UNIVERSITY OF
BATH

University of Bath

Feet Movement in Desktop 3D Interaction

Adalberto L. Simeone*   Eduardo Velloso†   Jason Alexander‡   Hans Gellersen§

School of Computing and Communications
Lancaster University, Lancaster, United Kingdom

ABSTRACT

In this paper we present an exploratory work on the use of foot movements to support fundamental 3D interaction tasks. Depth cameras such as the Microsoft Kinect are now able to track users' motion unobtrusively, making it possible to draw on the spatial context of gestures and movements to control 3D UIs. Whereas multitouch and mid-air hand gestures have been explored extensively for this purpose, little work has looked at how the same can be accomplished with the feet. We describe the interaction space of foot movements in a seated position and propose applications for such techniques in three-dimensional navigation, selection, manipulation and system control tasks in a 3D modelling context. We explore these applications in a user study and discuss the advantages and disadvantages of this modality for 3D UIs.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Input Devices and Strategies;

1 INTRODUCTION

3D interaction tasks are inherently multi-dimensional, requiring highly expressive input devices capable of providing at least three degrees of freedom [8]. Conventional input devices such as keyboards and mice are not designed for this purpose, so previous research has explored new interfaces that take into account the user's 3D spatial context to facilitate interaction [3]. In this context, our feet can provide an additional interaction space. When their movements are suitably tracked, the input can be combined with the input expressed by our hands to form a more expressive interaction intent.

Technological advancements such as depth cameras and high precision touch-sensitive displays made mid-air and multitouch gestures a reality for a wide audience outside research labs. Whereas much work has explored the advantages of hand gestures for 3D interaction, feet gestures have been underexplored, especially in desktop settings. We aim to contribute in this domain through an exploratory investigation focused on the use of feet movements in the desktop setting. As a testbed environment, we built interaction techniques supporting 3D modelling tasks. Our work is motivated by the fact that even though gestural input can provide the necessary degrees of freedom for 3D interaction, they present challenges for stereoscopic displays, such as breaking the illusion of 3D [4]. This illusion is not hindered by foot interaction, as the feet are out of the user's field of view.

In traditional interaction users often have to switch the function of the mouse between several modes of operation, to cope with the high number of degrees of freedom required. For example, in a 3D modelling application, the user may move an object with the mouse on a 2D plane, use the mouse to rotate the camera and move it along the third axis, again with the mouse. In this example, by delegating

---

*e-mail: a.simeone@lancaster.ac.uk

†e-mail: e.velloso@lancaster.ac.uk

‡e-mail: j.alexander@lancaster.ac.uk

§e-mail: hwg@comp.lancs.ac.uk

Figure 1: Experimental Setup. Participants performed navigation, manipulation, selection and system control tasks using their feet in a stereoscopic 3D environment.

camera control to the feet, users can enhance their performance by parallelising tasks that would normally be performed in sequence.

To investigate the possibilities of feet in 3D interaction, we developed a prototype of a feet tracker based on a Kinect camera mounted under the desk (see Figure 1). We then implemented four applications where the feet support each of the four fundamental 3D task [3] are performed with the feet. To better understand the advantages and disadvantages of such modality, we conducted an exploratory user study with 7 participants where they interacted with our application and provided feedback in an interview.

This paper thus contributes 1) the design of four interaction techniques supporting fundamental 3D tasks; 2) the implementation of a foot-tracking system enabling this interaction style; 3) a discussion of the users' feedback and guidelines for designing such interactions in the future.

2 RELATED WORK

The idea of using the feet to support three-dimensional manipulation was originally proposed by Choi and Ricci, who demonstrated an early application in which walking and leaning actions would rotate a cylinder in different axes [5]. Sangsuriyachot et al. controlled the rotation of a cube displayed on a tabletop with foot gestures tracked by a platform where the user stood [13]. Balakrishnan et al. used pedals and a foot mouse to select modes and control the camera in a 3D modelling application, while the hands manipulated a shape-sensitive tape that controlled an object on the screen [2]. This work builds on top of these by allowing for continuous and discrete input while freeing the feet from additional devices.

One of the primary functions of the feet in daily life is to support us while we walk, which makes navigation in virtual environments an intuitive application for them in HCI. Foot gestures that have been used for this purpose include leaning [6, 10] and walking in place [11]. While we are interested in three-dimensional virtual environments, the focus of this work is on augmenting traditional desktop interaction, where it is not possible to lean or walk.

Previous work that evaluated the physical capabilities of the feet concluded that the feet can be from 1.6 [12] to 2 times [9] as slow as the hands, but this difference can be reduced with practice [7]. Researchers have demonstrated that the feet are suitable for tasks such as mode selection [14], non-accurate spatial tasks [12] and performing secondary tasks whilst the hands are busy [1]. This work aims to draw on these strengths to support the hands in manipulating three-dimensional interfaces.

3 FEET-BASED SUPPORT TO 3D INTERACTION TASKS

Bowman et al. suggest four main categories of 3D interaction tasks [3]: navigation, selection, manipulation and system control. For each category we implemented and evaluated a prototype demonstrating an interaction technique addressing it.
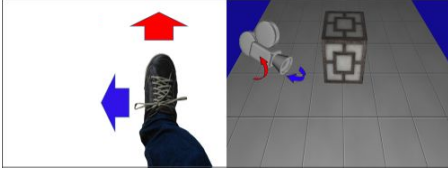
3.1 Navigation

Figure 2: Camera orbit with the feet. The horizontal offset (blue) changes the azimuthal angle and the vertical offset (red) changes the elevation angle around the currently selected object.

In 3D modelling applications, users can manipulate the camera viewpoint by controlling the position and orientation of the camera. Common camera controls include dolly, roll, truck, orbit and pan. In conventional setups, this is usually achieved by holding a modifier key and dragging the mouse. This forces the mouse to perform multiple functions, multiplexed over time.

In our implementation, the user moves the dominant foot to affect the camera's azimuthal and polar angles (see Figure 2). Foot tracking is relative to the starting foot location. The tracking is controlled by a key on the keyboard that toggles between an enabled and disabled state. Moving the foot causes the camera to orbit around the currently selected object. Horizontal movements affect the azimuthal angle, while vertical movements affect its polar angle. Using the foot to control the camera movement leaves the hands free to perform other tasks requiring higher precision. For example, positioning an object so that it matches to visual features in the environment that need to be observed from multiple orientations, e.g. an architectural application where the user needs to place a tree so that it does not occlude house windows. Foot-based camera manipulation can support these tasks by allowing users not to interrupt their interaction flow, performing in parallel tasks that would have been performed in sequence.

During pilot testing we observed that mapping camera XY-trucking/dolling to the non-dominant foot can easily lead to fatigue. Two mapping choices are possible: relative or absolute. With relative mappings, users would need to perform multiple forward/backward or right/left movements with their foot: a first movement causes the camera to advance; a second movement is needed to return to the starting position (e.g. by not tracking it while it is lifted) and perform another movement. Depending on the size of the environment and the rate of change associated to this movement, traversing it would lead to fatigue as multiple back and forth movements would need to be performed.

Absolute mappings, i.e. mapping the feet-tracked area to the whole environment so that moving the foot between the boundaries would result in placing the camera in the corresponding location. For this to be feasible, users would need to see an on-screen world-in-miniature representation of the environment. By visualizing the position of their non-dominant foot through a pointer in this representation, users would have a better understanding of the location they are about to move the camera to. We believe this approach to have more potential and we will explore it in future work.

3.2 Object manipulation

We wanted to explore whether feet-based mappings could be used to provide a way to manipulate an object while the user focuses his
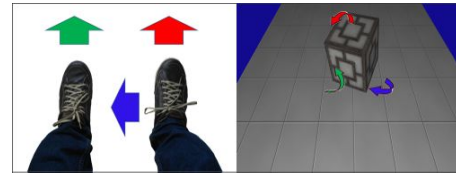

Figure 3: Object rotation with the feet. The horizontal and vertical offset of the right foot affect the object's yaw (blue) and pitch (red), respectively. The vertical offset of the left foot affects roll (green).

hands on tasks requiring higher precision. Analogously to the previous application, we mapped yaw and pitch to horizontal and vertical movements of the dominant foot, respectively (see Figure 3). In addition, roll was controlled by using horizontal movements by the non-dominant foot. We chose to map rotation to the feet as similar considerations as those made in the previous example can be applied on their use on manipulations affecting object translation. This can be combined with mouse and the keyboard to provide a full 6DOF manipulation alternative to specialistic input devices, so that the hands control translation while feet are mapped to rotation.
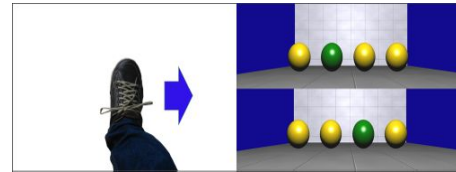
3.3 Selection

Figure 4: Object selection with the feet. The user iterates over objects by performing a swiping gesture either to the right or to the left.

Foot input can also be used to express discrete inputs by means of foot gestures, such as a swipe. A foot-based swipe gesture is a rapid gesture performed by the user either to the right or to the left of the starting position (see Figure 4). Foot swipes can also be used vertically. The gesture is detected by analysing the velocity of the movement and comparing it to an empirically determined threshold. A cooldown period of 250ms avoids unintended multiple gestures. During pilot testing we observed that horizontal swipes are more comfortable than vertical swipes. Indeed, horizontal foot swipes can be achieved by keeping the foot still and pivoting it left and right, whereas vertical swipes require the user to drag the foot forward or backward. In our implementation, horizontal swipes allow users to switch the currently selected object to the one whose 2D screen projection is to the right or to the left of the starting one.

This technique can be used in a wide range of scenarios where users would normally have to switch between a manipulation and a selection mode. In a modelling application, users need to move from object to object to perform different operations on them. Foot-swipes provide a convenient alternative for changing the currently selected object, without breaking the interaction flow. Foot-swipes can also be implemented to enable semantic selection within an object (i.e. selecting vertices, edges or faces) by mapping each foot to one of these two selection operations.

3.4 System Control

Finally, we designed a "radial foot menu" divided into four quadrants, each associated to a different colour parameter (e.g.: hue, saturation, brightness and alpha, see Figure 5). The dominant foot controlled a slider affecting the currently selected function from the menu, while the non-dominant foot selects the property to be
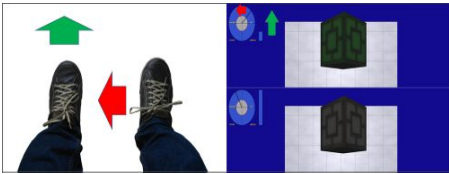
Figure 5: Radial foot menu. The position of the right foot selects the parameter to be controlled and the vertical offset of the left foot selects the value, in this case the saturation of the object's texture.

changed. Adjusting material properties is a task that requires the user to visualise the effect of the changes on the model it is applied to. If further changes are necessary, the user has to return to the editing environment and make adjustments. Foot-based movements can support this task by allowing users to visualise the effect that parameter adjustments can have on the model they are working on. In this way, the hands are free to manipulate the object and observe it from different viewpoints while continuing to make adjustments. Once satisfied with the results, a modifier fey can disable foot-tracking and finalise the operation.

4 SYSTEM OVERVIEW

To investigate the proposed techniques, we implemented a foot tracker prototype that extends the current capabilities of the Kinect sensor. System calibration is not user dependent, so it only needs to be calibrated when the setup changes. In this step the system takes a snapshot of the background without the user's feet. The user then draws on the image of the background a rectangle that will define the active area of the tracker. We use three corners of this rectangle (lower left, upper left, and lower right) to extract the floor plane and define the new basis for the coordinate system. The axes of the new coordinate system are formed by the normalised vectors parallel to the two edges formed by the lower left vertex with its adjacent vertices, and their cross-product. This vertex becomes the new origin.

We then convert the data of subsequent frames to this new coordinate system. This change of basis allows more freedom in the camera positioning, as the system knows the position and orientation of the camera relative to the floor, allowing the camera to be positioned in a more suitable height or orientation depending on the physical configuration of the desk.

To interact with the system, the user places their feet in the tracked area (Figure 6 (a)). For every depth frame captured by the system (b), we convert the coordinate system (c) and subtract the background frame, hence isolating the user's legs and feet (d). We
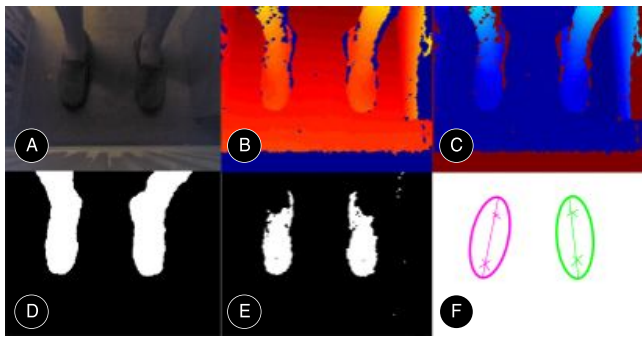

Figure 6: Feet detection algorithm. (a) Colour image, (b) Raw depth, relative to the camera, (c) Depth, relative to the floor, (d) Legs isolated from the background, (e) Feet isolated from the legs, (f) Ellipses fitted to the mask. The foci of the ellipse are used as the joint positions for the foot and ankle.

then isolate the feet from the legs by thresholding the depth data at 0.1m above the floor (e). We finally fit ellipses that have the same second moment to the thresholded image of each foot and use their foci as the positions for the feet and ankles (f).

A problem commonly encountered in vision-based systems is lighting conditions. In this regard, the area under the desk is particularly problematic as it is often very poorly lit. Because our approach does not rely on the colour feed from the camera, it works well in different lighting conditions. Another advantage is that we do not take previous tracking states into account, so if the system loses track of the feet momentarily (e.g. if the user leaves his desk), it can instantly recover from erroneous states. Also, it does not need to be calibrated for different users. Our algorithm was implemented in Matlab and runs at 20 frames per second on an Intel i7-3930K CPU at 3.20GHz with 16GB of RAM and an nVidia Quadro 4000.

5 USER STUDY

To better understand the advantages and disadvantages of such modality and to get user feedback on the interactions it affords, we conducted an informal user study with 7 participants (3M/4F), with ages varying from 19 to 30 (mean 26.3). All participants had prior experiences with 3D movies; 75% reported to regularly drive a car; none had any prior experience with a foot mouse or similarly foot-operated device. Our experimental setup (Figure 1) used an Asus VG278h 3D monitor to display the applications. The stereoscopic effect was obtained using DirectX 11.1's native support by computing the stereo pairs, without resorting to an automatic driver implementation. We used a custom 3D engine using SharpDX's DirectX11 C# port.

Participants were asked to use the four applications previously described and to provide feedback in a questionnaire and in a structured interview. In the *camera manipulation* task, they were asked to rotate the camera to read a number written on a hidden side of a cube; in the *object manipulation* task, users were asked to rotate a cube in its 3 axes so that the side containing the number was visible and upright (the orientation of the cube was reset so that the number was randomly hidden again); in the *selection* task, users were asked to select specific objects by foot-swiping left and right to iterate over a group of spheres in a 3D environment; in the *system control* task, users were asked to select the *saturation* property in a radial menu and change its value in order to make the selected object greyscale. After each task, we conducted an interview and participants filled in a Likert scale questionnaire. The experiment took approximately half an hour for each participant.

5.1 Results

Participants were able to quickly grasp how to operate the four different techniques. They reported low signs of frustration ($M = 2, SD = 1.05$, from 1 – very low to 5 – very high) and a low cognitive demand ($M = 1.93, SD = 1.12$). When asked whether they would be to coordinate their hands with one or both feet, they all replied positively ($M = 4.39, SD = 0.69$), with some distinctions. Two participants shared the opinion that feet are better used for coarse control tasks such as camera manipulation, whereas object rotation was deemed to require more precision. Thus, using feet-based movements for secondary tasks appears a more promising direction, as this leaves the hands free for fine-grained manipulations.

Participants observed that they were able to use their feet so to get the desired results. They confirmed our assumptions that pivoting is more comfortable than sliding one's foot across the floor. The floor surface used in our experiment consisted of carpet which, as noted by participant #1, introduced some friction. The main issues we observed in this regard concerned the difficulty in estimating whether feet were still in the tracked area and how to avoid unwanted movements, which we address in the next section.

On the suitability of foot-movements for 3D interaction, partici-pant #1 said: *"[They are] good for continuous 3D interactions that don't have to be perfect, you can do many different things at the same time"*. Participant #6 thought that *"camera manipulation felt easier than rotating the object, as I think that it is better when the object that you are working on is fixed; object rotation might be easier to perform by having each foot control one axis"*.

On the ease of use and comfort:, participant #2 said *"foot move-ments might help the health of the user, as it adds activity during computer work"*. Participant #4 stated that: *"simpler feet move-ments are easier to perform, so I would be able to coordinate [foot movements] better with hands movement"*. Participant #1 ex-pressed some concerns about unwanted movements: *"if i move my feet accidentally, and this is something I often do, it would proba-bly trigger some unwanted stuff. Also, if I have to move too much with my legs, it might be too exhausting"*. Participant #2 stated that *"rotation should not require the user to twist the foot too much"*.

6 DISCUSSION

During our study, we collected a number of observations and sug-gestions that we discuss in the following paragraphs.

Limitations. In our implementation, feet tracking could be en-abled or disabled by pressing an associated key. Once enabled, par-ticipants explored the interaction techniques without reverting to a non-tracked state. This introduced a form of the "Midas Touch" problem, where participants accidentally performed unintentional movements which affected the state of the system, causing some frustration. In a real application scenario, users need an unobtru-sive way to enable and disable the tracking. The keyboard key is one of the options we have explored.

An alternative approach consists in dividing the floor between tracking and rest areas. Only when feet enter in the tracked area, the system is affected. Further options consist in using a foot-tapping gesture (i.e. lifting one's toes and tapping the floor) to toggle be-tween tracked and untracked states. The microphone in the Kinect can be used to detect the audible impact and support the detection of the gesture. Another alternative consists in raising one's toe and pivoting the foot to enable tracking and, at the same time, affect the environment.

Estimating the tracked area's boundaries was another of the is-sues we observed. Participants were not aware of where exactly these boundaries were. We believe that this can be addressed by displaying a warning icon when the user is in proximity of the boundaries of the tracked area. Another issue we noticed is that, due to limitations in the depth-sensing technology, it became hard to distinguish when both feet were close together.

Design Guidelines. As we have previously highlighted, foot-based UIs can play an effective role in supporting primary tasks by delegating secondary ones to the feet. Users can use their hands to perform fine-grained manipulations with other input de-vices while feet can be used to control those aspects that would otherwise require frequent mode switching (i.e. between manipu-lation and system-control modalities). To avoid confusion, applica-tions should inform users when tracking is active and provide the option to undo feet actions directly through feet gestures.

Hand-Foot Coordination. Based on our observations, we be-lieve that in order to minimise users' cognitive burden, each modal-ity should be assigned to actions that can be performed indepen-dently, e.g. assigning hands to manipulation and feet to camera rotation. Mappings that require integration of DOFs split across modalities should be avoided. The extents to which foot and hands are able to work in parallel on different aspects of a common goal will need to be evaluated in future studies. Visualizing feedback of how feet are affecting the system might improve coordination.

Fatigue. In order to maximise the potential of this novel ap-proach to 3D interaction, we believe it is important to focus on the

design of foot mappings that are comfortable to the users. We have observed how horizontal foot movements were preferred by par-ticipants, whereas vertical foot movements were deemed to have higher potential for causing fatigue. In addition, pivoting move-ments were preferred over sliding movements. Thus, it appears that *in-place* movements hold more potential than *dragging* one's feet across the floor. Friction hinders the latter, while pivoting appears to be easier to control and requires less effort. Our observations indicate that short foot motions, in terms of amplitude and actual movement, were the most comfortable and less fatiguing ones.

7 CONCLUSION

In this paper we have presented an exploration of feet-based inter-action techniques supporting 3D tasks. Through an informal user study we determined that foot-movements are easy to learn and can be used to perform 3D manipulations or affect the system. We de-scribed how it is possible to address the limitations emerged during the study. In future work we will extend the system to detect a richer set of foot-gestures and perform an in-depth study evaluating coordination of hands and feet together in a 3D environment.

REFERENCES

[1] J. Alexander, T. Han, W. Judd, P. Irani, and S. Subramanian. Putting your Best Foot Forward: Investigating Real-world Mappings for Foot-based Gestures. In *Proc. CHI '12*, pages 1229–1238. ACM, 2012.

[2] R. Balakrishnan, G. Fitzmaurice, G. Kurtenbach, K. Singh, and K. S. East. Exploring interactive curve and surface manipulation using a bend and twist sensitive input strip. In *Proc. I3D '99*, pages 111–118. ACM, 1999.

[3] D. A. Bowman, E. Kruijff, J. J. LaViola Jr, and I. Poupyrev. An intro-duction to 3-d user interface design. *Presence-Teleop. Virt.*, 10(1):96–108, 2001.

[4] G. Bruder, F. Steinicke, and W. Sturzlinger. To touch or not to touch?: comparing 2d touch and 3d mid-air interaction on stereoscopic table-top surfaces. In *Proc. SUI '13*, pages 9–16. ACM, 2013.

[5] I. Choi and C. Ricci. Foot-mounted gesture detection and its appli-cation in virtual environments. In *Proc. SCM '97*, pages 4248–4253. IEEE, 1997.

[6] F. Daiber, J. Schöning, and A. Krüger. Whole body interaction with geospatial data. In *Proc. SG '09*, pages 81–92. Springer, 2009.

[7] F. P. Garcia and K.-P. L. Vu. Effects of Practice with Foot-and Hand-Operated Secondary Input Devices on Performance of a Word-Processing Task. In *Proc. HCII '09*, pages 505–514. Springer, 2009.

[8] K. P. Herndon, A. van Dam, and M. Gleicher. The challenges of 3D interaction: a CHI '94 workshop. *SIGCHI Bull.*, 26(4):36–43, 1994.

[9] E. R. Hoffmann. A comparison of hand and foot movement times. *Ergonomics*, 34(4):397–406, 1991.

[10] J. J. LaViola Jr, D. A. Feliz, D. F. Keefe, R. C. Zeleznik, and J. L. Jr. Hands-free multi-scale navigation in virtual environments. In *Proc. I3D '01*, pages 9–15. ACM, 2001.

[11] N. Nilsson, S. Serafin, M. Laursen, K. Pedersen, E. Sikstrom, and R. Nordahl. Tapping-in-place: Increasing the naturalness of immer-sive walking-in-place locomotion through novel gestural input. In *Proc. 3DUI '13*, pages 31–38. IEEE, 2013.

[12] T. Pakkanen and R. Raisamo. Appropriateness of foot interaction for non-accurate spatial tasks. In *CHI '04 EA*, pages 1123–1126. ACM, 2004.

[13] N. Sangsuriyachot, H. Mi, and M. Sugimoto. Novel interaction tech-niques by combining hand and foot gestures on tabletop environments. In *Proc. ITS '11*, pages 268–269. ACM, 2011.

[14] A. J. Sellen, G. Kurtenbach, and W. A. S. Buxton. The prevention of mode errors through sensory feedback. *Human-Computer Interaction*, 7(2):141–164, 1992.