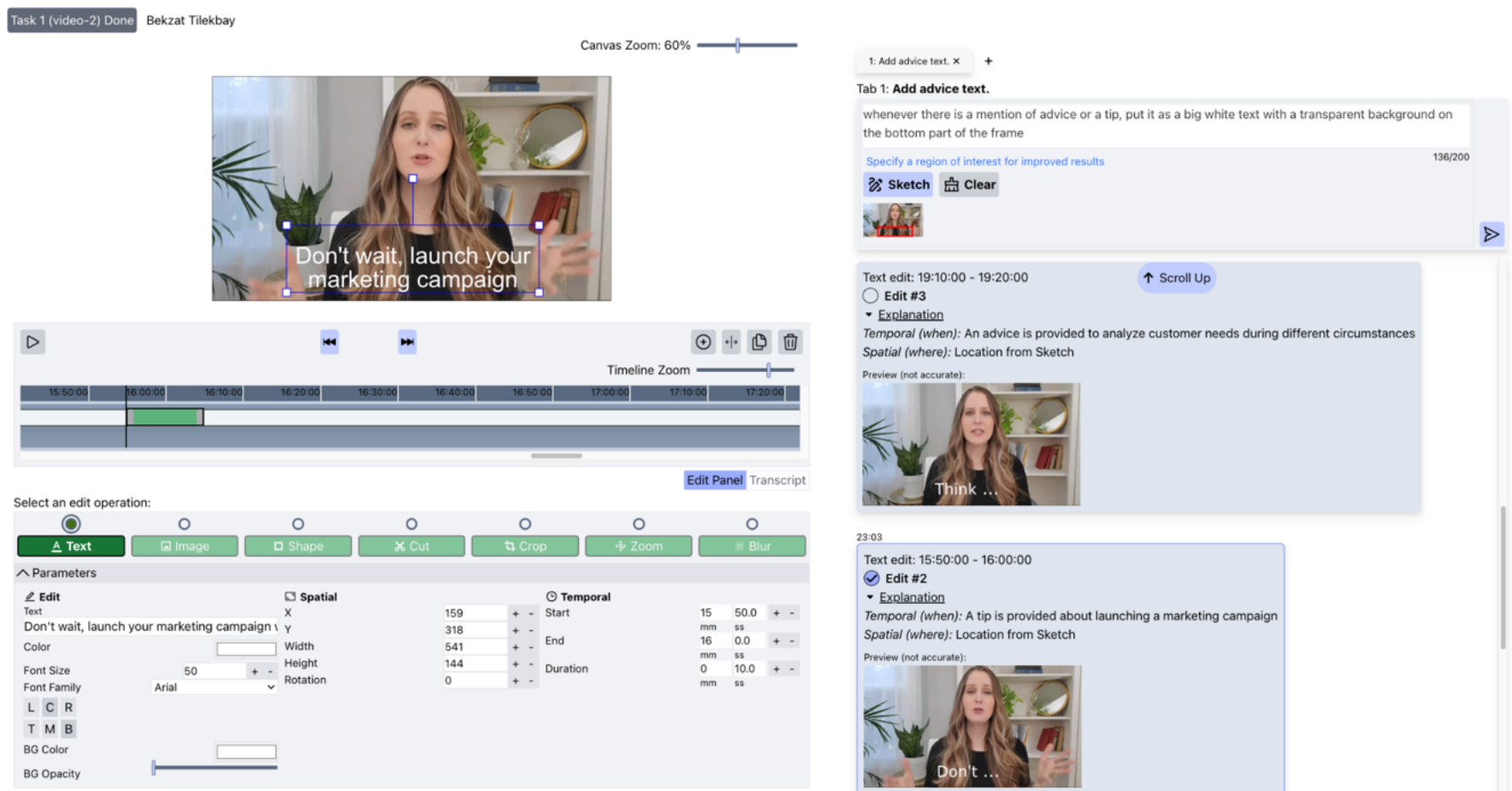


ExpressEdit: Video Editing with Natural Language and Sketching

Bekzat Tilekbay, Saelyne Yang, Michal Lewkowicz, Alex Suryapranata, Juho Kim



We present ExpressEdit, a system that enables editing videos via NL text and sketching on the video frame. Powered by LLM and vision models, the system interprets **(1) temporal, (2) spatial, and (3) edit operational references** in the NL command and **spatial references from sketching**. This work offers insights into building multimodal interfaces for video editing.

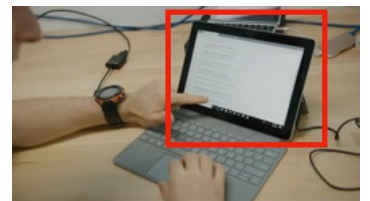


Project website:
expressedit.kixlab.org

Natural Language Text

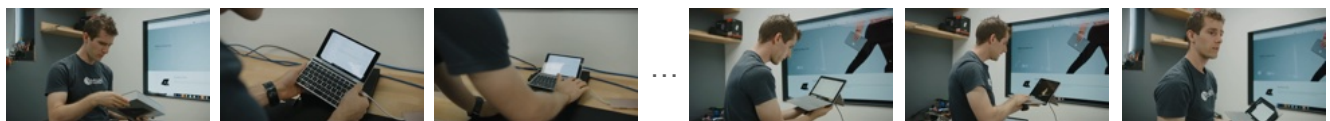
"Whenever a display appears, add a slow zoom in towards the display."

Sketch



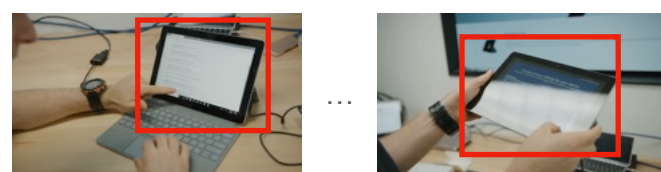
1. Temporal Interpretation

"**Whenever a display appears**, add slow zoom in towards the display."



2. Spatial Interpretation

"**Whenever a display appears**, add slow zoom in **towards the display**."



3. Edit Interpretation

"**Whenever a display appears**, add **slow zoom in towards the display**."

