Input → **Video Generation & Pose Extraction** → **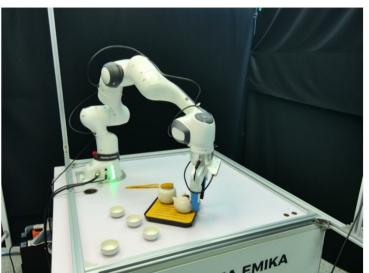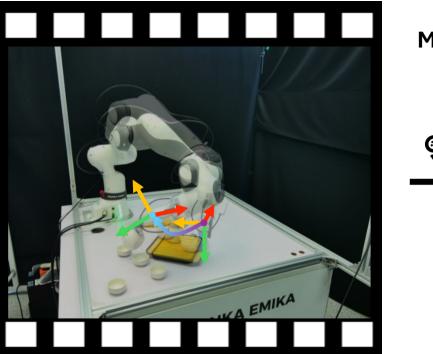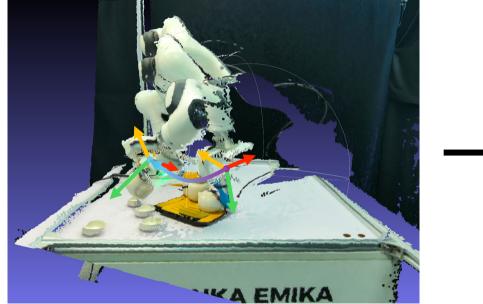Metric Depth / Lift to 3D** → **Transform to base frame** → **Execute**