I. INTRODUCTION

A major limitation of 3D point cloud-based representations is that they often fail to capture an object’s functional features – a 3D model of a stapler does not specify where the stapler should be pressed. For many tools, successful manipulation requires that the robot detects the specific location that needs to be actuated (e.g., a trigger or a button) to turn on the tool. Yet, to date, virtually all methods for robotic tool use rely on 3D object models and require the programmer to specify what part of the object needs to be actuated.

Recently, interactive perception approaches have been proposed for problems ranging from discovering the degrees of freedom of an object [1] to autonomous learning of manipulation skills [2]. Inspired by these lines of work, we propose a method for a robot to interactively detect the functional parts of tools using tactile and auditory percepts coupled with exploratory behaviors. The approach consists of detecting auditory and tactile events while manipulating an object and mapping those events onto the object’s 3D model. The result is a novel visual-audio-tactile point cloud representation which not only captures the object’s shape, but also encodes its functional components (e.g., buttons) that afford actuation.

II. EXPERIMENTS AND RESULTS

The proposed method, shown in Fig. 1.a), was tested with a humanoid robot and three tools: a drill, a flashlight, and a stapler. The flashlight and the drill were explored using a grasp-squeeze sequence during which the robot detected tactile and auditory feedback (e.g., whether the drill was turned on or off). The robot explored the stapler using a touch-press sequence during which multi-modal feedback was captured as well. The starting position of each behavior execution was manually set so that it varied from trial to trial. The robot recorded 3D auditory and tactile point clouds by correlating the forward kinematics positions of its fingertips and palm with the detected auditory and tactile events. The resulting tactile and auditory 3D point clouds were mapped onto each object’s 3D model to detect its functional feature.

The results, visualized in Fig. 1, show that the method is able to successfully annotate each tool’s 3D model with the probability of successful actuation. In conclusion, the integration of non-visual sensory percepts with 3D object representations could greatly bridge the gap between human and robotic perception of objects.

REFERENCES