The KidsRoom

∽ Aaron F. Bobick, Stephen S. Intille,
James W. Davis, Freedom Baird, Claudio
S. Pinhanez, Lee W. Campbell,
Yuri A. Ivanov, Arjan Schütte,
and Andrew Wilson

Computer vision sensing technologies turn a child's bedroom into a dreamy wonderland.

The KidsRoom is a fully automated and interactive narrative playspace for children developed at the MIT Media Laboratory. Built to explore the design of perceptually based interactive interfaces, the Kids-Room uses computer vision action recognition simultaneously with computerized control of images, video, light, music, sound, and narration to guide children through a storybook adventure. Unlike most previous work in interactive environments, the Kids-Room does not require people in the space to wear any special clothing or hardware, and the KidsRoom can accommodate up to four people simultaneously. The system was designed to use computational perception to keep most interaction in the real, physical space even as participants interacted with virtual characters and scenes.

The KidsRoom, designed in the spirit of several popular children's books, is an interactive child's bedroom that stimulates imagination by responding to actions with images and sound to transform itself into a storybook world. Two of the bedroom walls resemble the real walls in a child's room, complete with real furniture, posters, and windows. The other two walls are large, back-projected video screens used to transform the appearance of the room environment. Four speakers and one amplifier project steerable sound effects, music, and narration into the space. Three video cameras overlooking the space provide input to computer vision people-tracking and action recognition algorithms. Computer-controlled theatrical lighting illuminates the space, and a microphone detects the volume of enthusiastic screams. The room is fully automated.

During the story, children interact with objects in the room, with one another, and with virtual creatures projected onto the walls. Perceptual recognition makes it possible for the room to respond to the physical actions of the children by appropriately moving the story forward thereby creating a compelling interactive narrative experience. Conversely, the narrative context of the story makes it easier to develop context-dependent (and therefore more robust) action recognition algorithms.

The story developed for the KidsRoom begins with a normal-looking bedroom. Children enter after being told to find out the magic word by asking the talking furniture that speaks when approached. When the children scream the magic word loudly, sounds and images transform the room into a mystical forest. The story narration prods the children to stay in a group and follow a path to a river (see the stone path (a) in the figure). Along the way, they encounter roaring monsters and must hide behind the bed to make the roars subside. After a short walk, the children reach the river world, and the narrator informs them the bed has become a magic boat that will take them on an adventure. The children climb on the "boat" and paddle to make it move, which is represented by images of the river flowing by on the screens. To avoid obstacles in the river, the children must row collaboratively on the appropriate side of the bed. Finally, the children reach the monster world. The monsters appear and teach the children some dance steps, and then the monsters mimic the children as the children perform these steps. The story ends when an insistent, motherly voice off in the distance urges the children to return to bed, at which point the room transforms back to a normal bedroom. A typical interaction runs nearly 12 minutes.

Throughout the adventure, the computer system tracks the positions of the movable bed and up to four children. The system detects and responds to events like "Is everyone on the bed?" "Is everyone near the chest?" "Are the children in a group?" and "Are the children following the path?" The music, sound, and narrative of the story change depending upon what the children are doing. For example, if the children fail to get on the bed, characters in the story encourage them to do so. The vision systems use the context established by the story (for example, that everyone is on the bed) for robust initialization and performance. Although the storyline is linear, the room continually reacts to the children's actions, giving the environment an interactive feel. During the river scene, the vision system determines the side of the bed with the highest motion energy and uses this information to "steer" the bed as the children use their arms to row down the virtual river. In the monster world, the still-frame animated cartoon monsters teach the children four different dance moves (for example, "spin around like a top"), after which the children can

perform any step. The vision system is trained to recognize these dance moves, which then triggers the corresponding animations of the monsters with encouraging character narrations. When the vision processing requires constraints (for example, people in certain positions), they were built naturally into the storyline. For instance, the monsters tell the kids to stand on particular rugs "so's we can see ya;" this storyline device actually ensures that each camera has a nonoccluded view of each child.

The KidsRoom demonstrated that nonencumbering, computer–vision sensing technologies can be used to automatically create new types of physical interactive experiences in real environments by integrating sensing and narrative control. We believe the KidsRoom is the first multiperson, fully automated, interactive, narrative playspace ever constructed, and the experience we acquired designing and building the space has allowed us to identify some major questions and to propose a few solutions to simplify construction of more complex spaces in the future. **C**

(a) A view of the KidsRoom showing the two projection screens and the movable bed.

(b) A child and mother rowing the boat together. Rowing was detected using story context and motion energy.

REFERENCES
1. A.F. Bobick, S.S. Intille, J.W. Davis, F. Baird, L.W. Campbell, Y. Ivanov, C.S. Pinhanez, A. Schütte, and A. Wilson. The KidsRoom: A perceptually-based interactive and immersive story environment. *PRESENCE: Teleoperators and Virtual Environments 8*, 4 (Aug. 1999), 367–391.

AARON BOBICK (afb@cc.gatech.edu) is an associate professor of Computer Science in the College of Computing and Associate Director of the GVU Center at the Georgia Institute of Technology in Atlanta, Ga.

STEPHEN INTILLE (intille@mit.edu) is a research scientist at the Massachusetts Institute of Technology in Cambridge, Mass.

JIM DAVIS (jdavis@media.mit.edu) is a Ph.D. student in the MIT Media Lab's Perceptual Computing section in Cambridge, Mass.

LEE CAMPBELL (elwin@media.mit.edu), is a Ph.D. student in the MIT Media Lab's Perceptual Computing section in Cambridge, Mass.

CLAUDIO PINHANEZ (pinhanez@us.ibm.com) is a research scientist at the IBM TJ Watson Research Center in Yorktown Heights, New York.

FREEDOM BAIRD (baird@media.mit.edu) designs and builds instruments and songs for a band in Cambridge, Mass.

YURI IVANOV (yivanov@media.mit.edu) is a Ph.D. student in the MIT Media Lab's Perceptual Computing section in Cambridge, Mass.

ARJAN SCHÜTTE (arjan@mulchmedia.com) is partner with Mulch Media and a consultant to the Internet education industry.

ANDY WILSON (drew@media.mit.edu) is a Ph.D. student in the MIT Media Lab's Perceptual Computing section in Cambridge, Mass.