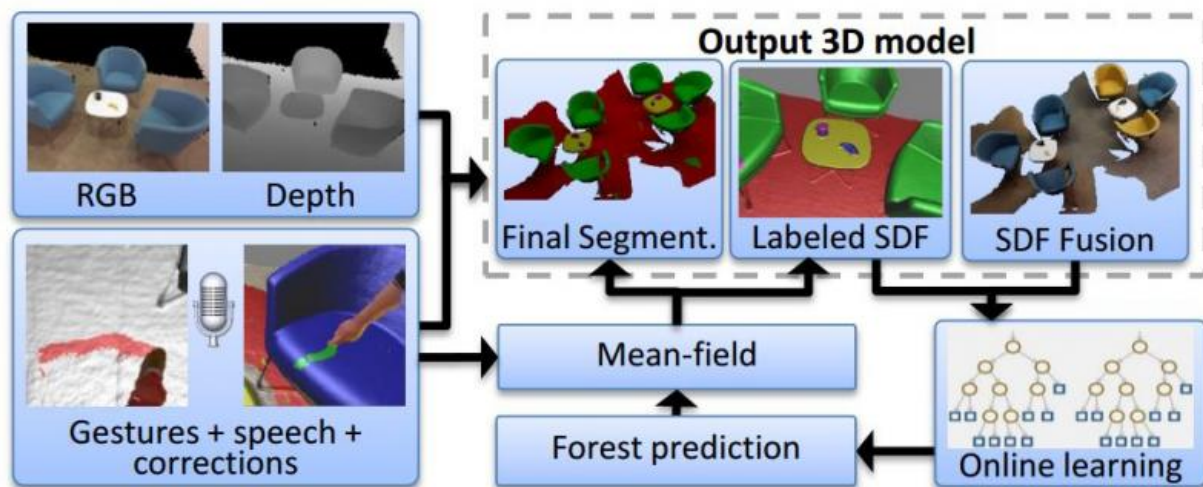# SemanticPaint system labels environment quickly online

July 3 2015, by Nancy Owano

Overview of 3D semantic modeling pipeline. Credit: Julien Valentin, et al.

Ten researchers from University of Oxford, Microsoft Research Cambridge, Stanford, and Nankai University have presented a new approach to 3D scene understanding with a system which they dubbed SemanticPaint. "Our system offers a new way of capturing, labeling and learning semantic models, all in an online manner." The key word is *online*.

In offline systems, performing capture, labeling and batch learning "often takes hours or even days," they wrote in "SemanticPaint:

Interactive 3D Labeling and Learning at your Fingertips," while their approach enables users to get "continuous live feedback of the recognition during capture." That way they can correct errors immediately in segmentation and/or learning. The system, they said, is capable of labeling new unseen parts of the environment.

The authors presented a use scenario:

"The user walks into a room with a number of chairs, tables, and smaller objects placed on the tabletops. The user is holding a consumer depth camera, such as a Kinect. Immediately they are able to capture the geometry of room, and generate a dense, globally consistent, 3D model, which can be viewed on a tablet screen or using heads-up displays."

The user can touch surfaces in proximity and perform voice commands to label the objects. Initially, the user points the camera down to the ground, extends a foot with a 'stroke' gesture across the floor. The user says 'floor' and the system automatically propagates this label across the [floor](#).

A video about their work also shows the process and demonstrates the labeling of a mug, banana (simple strokes across smaller objects will do) and chair (in the video, colored green). The video explains that this is where the 3D recognition engine kicks in. Other chairs are also labeled as in the first with the same color green. And so other objects with similar geometry and appearances get automatically labeled. The operator can easily correct recognition errors such as if, say, a chair looks "confused" with a blue smudge interrupting its green.

"In just under two minutes," said the video, "we captured a full, dense 3D model of our environment, which is broken down into separate object classes."

The Microsoft Research site said that SemanticPaint is part of an ongoing collaboration between Microsoft Research and University of Oxford; their work is to be presented at SIGGRAPH '15.

The authors commented on what research of this nature might accomplish: It might help computers not only reason about space around them, but give the geometry which they observe semantic meaning.

"Our system hopefully brings us closer to a future where people can create semantic models of their environments in lightweight ways, which can then be used in a variety of interactive applications, from robot guidance, to aiding partially sighted people, to helping us find objects and navigate our worlds, or experience new types of augmented realities."

They added that "Our system also hopefully moves us closer to the vision of life-long learning: where semantic models adapt and extend to new object classes online, as users continuously interact with the world."

Writing in *WinBeta*, Sean Cameron provides additional perspective in what their research contributes. "Typically, machines are only 'intelligent' in a very selective sense; they can execute given commands but cannot 'improvise'. Projects such as Semantic Paint represent a small but significant step in the advance towards further autonomy for machines. As the user is able to correct the machine simply, and in real-time, it is the interactivity of this project that is truly innovative."

  More information: "SemanticPaint: Interactive 3D Labeling and Learning at your Fingertips," (PDF) research.microsoft.com/en-US/p … 015semanticpaint.pdf