et al. explored using a variety of graphical representations through 'ghosting' to enable the discovery and completion of gestures [32]. While these approaches are promising, head-mounted displays can be cumbersome for users to wear and diminutive screens can constrain the user experience.

Augmenting Environments with Projectors

Recent advancements in projection technology have made it possible to imbue user's environments with projection capabilities [24,26,27,33]. For example, Wilson and Benko explored using a series of depth-sensing cameras and projectors to transform a room-sized environment to enable un-instrumented surfaces (e.g., a desk) to emulate interactive displays.

In addition, the emergence of miniaturized projection technology has opened up the possibility of appropriating the user's body as a display surface where even a user's hand alone contains more surface area than a typical smart phone [12,13]. In addition to body-worn projection systems, handheld [5,6] and head-mounted projectors [15] also allow users to be mobile, without requiring their environments to be permanently instrumented with projectors and cameras [18,20,25]. All of these projection-based approaches are similar to our approach using a depth-sensing camera for tracking and a projector to turn an arbitrary surface into an interactive display.

Projection-Based Guidance

The use of projection-based augmented reality AR for guiding users through tasks has been a research vision in recent years [10]. Kirk et al. looked at using projection in remote collaborative scenarios (e.g., remote Lego building) with real-time projection guidance co-located next to the user's hand on a static desk [19]. Similarly, Rosenthal et al.

found that combining static text and pictorial instructions on a screen with micro-projection based guidance on physical objects improved overall task performance [28]. In contrast to prior literature, we explore projecting visual-hints with real-time feedback directly on the user's hand that is tracked in mid-air for movement guidance.

DESIGN CONSIDERATIONS FOR GUIDING MOVEMENT

To provide in-situ guidance for the user's movement, visual hints need to *convey* a sense of where to move next. We are motivated by the idea that one can co-locate the instruction for the movement with the body part that needs to be moved along a desired path. To inform the design of such hints and the validity of the overall approach, we focused this work on projected hints on the user's hand as it moves freely in space. We believe that our approach allows the user to focus their attention on body part and the movement itself. Through our initial exploration as well as leveraging prior literature, we highlight six critical aspects that need to be considered when designing on-body guidance hints: (1) feedback, (2) feedforward, (3) scale, (4) dimension, (5) perspective and (6) timing.

Feedback

Feedback components provide information about the current state of the user during the execution of the movement. This feedback can come in the form of a user's current position, the path the they took (e.g., [3]), or their error or deviation from the path, to name a few. For example, with position, the feedback can either be relayed to the user in a relative sense (e.g. a user's projected "progress" along a movement path) or in an absolute sense (e.g. a user's absolute deviation from a movement path).

Feedforward

Feedforward components provide information to the user about the movement's shape *prior* to executing the movement. As described in [3,11], the feedforward can come in the form of showing the user where to go next, a segment of the movement path ahead, or simply show the user the entire movement path. One possible downside for showing the whole movement is for sufficiently complex paths, path self-occlusions may obstruct a user's view of where to move next.

Scale

To gain insight into how to convey scale, we consider Steven's Power Law which describes a relationship between magnitude of a stimulus (e.g., visual length, visual area, visual color) and its perceived intensity or strength (projected line, projected circle, projected intensity) [2]. That is, the relationship allows us to understand how users perceive visual cues, e.g., the area of a circle, color, or the length of a line, and describes how well they convey the scale of a movement (e.g., what is the distance I should move my hand to get from point A to point B when a projected line denotes distance versus using area or color to denote distance?).

Dimension

As found in [17] the way in which the user perceives the structure of the task greatly affects their performance for high-dimensional input. As such, how we convey where to move in three-dimensions depends on how intuitive the user finds the visual hint. For certain users, the most intuitive way to get from point A to point B may be in the form of a visual hint that is broken down into two distinct components, e.g. where to move horizontally and vertically. In contrast, for others, a single metaphor hint may be the most perceptually intuitive, e.g. go from point A to B all in one simultaneous task.

Perspective

One aspect of conveying an on-body visual hint is to explore egocentric and exocentric viewpoints [30] (e.g. a first person and third person perspective, respectively) With an egocentric viewpoint, we want users to get a greater sense of presence where the hints become a *natural extension* of their bodies, reinforcing guidance by "tugging" the user's hand along the movement path. In contrast, with an exocentric viewpoint, rather than seeing guided hints embodied in the user, they are seen at an overview (e.g., a video).

Timing

In our design of an on-body visual hint, we feel that there are two main approaches that may effectively communicate timing in motion: *system imposed timing* and *self-guidance.* For system imposed timing, users follow a visual hint that is displayed at a system specified speed. A visual hint can convey a range of dynamics, such as in keeping the speed constant or changing it dynamically throughout the movement. For self-guidance, the user can see a visual hint and choose the pace at which they react to the hint.

LIGHTGUIDE PROJECTED HINTS

We describe a set of visual hints that follow important aspects of the design space we have highlighted. Our visual hints can be used to help guide a user's movement in all three translational dimensions. To our knowledge, this is the first implementation of on-body projected hints for real-time movement guidance. While this is a rather large design space with many possible solutions, our iterative design process included an analysis of 1D, 2D and 3D visual hints and offers a set of compelling solutions that can inform future designs. We focus our descriptions on the final hint design which resulted from our iterative process, but encourage the reader to see the accompanying video for a more complete reference of alternatives.

In this initial exploration we have chosen to focus and verify our ideas by tackling hand translation first (i.e., movements in the *x, y* and *z* dimensions), without any rotations of the hand. As such, we leave a visual vocabulary for 3D rotations to future work.

Follow Spot

The Follow Spot can be seen in Figure 3(a)-(b). Through our initial pilots, we found the most intuitive metaphor for

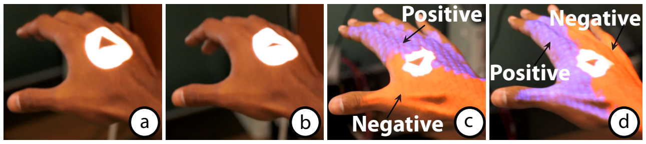

Figure 3. In (a)-(b) the *Follow Spot* shows a user a white circle and a black arrow reduces in size when user moves their hand up, (c)-(d) the *Hue Cue* shows positive coloring (blue) which represents the direction a user should follow horizontally while moving away from the negative coloring (red).

users was to use 1D visual length (e.g. distance), which is reflected in the mapping specified by Steven's Law [2]. To specify feedback in depth, the 1D arrow points away from the user to signal moving up and points the arrow towards the user to signal moving down. The size of the arrow dictates the distance to the target depth position communicating the scale of the movement. That is, as the user moves up in the z-direction to hit a target depth as specified by a large black arrow pointing away from the user, the tip of the arrow decreases in size until it becomes a black horizontal line. The visual hint otherwise contains no feedforward mechanism.

Hue Cue

We create a visual hint that utilizes negative and positive spatial coloring to indicate direction and the space a user should occupy, shown in Figure 3(c)-(d). The cue uses a combination of spatial coloring in x and y and depth feedback in z to guide a user's movement in three dimensions. The feedforward component is conveyed in the positive coloring, shown in blue and the negative coloring for feedback in red. To perform the whole movement, a user can continuously move toward the blue and away from the red. In order for a user to see if they are moving at the correct depth, a *Follow Spot* hint is projected in the middle of the hand.

3D Arrow

We create a more direct mapping to visualize direction by conveying a simple 3D Arrow to the user, shown in Figure 4(a)-(b). The benefit of using a *3D Arrow* is that direction for all three dimensions, *x, y and z* can be conveyed in a single metaphor. Additionally, to engage the user's egocentric viewpoint, we render the *3D Arrow* from the user's perspective and add shading to emphasize its 3D shape.

3D Pathlet

We create a *3D Pathlet* metaphor where users are shown a small segment of the path ahead in the movement. This visual hint allows users to see a segment of the path, denoted in blue in Figure 4(c)-(d) as a form of feedforwad. The red dot provides users with their relative position, projected on the movement path. The benefit of the *3D Pathlet* is that users can see changes in direction of curved motions along the path well before they execute the movement. Figure 4(c)-(d) shows a user completing a movement shaped in the form of the alphabet letter 'N' displayed at a 45-degree angle. Additionally, for perspective, a similar shadow is used to emphasize the *3D Pathlet*'s shape. As shown in Figure 5, when the user distorts their hand significantly, the 3D illusion is diminished.

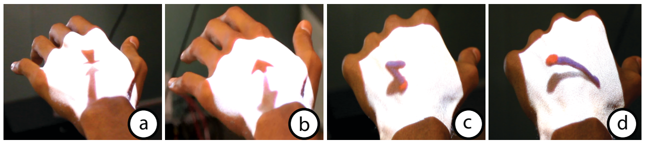

Figure 4. In (a)-(b), the *3D Arrow* is shown pointing down and up, (c)-(d) the *3D Pathlet*, shows the user (red dot) a small segment of what is ahead in the path (denoted in blue).

Movement Guidance Algorithm

LightGuide can replay any pre-recorded movement (e.g., recorded with a depth sensor) or ideal generated path (e.g., parametric wave). For a path, we summarize our algorithm (Figure 6) as follows: The path is first pre-processed into segments, where a segment is composed of two points in the order with which we wish to guide the user. The path is then translated to the user's current hand position where the visual hint is rendered. As the user follows a visual hint, any deviation from the path can result in an absolute, relaxed-absolute or relative projection (Figure 6(a)-(c)). The user continues through the path using one of these three approaches until the path is complete.

The absolute projection results in a visual hint that immediately guides the user back to the movement path once deviated, the relaxed-absolute movement slowly guides the user back to the movement path and the relative projection simply shows the user the next direction of the movement without requiring the user to be directly on the path. Each projection type is task dependent. For example, a dancing movement may be less stringent about following the exact path and could thus use a relative projection. In contrast, an exercise movement where a user can potentially strain a muscle if done incorrectly may use an absolute or relaxed-absolute projection. Based on our initial pilots, we chose to have the *Follow Spot* use an absolute mapping, the *Hue Cue* to have a relative mapping in *x and y* and an absolute mapping in *z,* while the *3D Arrow* and *3D Pathlet* use a relative mapping.

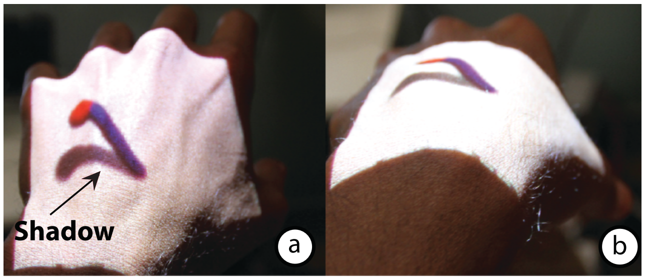

Figure 5. In (a), the *3D Pathlet* creates an illusion of the path extending beyond your body where the shadow emphasizes the 3D nature of the hint. In (b), the illusion is diminished when the user's hand is orientated at an extreme angle.

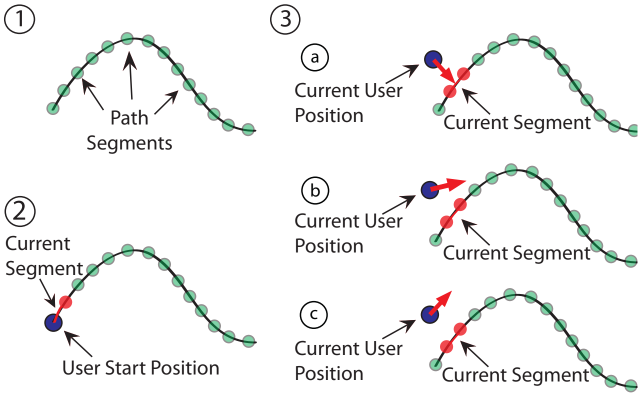

Figure 6. Our algorithm first breaks down the path into smaller segments. The path is translated to the user's current hand position and the visual hint is rendered to begin guiding the user. When the user deviates from the desired path, the visual hint can, (a) direct the user back to the closest point on the path, (b) incrementally bring the user back to the path, or (c) guide the user through a relative movement.

Dynamics

For *system imposed timing*, LightGuide can replay the visual hint so that it follows the movement path automatically in space at any speed. To ensure that the visual hints do not move off of the user's hand, we followed the same procedure as [12] in which we compute a derivative map of the depth image to check for large changes in the boundaries at the contours of the hand. That is, if the visual hint reaches the contour of the hand, it stops moving until a user has adequately caught up to the path. For the *self-guidance* approach, the system relies on the user to direct themselves through the movement. A visual hint describes the motion trajectory through feedback and feedfoward cues and a user can choose their own pace.

LIGHTGUIDE IMPLEMENTATION

Our proof-of-concept LightGuide system, seen in Figures 7, consists of two primary components. First is a commercially available Microsoft Kinect Depth Camera, which provides 640x480 pixel depth images at 30Hz. The second component is a standard off the shelf InFocus IN1503 wide-angle projector (1280x1024 pixels) [16]. The depth camera and projector are both rigidly mounted to a metal stand positioned above the user. This ensures that we could adequately see the user's hand motions as well as to ensure that our projected visual hints would fully cover the user's hands.

The visual hints are rendered from a fixed perspective that assumes a user is looking down a 45-degree angle towards their hand. While occlusion (particularly self-occlusion) is a fundamental problem with all projector-camera systems, we do not feel that this played a significant role in users' interactions. In the future, multiple projectors and cameras can be used to help reduce the effects of occlusions on more complex unconstrained movements.

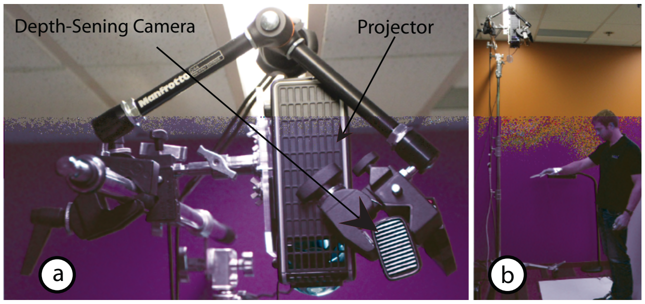

Figure 7. (a) LightGuide uses a single projector and depth sensing camera, (b) the projector and depth camera are fixed over the user's body.

Projector Camera Calibration

For the visual hints to be correctly projected on a user's hand, we must first unify the projector and camera into the same coordinate space. We calibrate our projector to the depth camera as the camera already reports real world coordinates (mm). The intrinsic parameters of the projector can be modeled using the diagonal field of view and the center of projection. To compute the extrinsic parameters, we require four non-coplanar correspondences between points that can be seen in the depth camera and projector. Once we establish correspondences between the 2D points of the projector and the 3D points of the camera, we use the POSIT algorithm [7] to find the position and orientation of the projector.

Hand Tracking

The prototype system first transforms every pixel in the input image into world coordinates and then crops pixels outside of a volume of 1 cubic meter. This removes the floor, walls and other objects in the scene (e.g. a desk). The prototype then identifies the user's arms by determining continuous regions along the depth image. The system then finds the farthest point along the entire arm by tracing through the continuous region eventually reaching the most distant point along the hand. To extract the user's hand, we assume a constant hand length [14] which worked well in our tests. A distance transform [9] is then used on the resulting image and the maxima is assumed to be the center position of the hand.

USER STUDY

The purpose of this study was to demonstrate the feasibility of our approach and to determine if our prototype is capable of guiding a user's hand in mid-air. Specifically, we wanted to know how accurately users follow on-body projected visualizations. We also wanted to investigate how the accuracy and behavior of a user changes for paths at varying depth levels. In addition to following, we also explored the accuracy and speed of self-guided movements where users dictate their own pace of a movement.

To place LightGuide's performance in context, we compared our method to video as we felt it was representative of a resource that users currently utilize. The *video condition*, shown in Figure 8, is comprised of a 3D

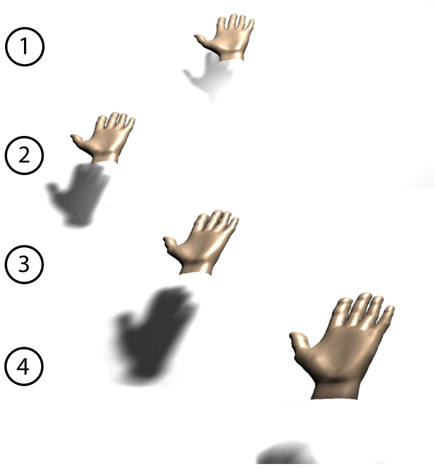

Figure 8. A rendering of the 3D hand that is used in our video condition. The motion is an arc that moves towards the user and gradually increases in depth.

model of hand that follows an ideal, system-generated path. Although our animated video does not provide nearly as much visual context to participants as a real life video, a system controllable video allowed us to remove the effects of any human or tracking error that could affect the movement paths. More importantly, the animated video allowed us to control the perspective of the video (e.g. rendered from the user's perspective) as well as precisely control the speed and timing of replayed movements. While we feel that the best performance with our system can be attained by using both video and on-body hints, our comparison independently measures the effect of our visual hints and video for movement guidance.

Participants

We recruited 10 right-handed participants from our local metropolitan area (2 female) ranging in age from 18 to 40. All participants were screened prior to the study to ensure their range of motion was adequate to perform our tasks. The study took approximately 90 minutes and participants received a gratuity for their time.

Test Movements

Our goal was to support interactions on a variety of movements. For our user study, we included five different paths: a line which must be traced back and forth, a square, a circle, an 'N', and a line plus a curve (Figure 9). These paths share similar characteristics to the types of movements patients are asked to perform in physical therapy sessions (see Motivation). The paths, seen in Figure 9, range in length from 300 to 630mm (mean = 438.1 mm, SD = 130.6mm). To ensure that we adequately tested a variety of depth levels, we vary the paths at three different angles: 0°, 45° and 90° with respect to the horizontal plane in the participant's frame of reference.

Procedure

During the experiment, participants were instructed to stand at a comfortable position underneath the overhead projector and depth-sensing camera. Prior to starting, we verified that each participant had enough room to move their hand while being adequately tracked by the system.

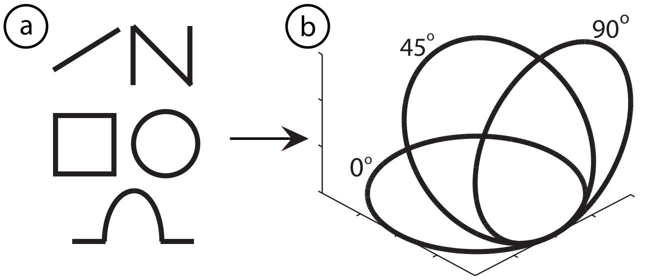

Figure 9. In (a) the test paths used in our study, (b) each path is oriented at 0°, 45° and 90° (only a circle path is shown).

The primary task consisted of a participant moving their hand in space following specific hand guidance visual hints. By 'following', we mean that a visual hint would begin moving in space at a speed of 30 mm/sec. and participants would follow the hint and respond to its cues. Our choice of 30mm/sec for visualization speed was chosen through informal pilot studies that had users try out a variety of speeds. 30 mm/sec was chosen to be the most comfortable constant speed while still producing reasonable hand motions. To quantify how users perform a movement at their own pace, a secondary task was included where the same *3D Arrow* was used without any system imposed timing. That is, the *3D Arrow* would only change position if the user responded to the direction indicated by the *3D Arrow*. We refer to this as *self-guided.*

We performed a within subjects experiment and in total, we tested 6 visual hints: *Follow Spot, 3D Follow-Arrow, 3D Self-Guided Arrow, 3D Pathlet, Video on Hand,* and *Video on Screen.* Here on, we refer to our two *3D Arrow* conditions as *3D F-Arrow* and *3D SG-Arrow.* All except the *Video on Screen* condition were projected on the participant's hand. Our baseline *Video on Screen* condition was shown to a participant on a computer monitor situated directly in front of the user. Importantly, participants were told to keep their hands flat (facing down) during the entire experiment to ensure that the visual hints would consistently appear on their hands between trials as well as to ensure consistent hand tracking performance by our system.

To provide consistent start location for each movement, we marked the desired starting hand location with markers on the floor in front of the participant and asked them to return to the marker before beginning each new trial. In each trial, participants were instructed to hold out their hand and follow the guidance cues completing a single path as accurately as possible. We asked the participant to keep the visual hint at the center of their hand. Once the path was completed, the system would sound a 'chime' and a red circle would appear on the participant's hand signaling the user to return to the start position. In total, participants were asked to follow a single visualization over our 15 test paths;

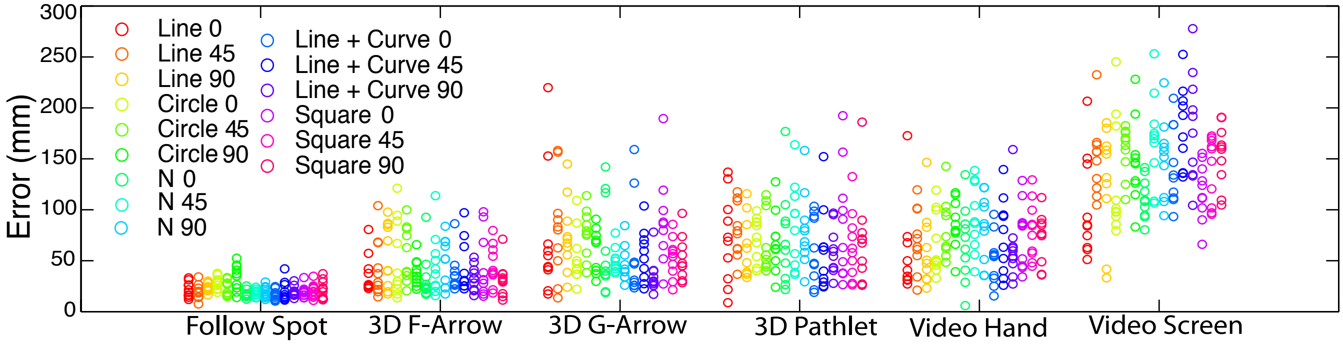

Figure 10. Overall distribution of unscaled deviations from a path. The circles denote users while colors show the 15 unique paths.

presentation order was randomized. The procedure was repeated for each of our conditions.

Before each measurement phase, participants were allowed to practice using the visual hints to move through a path. Each condition lasted approximately 10 minutes, of which 5 minutes was used for practice and 5 minutes for measurement. Between conditions, we allocated 5 minutes for participants to rest in order to reduce the effects of hand fatigue.

Each session produced 90 trials (6 conditions x 5 paths x 3 angles) per participant. To counter-balance the conditions, the presentation of each condition was randomized to remove the effects of ordering. Users were interviewed after each session followed by a short post-study interview. We recorded video of the participants and measured their position, hand-orientation and time.

RESULTS

We separate our analysis into two components: *Movement Accuracy and Movement Times.*

Our 10 participants produced a total of 900 movement trials on 15 unique paths. During the study, we experienced only a single type of outlier relating to the tracking of a user's hand. The tracking results would change depending on if the user would self-occlude their hand (e.g., rotate towards the principal axis of the camera). Additionally, we experienced 21 trails (2%) where users would lean their bodies into the capture volume, leading to momentary erroneous hand measurements that would only appear in the outer extents of the capture volume. The erroneous measurements in the outer extents were filtered in post-data analysis allowing us to use all trial measurements in our final analysis.

Movement Accuracy

We take a two-fold approach on measuring the accuracy of movements: deviation from the path and fit (e.g., see [14]). In both cases, to determine accuracy, we use the absolute Euclidian distance from the closest point as an error metric.

As in prior literature [12], we highlight two sources of systematic error: 1) non-linearity and improper calibration of the projector and camera (e.g., the location of the projected visualization differs from where the camera expects it to be) and 2) inaccuracy in the hand tracking, especially when the user's hand begins to leave the capture volume. Overall, we found a small global systematic offset between the camera and projector where the average X-offset across users was 9.02mm to the left of a path and a Y-offset of 1.05mm below a path, which is in agreement to findings in previous literature [12,14].We did not apply these global X/Y offsets, as participants would compensate for the system inaccuracy in the 'following' conditions by moving their hand until the visualization appeared at the center of their hand. In the self-guided condition, the location of the *3D G-Arrow* was sufficiently well placed in all our trials so that participants could see the visual hint.

Movement Deviations

We analyzed the average deviations of users across all paths and visualizations using their raw, unscaled, distances to the closest point on the path, (see full distribution in Figure 10 and a single user's performance in Figure 11). Using a standard ANOVA, we found that there was a significant difference between our visual hints ($F_{[5,\ 894]} = 276.5, p < .001$) . A post-hoc Bonfferroni-corrected t-test on the *Follow spot* and *3D F-Arrow* performed significantly better than both video conditions with average deviations of 24.6mm (SD = 9.0mm) and 49.9mm (SD = 29.17mm) respectively ($t_{16} = 25.6, p < .001, t_{26} = 122.5, p < .001$).

Additionally, the distribution highlights the difficulties users had in perceiving scale for our animated videos. Surprisingly, a Bonfferroni-corrected t-test comparing the accuracy of our video conditions show that significantly smaller deviations can be achieved by showing an identically rendered video, on the user's hand ($t_{56} = 93.0, p < .001$).

Movement Shape

Although the unscaled distribution in Figure 10 shows that our users were not able to achieve the desired scaling on a path with the *video screen* condition, the results do not explain how well users do at performing the shape of the movement. To help analyze shape, we use the Iterative Closest Point (ICP) algorithm to register the user's movements to our model paths [34]. With ICP, we have the flexibility of rotating, translating and scaling an object in all three axes to find the best match. For our purposes, we

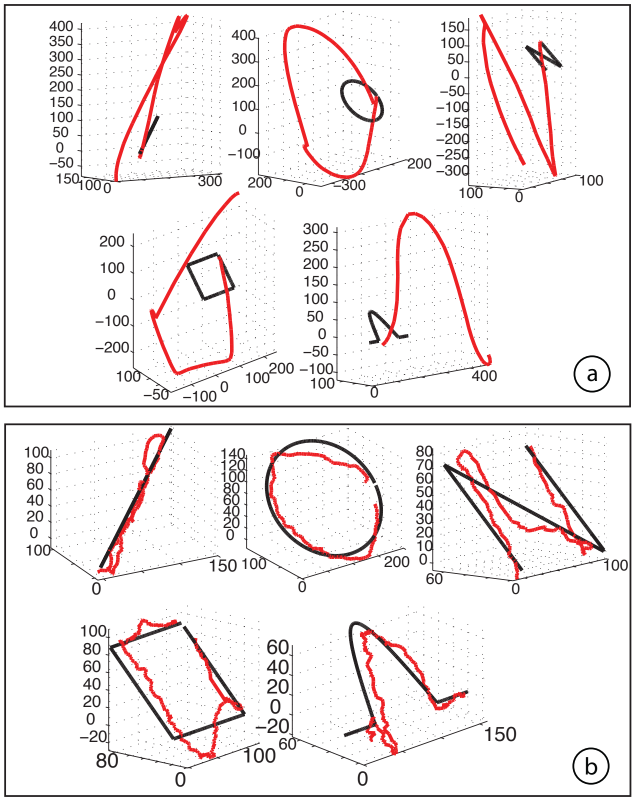

Figure 11. A single user's performance on paths oriented at 45° using the *Video Screen* (top) and *Follow Spot* (bottom row) visual hint. The ground truth is denoted in black and the user's movement is shown in red. Axis units are in mm.

exclude rotation from our ICP transformation as our path's unique characteristics are defined by their angle of rotation. That is, we wanted to see how well users perceived angles in video and excluding rotation allowed us to analyze deviations from angled motions.

Figure 12 shows results on the change in deviation when a user's path is scaled and translated with ICP. On average, participants using the *video screen* condition deviated from the desired path by 25.1mm (mean SD = 7.3mm), while the *video hand* condition faired comparably. Participants using the *Follow Spot* condition showed significantly less deviation at 13.7mm (mean SD = 6.6mm) ($t_{16} = 11.4$, $p < .001$).

Additionally, our results indicate that there was a significant performance difference in orientation of the paths in the *video screen* condition ($F_{[2,\ 147]} = 24.6$, $p < .001$). On average, participants performed angled movements with an average deviation of 43.2mm (SD = 9.3mm), approximately 40% less accurately than flat or vertical movements.

Movement Times

We break down movement measurements into two components: self-guided times for the *3D SG-Arrow* compared to the video conditions and distances ahead or behind a path for each of our visual hints.

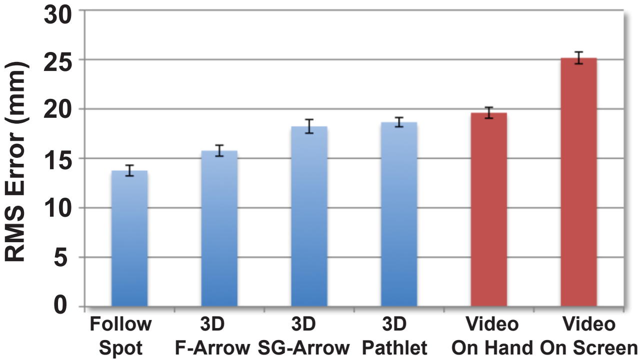

Figure 12. The Iterative Closest Point Algorithm is used to analyze the performance of a user's shape. A user's movement is translated and scaled iteratively until their motion converges to the ideal path. Error bars encode standard error of the mean.

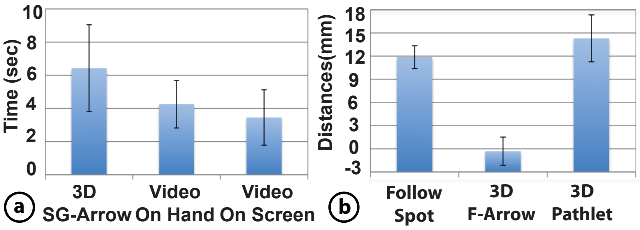

Figure 13. (a) Participants movement times were analyzed in the 3D SG-Arrow condition and compared to video on hand and video on screen and (b) shows participant's average distance behind each projected visualization.

Self-Guided Times

The average movement times across all users and paths for the *3D G-Arrow, video screen,* and *video hand* are visualized in Figure 13(a). With the *3D SG-Arrow*, although participants were able to perform the movements with more accuracy over both video conditions, movement times for video were significantly faster ($F_{[2,\ 447]} = 54.9$, $p < .001$).

On average, participants performed *video screen* movements with a mean of 3.45s (SD = 1.67s), nearly twice as fast as the *3D SG-Arrow*. These results reflect our observation that participant's tendencies were to first see the whole path conveyed on video, where users acquire the gist of the entire movement. In contrast, users with the *3D SG-Arrow* would perform movements in situ, figuring out direction as they moved along the path.

Distance Ahead/Behind Paths

Figure 13(b) displays the average distance (mm) participants were in front, or behind each of the visualizations in the 'following' conditions. To illustrate how participants follow a *3D F-Arrow*, Figure 14 displays a single participant's movement on a circle that is oriented at 45 degrees with the respect to a canonical horizontal X-Y plane.

User Feedback

In the video condition, users were able to quickly perform movements, but often expressed frustration with the lack of

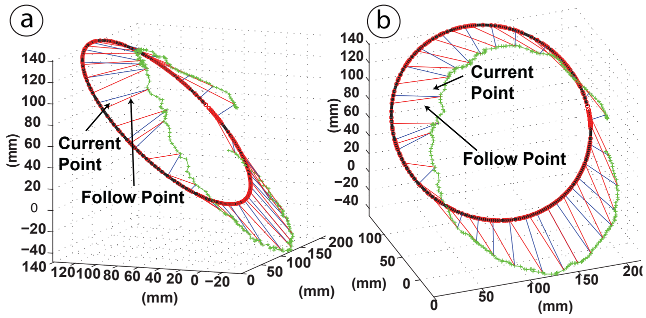

Figure 14. The plots show the same movement from two perspectives of how far in-front or behind a user's hand is compared to the projected visual hint. Green denotes a participant's path, the red line shows the actual position of the hint and the blue line shows the projected point on the path.

feedback. As one participant described for video, "It was harder to reproduce subtle movements, then to follow. It was also harder to judge elevation based on the size of the hand." Importantly with video, users also described the lack of feedforward hints. As one participant said, "With video, you have global features. You just never know what's coming next."

With the *Follow spot* visualization, users commented on the general ease of understanding of the visualization. For example, as a participant explained, "The circle one was simplest, it was only telling you up or down. Less displayed info made it easier." Similarly, another participant noted, "For me, the best visualization was probably the circle with the arrow, as once I was used to the mechanics of it, it became somewhat second nature." With the *3D Pathlet*, users commented on the benefits of knowing what was coming up ahead in the movement. As a participant described, "The feedback was great and I liked seeing where I was going." Although occlusions were not prevalent in all paths, users occasionally commented on a "disappearing red ball." As a result, participants would tend to overshoot a path, as they were unable to see how much of the path they had consumed.

Following and Leading

A majority of our users in our interviews (8/10) said they preferred the 3D *SG-Arrow* over all other visualizations. The ability for users to shape their own tempo was important to their overall satisfaction with the visualization. As a participant noted, "Creating my own tempo made it easier to concentrate on where I was moving." Another participant described, "If I go faster, I feel like I can do it better. Moving at my own speed lets me concentrate on what the system wants me to do. Because it's reacting to me, I can focus on the shape of the path. I didn't have to follow a slower system when I could do better."

DISCUSSION

The results of our study support our approach of guiding users' movements with on-body projections in that users were able to perform more accurately with our system over video. Our reasoning for using an animated video was to adequately control for perspective, tracking error and speed. However, we hypothesize that the reason for the large difference in scale was that we only provided users with a single shadow on a white horizontal plane for visual context. Thus, a more representative measure of accuracy with our hints can be seen in our analysis with ICP, where the *Follow Spot* visual hint did significantly better than all other conditions. In general, users' qualitative feedback also reflected our empirical findings where general comments positively reflected the ease of understanding the hint.

However, this accuracy comes at a cost. With our *3D SG-Arrow*, users were able to accurately guide themselves through a path, but were unable to do so at the same speed as video. One reason for this behavior may be attributed to users being able to see the movement before completing the path, getting the gist of the motion. Although we have highlighted scenarios where users are no longer in the position to look at a video screen (e.g., when they are using body-worn projection systems), a more beneficial scenario for on-body projections may be attained when combined with video. One strategy users can take is to view the gist of the motion and see the visual context in video, and then use our on-body hints to perfect the motions.

Our findings also showed that when the exact same video was moved from the screen to the hand, there was a significant performance difference in scale. While this may be attributed to the lack of visual context, or a change in scale (e.g., the video on the hand was smaller), another possibility could be that users were able to more accurately calibrate for the desired movement of their hand when the video was rendered in the same location as the physical body-part we were attempting to guide.

Surprisingly, at times users would become so immersed in the visual hints that it became unclear to them whether they were moving their hand or if it was the visualization that was moving in space. This reaction reflects similar findings in previous literature [27] where projected light was used to 'trick' viewers into thinking a static car was moving along a road.

Among the visual hints, the *3D F-Arrow* showed the most promising behavior with regard to user's consistently keeping pace with system imposed timing. That is, by simply conveying a sense of the next point along the path through direction alone, users were able to more accurately predict where to move next. Our user opinions suggest that when the task is not rhythmic in nature or requires a fixed speed/accuracy, the most benefit may be obtained by allowing the user to dictate their own pace. This allows users to have the flexibility to decide how they want to interpret and react to the visual hints.

CONCLUSION AND FUTURE WORK

In this paper, we described and evaluated four on-body projected visual hints to help guide a user's movement in mid-air. In addition, we introduced LightGuide, a proof-of concept system that uses an overhead projector and camera to display our visual hints which can replay movements at a user guided or system imposed speed. Our results suggest that users can follow our on-body hints accurately for movement of a single body part in space and can do so at a system controlled speed.

While our chief goal with the present work was to demonstrate that on-body projected hints could be used for movement guidance, we have only tested these visual hints on a user's hand. For example, these visual hints could just as easily be shown on the rest of your arms, torso and legs assuming the rest of the user's body is tracked (e.g. with the Kinect skeletal tracker). Furthermore, we have yet to explore guidance of two or more body parts (e.g. two hands) simultaneously. Another important question moving forward is how to adapt or add rotational visual hints to allow guiding a user's full range of motion.

Finally, there are many fascinating cognitive questions we would like to investigate. For example, does projecting the same visual hint on a screen allow the user to perform the movement just as accurately as projecting the hint on your body? Can on-body visual hints be used to distort the user's sense of their space, allowing us to control their range of motion? While we have helped to define what the design space of projected hints might look like to explore guidance first, we have yet to answer the question of how well users *learn* a particular movement. In addition, we have yet to explore how well users perform the dynamics of a movement, particularly when users are guided through alternating speeds. Our work has allowed us to answer the fundamental question of whether or not on-body projected hints are possible, but many interesting questions lie ahead.

ACKNOWLEDGMENTS

We thank the Microsoft Research VIBE and ASI groups, as well as the 2011 Microsoft Research interns, for helping us improve LightGuide.

REFERENCES

1. Baird, K.M. and Barfield, W. Evaluating the effectiveness of augmented reality displays for a manual assembly task. *Virtual Reality 4*, 4, 1999, 250-259.
2. Banaji, M.R. Stevens, Stanley Smith (1906–73). *Journal of Personality and Social Psychology*, 1994.
3. Bau, O. and Mackay, W.E. OctoPocus: a dynamic guide for learning gesture-based command sets. *Proc. UIST*, 2008, 37–46.
4. Britannica. http://britannica.com/EBchecked/topic/150794/dance-notation.
5. Cao, X. and Balakrishnan, R. Interacting with dynamically defined information spaces using a handheld projector and a pen. *Proc. UIST*, 2006, 225-234.
6. Cao, X., Forlines, C., and Balakrishnan, R. Multi-user interaction using handheld projectors. *Proc. UIST*, 2007, 43-52.
7. Daniel Dementhon, L.D. Model-Based object pose in 25 lines of code. *Proc. of IJCV 15*, 1 (1995), 123-141.
8. Feiner, Steve, Blaire Macintyre, D.S. Knowledge-Based augmented reality. *Commun. ACM 36*, 7, 1993, 53-62.
9. Felzenszwalb, P.F. and Huttenlocher, D.P. *Distance transforms of sampled functions*. 2004.
10. Flagg, M. and Rehg, J.M. Projector guided painting. *Proc. UIST*, 2006, 235-244.
11. Freeman, D., Benko, H., Morris, M.R., and Wigdor, D. ShadowGuides: Visualizations for In-Situ Learning of Multi-Touch and Whole-Hand Gestures. *Proc. ITS*, 2009, 165-172.
12. Harrison, C., Benko, H., Wilson, A.D., and Way, O.M. OmniTouch: Wearable multitouch interaction everywhere. *Proc. UIST*, 2011, 441-450.
13. Harrison, C., Tan, D., and Morris, D. Skinput: Appropriating the body as an input surface. *Proc. CHI*, 2010, 453-462.
14. Holz, C. and Wilson, A. Data miming: inferring spatial object descriptions from human gesture. *Proc. CHI*, 2011, 811–820.
15. Hua, H., Brown, L.D., and Gao, C. Scape: Supporting Stereoscopic Collaboration in Projective Environments. *Proc. CG*, 2004, 66-75.
16. InFocus. http://www.infocus.com.
17. Jacob, R. The perceptual structure of multidimensional input device selection. *Proc. CHI*, 1992, 211-218.
18. Kane, S.K., Avrahami, D., Wobbrock, J.O., and Harrison, B. Bonfire: A Nomadic system for hybrid laptop-tabletop interaction. *Proc. UIST*, 2009, 129-138.
19. Kirk, D. and Stanton Fraser, D. Comparing remote gesture technologies for supporting collaborative physical tasks. *Proc. CHI*, 2006, 1191-1200.
20. Mistry, P. and Maes, P. WUW-wear Ur world: a wearable gestural interface. *Proc. CHI*, 2009, 4111-4116.
21. Motokawa, Y. and Saito, H. Support system for guitar playing using augmented reality display. *Proc. ISMAR*, 2006, 243-244.
22. Neumann, U. and Majoros, A. Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance. *Proc. of VR*, 1998, 4-11.
23. Palmiter, S. and Elkerton, J. An evaluation of animated demonstrations of learning computer-based tasks. *Proc. CHI*, 1991, 257-263.
24. Pinhanez, C. The Everywhere Displays Projector: A Device to create ubiquitous graphical interfaces. *Proc. UbiComp*, 2001, 315-331.
25. Raskar, R., Beardsley, P., Van Baar, J., et al. RFIG Lamps: interacting with a self-describing world via photosensing wireless tags and projectors. *Proc. of SIGGRAPH*, 2004, 406–415.
26. Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., and Fuchs, H. The office of the future: A unified approach to image-based modeling and spatially immersive displays. *Proc. SIGGRAPH*, 1998, 179-188.
27. Raskar, R., Welch, G., Low, K.L., and Bandyopadhyay, D. Shader lamps: Animating real objects with image-based illumination. *Proc. Eurographics*, 2001, 89-102.
28. Rosenthal, S., Kane, S.K., Wobbrock, J.O., and Avrahami, D. Augmenting On-Screen Instructions with Micro-Projected Guides: When it works, and when it fails. *Proc. UbiComp*, 2010, 203-212.
29. SynchronousObjects. http://synchronousobjects.osu.edu.
30. Tan, D.S., Pausch, R., and Hodgins, J. Exploiting the cognitive and social benefits of physically large displays. 2004.
31. Watson, G., Curran, R., Butterfield, J., and Craig, C. The effect of using animated work instructions over text and static graphics when performing a small scale engineering assembly. *Proc. CE*, 2008, 541–550.
32. White, S., Lister, L., and Feiner, S. Visual Hints for tangible gestures in augmented reality. *Proc. ISMAR*, 2007, 1-4.
33. Wilson, A.D. and Benko, H. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. *Proc. UIST*, 2010, 273–282.
34. Zhang, Z. Iterative point matching for registration of free-form curves and surfaces. *JCV 13*, 2 1994, 119-152.