

Project Natal as an off-the-shelf gesture capture system for the Dancing the Earth: a mixed-reality science center program based on whole-body interaction and virtual docents

Introduction:

Throughout the Fall 2009 course Interactive and Applied Storytelling we have been exploring the many technical and social aspects of implementing a mixed-reality science center program called Dancing the Earth. This paper focusses on one particular aspect of the project, namely using Project NATAL's gesture capture hardware to make the whole-body interactions of the Dancing the Earth program possible.

Why whole-body interaction?:

Before describing Project NATAL in greater detail, it is important to understand why whole-body interaction is even a goal of the Dancing the Earth program (DE). Over the last decade researchers have begun focussing on the role mixed-reality and virtual environments can play in academic settings. Of particular interest to our project is research showing that elementary school aged children construct knowledge through their experience of the world. More pointedly, children need to learn through the uses of as many of their five sense as possible and through physical activity (Winkler, Herczeg & Kritzenberger, 2002).

The DE program seeks to embody these findings by constructing a gesture-controlled, interactive wall- and floor-projected virtual scenario. One which, like other augmented reality museum instillations (Kondo 2006), posits that immersive environments can demonstrably improve the participant's ability to identify a problem, formulate a hypothesis, and test it. Doing so utilizing as many of the participant's senses as possible should demonstrate a greater degree of knowledge

acquisition than more traditional museum exhibits.

Why collaborative?:

Beyond the importance of learning through the whole-body interactions described above, a powerful driver of the cognitive processes comes through the interactions which take place in collaborative learning environment. Research in this area points out that the collaborative nature of the DE program will help facilitate an important learning synergy in the participants (Schaf, Müller, Bruns, Pereira & Erbe, 2009). Humans are a communal species and learning in a shared collaborative environment has show to be particularly more effective that learning in isolation.

The DE program takes this a step further and offers a rich mixed-reality environment for these important social interactions to take place. The interaction with the virtual docent, for instance, overcomes the limitations of more conventional video-based telecommunication modalities, and simulates a face-to-face meeting, where gestures, gaze awareness, and realistic responsive images create an immersive experience despite the physical distance separating the participants and the docent (Sang-Yup, Hyoung-Gon, & Myotaeg 2006).

A Hardware Link - Project NATAL:

I believe that Microsoft's Project NATAL is a piece of hardware uniquely qualified to provide a link between the two concepts mentioned above: Whole-body interaction in a collaborative environment.

Project NATAL (PN) is the code name of a controller-free gaming and entertainment

experience under development by Microsoft for their current and future XBox 360 gaming systems.



The PN hardware interface combines an RGB camera, an IR depth sensor, a multiarray microphone and a custom processor running the drivers that make the hardware magic happen. With the ability to track up to 48 points on the human body, PN enables 3D Full-body gesture capture (Lange & Edwards, 2009). Thereby, controlling virtual environments without the need of a physical controller.

The Hardware Spec:

The RGB Camera: Provides the unit not only with a means of video and still image input to the system, it also give the system the ability to recognizes faces.

The Depth Perception Sensor: Infrared is projected from the unit and a monochrome CMOS chip senses the reflected IR. This allows PN to see the surrounding environment in 3D. Even in low-light environments the IR/CMOS combo can work flawlessly.

Multiarray microphone: Provides not only audio input into the system, but more importantly has the ability to audio-locate the users in the room. Additionally built into the system is an ambient noise canceling filter.

Custom Processing: The benefit of having this on-board processing is that thePNunit could be hacked to work with other gaming/computing platforms without loosing functional capabilities.

Dancing Earth x PN implementation:

As it stands, there is no motion or form of interaction proposed in our class' Wood Stork scenario that could not be handled by the PN unit. That is however, with the caveat that the PN units will work best in a scenario in which users do not roam the exhibit floor but are restricted to certain hot zones or active zones. From these zones they would control their avatars in the virtual environments projected on the walls and their user specific information would be displayed around them via floor projection systems. Perhaps a PN unit, or several, could be placed above the users and configured to enable free roaming gesture capture, but it would negate the facial recognition aspects as the units would be looking at the top of the users heads. In conjunction with other traditionally placed PN units, however, this issue could be alleviated.

A great benefit of the unit is the freedom it offers museum administrators. There is no need for participants to handle hardware that needs upkeep and monitoring. Using controllers like the Wii remotes that can be dropped, stolen or misplaced creates an additional layer of concern for the volunteers administering the scenario. Additionally, there is no need for users to tote sticks with florescent ends, wear reflective gear or be

restricted to only a handful of big full limb motions.

An area where I believe the PN gesture capture system provides a perfect fit is in the multiple high to low light scenarios of the exhibit. My impression of the proposed exhibit floor is that it will be a predominately mixed-light environment. While many mocap systems can handle lower-light environments, the fact that PN uses infrared for its gesture capture means that it will adapt to any lighting conditions. From scenario to scenario the unit will not need to be calibrated. For example a scenario about Florida's sunny coastline could easily follow a darker scenario set in the dark shadowy cypress swamps of the Okefenokee.

One of the remarkable capabilities of the unit is its ability to automatically calibrate its depth sensors to accommodate existing objects in its visual field. In a DE application with multiple PN units at work auto-sensing could be an issue. Thankfully the sensing range of the IR/CMOS sensor depth is adjustable. With a little tweaking we can have multiple units running with predefined action zones without having units picking up the motion of other participants (Wilson & Buchanan, 2009). Additionally the sensing range can be change per game-play scenario. Therefore the active zone can change with each new scenario without on-site volunteers having to interface with the system.

A single PN unit can simultaneously track four users in the active zone (Gibson, Ellie 2009). While our class' proposed DE scenario does not call out for four users to participate in the same active zone, other scenarios could benefit from this functionality. As it stands one PN unit could be used to track two users per wall projection. Each of the three walls of the environment would be assigned one PN unit and track two

users each. With the ability to track up to 48 skeletal points, including fingers, much more intricate actions can be captured than we have even proposed thus far with our stork scenario.

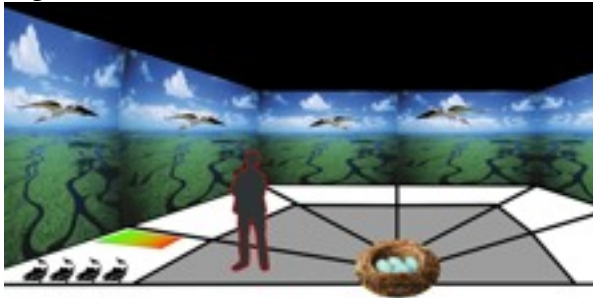
With the ability to do multi-user facial and vocal recognition, the PN hardware can be used to log individual users and keep relevant statistics that would otherwise be tedious to record (Totilo, 2009). For example the first time a user participates in a scenario the facial recognition system would log them and which scenario they participated in. In future scenarios they will be recognized by the system and docent can verbally acknowledge this fact, regardless of whether the virtual docent actually was connected the last time the visitor was at the exhibit. Thereby lending some continuity to the user experience. I.e. the docent could even say "Oh I see we have some returning Jr. Rangers. That is excellent. You two will be able to help me demonstrate to the trainees".

Depending on how things are designed by the audio team it would be interesting to experiment with PN's ability to do audio-location. Why I think this could be useful is that it would alleviate the need for the docent to control his/her avatar to follow the face of whoever is speaking. By programming their avatar to automatically look to wherever the PN interface is sensing the guest speaking from, the virtual docent could focus their attention on the more important task of controlling the scenario and engaging with the guests.

A Wood Stork Scenario Walkthrough:

To roll this into a tangible walkthrough let's look at how the users will interact with the PN interface during our intro and training phases of the Wood Stork scenario.

Figure 1.



To understand the environmental layout look at the image above. The above configuration would provide an environment where only 3 PN units could capture the full-body interactive motions of up to 6 participants (2 per wall). For other scenarios those 3 units could handle up to 12 simultaneous participants. The PN units would be placed at either the top or bottom border of the wall screens where the individual viewing areas intersect.

As the On Site Volunteer (OSV) organizes the next scenario he or she will instruct the six participants to move into one of the six grey sections which represent the participants “active zones”. The PN units will have their depth sensors to stop just past the active zone so as to not bleed into the other participants “active zones”.

Once the users are in their zones the OSV will ask them to face their respective wall projections at which points the PN will do a quick facial recognition. Then as the OSV introduces the participants to the Virtual Docent (VD) each participant could say hello to the VD and state their name, thus giving the PN system a chance to run voice recognition on each user.

“Learning to fly without wings”:

What I love about the PN’s abilities is that we can give participants more movement options that have previously envisioned. For instance

I have devised several new movements to make navigation more intuitive and less clunky that we originally proposed.

Flapping: Participants can flap their arms like a bird to 1.) *gain altitude* and 2.) *gain speed*.

Gliding: By holding their arms out the participants can make their bird avatar glide. They will slowly *lose altitude*, but currently we don’t have a way for the birds to descend or ascend.

Steering: By placing one wing higher than the other the participants can steer their avatars to the left or right, and even in full circles.

Navigating by sight: Because of the advanced nature of the PN units I would like to propose a means of navigating that tracks the direction the participants head is facing. This could help provide more precise descents when, for instance, a participant’s avatar is at max altitude and needs to make a steep descent to the edge of a pond. NOTE: Otherwise they will have to glide in circles until the bird descends all the way down. While this is how birds in the wild descend it could be potentially very disorienting for the participant to have their wall projecting whirling in circles.

Foraging:

The OSV will show the participants that when they are in the shallows of the river or ponds that by closing their arms like a Gator chomp they can eat fish. Since I am proposing to not use force feedback hardware controllers like the Wii remotes or similar devices the fishing techniques have changed some.

While at the water’s edge the participants view changes from a 3rd person over the

shoulder view to a 1st person view where they only see their beak out in front of them. While Wood Storks do eat by a tactile feeding mechanism it isn't crucial that we simulate this in order that the participants get the scientific take-away of the scenario.

There are two methods then for fishing. Either we can have participants actually see the fish below the water and clap their arms to catch them, or like other birds of prey, the participants could see a splash in the water where a fish has surfaced and if they bite there within a certain amount of time they could catch a fish.

The sensitivity of the PN unit would allow the user to catch fish not just directly in front of them, but anywhere within arm's (err beak's) length. Rather than just waiting for a controller in their hands to buzz, they are able to take an active, physical role in the foraging.

The Shuffle: While the participants are waiting for a splash the OSV will mention to them that if they shuffle their feet they will stir up more fish. The second they start doing so the PN unit will pick up this unit and the "splash/second" ratio will go up and the participants will catch more fish whenever they are actively shuffling.

The HUD: On the floor in front of each participants active zone can show the pertinent data for the participant. The avatar's energy bar and fish count are displayed in the example image above, but this area could be used to display any individual participant's key data for a scenario.

Feeding the Young:

The example environment of the image above the central nest serves as a central place for the OSV to coordinate the participants but

does not represent the actual location of any one's virtual nest. When the participants begin their session they will take control of their Wood Stork avatar while it is perched in its respective virtual nest. While the PN unit could track the participants if they turned around to feed the chicks in the center nest, the participants would not be able to actually see their own flight as they would have turned their back on the wall projection.

I propose that on screen prompts (arrows) in conjunction with a floor displayed mini map in their HUD can serve to help them navigate back to their virtual nests without having to converge on a center of the room floor projected nest. While the PN units could be placed above the participants and the scenario could be floor and wall projected, there is always going to be issues with how to hand the borders between wall and floor. My proposal eliminates this by never having the participant interact with the floor projection images as they are only there to provide non-interactive story elements and data.

Therefore the participants will use the full-body motions of flight and navigation to return to their nests once their stomach meter in the HUD is full. As they arrive at the nest they will return to a first person view again and be able to see their beaks out in front of them. Then by placing their arms out in front and aiming them at the cheeping chicks they can make a barfing noise that will be picked up by the PN voice recognition system and a fish will be disgorged into the hungry chicks open mouth.

Conclusion:

It is clear that the Project NATAL system being developed by Microsoft could serve as an excellent off-the-shelf hardware solution to handle the full-body gesture capture needs of our Dancing the Earth program.

Its ability to scale to handle multiple users, provides a learning framework where the synergies of leaning in a collaborative environment can be maximized. From 2 participants to 12, PN can handle full-body motion capture for a exhibit full of participants.

Additionally PN would enable the Dancing the Earth exhibit to provide an unparalleled interactive experience, where the knowledge acquisition benefits of learning through full-body interaction could be given its fullest expression.

Finally it is important to note that the potentially large community of developers for the PN hardware will give us access to a broader source of support than were we to try and develop a proprietary mo-cap system. If we lost an engineer on the project, it would be less of a headache to find a replacement than to have to bring someone up to speed on a proprietary system.

In closing I would strongly suggest that the research team take a serious look at the Project NATAL as a possible motion capture system for implementation in the Dancing the Earth exhibit. Even if it is decided that PN is not a perfect fit, there are many lessons disclosed in this paper that should be included in any system buildout for this project.

REFERENCES:

Gibson, Ellie (2009). "[E3: Post-Natal Discussion](http://www.eurogamer.net/articles/e3-post-natal-discussion-interview)". *Eurogamer*. Eurogamer Network. pp. 1–2. <http://www.eurogamer.net/articles/e3-post-natal-discussion-interview>. Retrieved 2009-11-09.

Kondo, T. (2006). Augmented Learning Environment using Mixed Reality Technology. In T. Reeves & S. Yamashita (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2006* (pp. 83-87). Chesapeake, VA: AACE.

Lange, E. & Edwards, N. (2009). Project Natal Fact Sheet. Microsoft. Retrieved 11/03/2009 from: <http://download.microsoft.com/download/A/4/A/A4A457B3-DF5D-4BF2-AD4E-963454BA0BCC/ProjectNatalFactSheetMay09.zip>

Sang-Yup, L., Ahn, S., Hyoung-Gon, K., & Myotaeg, L. (2006). Real-time 3D video avatar in mixed reality: An implementation for immersive telecommunication. *Simulation & Gaming*, 37(4), 491-506. doi:10.1177/1046878106293679.

Schaf, F., Müller, D., Bruns, F., Pereira, C., & Erbe, H. (2009). Collaborative learning and engineering workspaces. *Annual Reviews in Control*, 33(2), 246-252. doi:10.1016/j.arcontrol.2009.05.002.

Totilo, Stephen (2009). "[Microsoft: Project Natal Can Support Multiple Players, See Fingers](http://kotaku.com/5279531/)". *Kotaku*. Gawker Media. <http://kotaku.com/5279531/>. Retrieved 2009-11-06.

Wilson, Mark; Buchanan, Matt (2009). "[Testing Project Natal: We Touched the Intangible](http://gizmodo.com/5277954/testing-project-natal-we-touched-the-intangible)". *Gizmodo*. Gawker Media. <http://gizmodo.com/5277954/testing-project-natal-we-touched-the-intangible>. Retrieved 2009-11-06.

Winkler, T., Herczeg, M. & Kritzenberger, H. (2002). Mixed Reality Environments as Collaborative and Constructive Learning Spaces for Elementary School Children. In P. Barker & S. Rebelsky (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2002* (pp. 1034-1039). Chesapeake, VA: AACE.