

# Stagehand Robot: Human-Robot Collaborative Stop-Motion Filmmaking

Sieun Choi  
Dongguk University

Junseo Kim  
Delft University of Technology

Jean Oh  
Carnegie Mellon University

**Motivation:** Stop-motion movies present a unique tactile aesthetic of handmade crafts, but production is constrained by frame-to-frame continuity, and scenes are created through repeated physical staging, micro-adjustments, and image capture. This manual loop is labor-intensive and error-prone. Small pose errors accumulate across long sequences, and continuity fixes can cascade into retakes and delays. Due to its time-consuming, labor-intensive nature, the stop-motion community has been facing declining participation. To promote the field, we present Stagehand Robot, a human-in-the-loop creative robotics pipeline in which we develop robotics technologies that serve as stagehands for physical set-making, translating natural-language shot intent into repeatable, camera-ready staging actions with closed-loop visual feedback.

**Approach:** The proposed system integrates language-to-keyframe planning, perception-guided manipulation, and continuity-aware frame capture into a single end-to-end workflow. Given a short script (e.g., “Harry walks toward the center; Sally approaches; they meet and hug”), a vision-language planner maps language into a constrained symbolic program and object-level keyframes. Each keyframe specifies a planar pose in  $SE(2)$  ( $x, y, \theta$ ), explicitly encoding both position and facing direction so cinematic intent, such as turning toward a partner or maintaining gaze, can be staged consistently. A trajectory module interpolates keyframes into per-frame pose targets aligned with director-specified timing. To realize each target, the robot executes a two-step staging loop under overhead visual feedback. First, it performs coarse placement via pick-and-place to bring an object near the desired pose. It then refines alignment with gentle planar pushes that correct residual translation and orientation errors. After each adjustment, the system estimates the object pose per frame and iterates micro-corrections until the pose falls within a tolerance, at which point it triggers image capture. Repeating this loop across targets produces a consistent sequence of staged frames that are automatically assembled into a stop-motion clip, enabling rapid iteration on blocking, spacing, and timing while reducing repetitive manual work.

**Artistic Values:** We aim to form a question regarding the ownership of artwork via two types of artifacts. On the one hand, we present the film as an artifact of the proposed system. On the other hand, we present an in-the-making video in which the robot is portrayed as a collaborative stagehand that executes precision staging while the human director retains the authorship of narrative and pacing. Through this work, we raise the question to the audience about the ownership of the creativity and the artifact. We also hope to raise awareness on the importance of real-world, tactile interaction in art creation.

**Presentation:** We propose a dual-screen video presentation in which the final film is screened on one display while a behind-the-scenes version featuring the robot and production process is shown side-by-side. Our preferred mode of presentation is video screening, and we request two monitors/TVs (preferred) positioned side-by-side, necessary cables/adapters and power, and audio playback with speakers.

## MEDIA



- *Colors* (Halloween-themed stop-motion short film): <https://youtu.be/B20y1IBAYo8?si=Gg6T3WZOvnOYbPvO>
- Behind-the-Scenes video: <https://youtu.be/8rHcxtF5ne4>