# SurgSync: Time-Synchronized Multi-Modal Data Collection Framework and Dataset for Surgical Robotics


Haoying Zhou[1,2], Chang Liu[1], Yimeng Wu[1], Junlin Wu[1], Zijian Wu[3], Yu Chung Lee[3], Sara Martuscelli[4], Septimiu E. Salcudean[3], Gregory S. Fischer[2], and Peter Kazanzides[1]

[1] Johns Hopkins University [2] Worcester Polytechnic Institute
[3] University of British Columbia [4]Politecnico di Milano


ICRA IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION 2026 VIENNA

Project Website

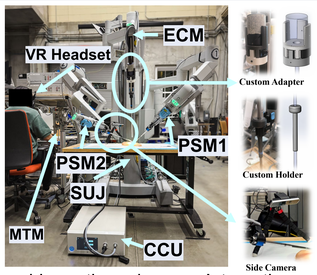
## 1. Introduction

**Problem**

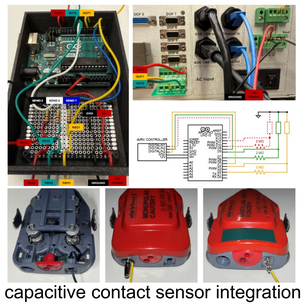
- Surgical AI needs high-quality real-world data
- Existing datasets often have
  - weak time synchronization
  - suboptimal visual fidelity
  - narrow task coverage

**Our contributions**

- Dual-mode synchronized recording
- Modern chip-on-tip stereo endoscope from Cornerstone Robotics (CSR) Ltd. integration
- Capacitive contact sensor integration
- Scalable post-collection processing toolbox
- Multi-task user-study dataset from surgeons, experienced researchers and novice students




chip-on-tip endoscope Integration with the dVRK-Si system



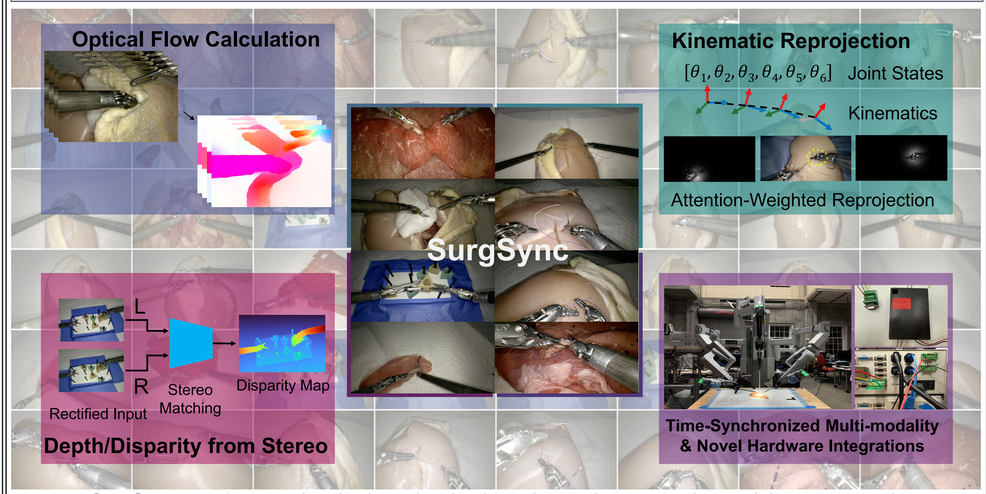
capacitive contact sensor integration

## 2. Overview

**SurgSync** synchronizes
- Video streams
- Robot kinematics
- Timestamps
- Contact signals

Support both:
- **Online matching**: real-time workflow
- **Offline matching**: training dataset generation

Provide downstream tools for
- Depth/disparity estimation
- Optical flow
- Kinematic reprojection
- Annotation




**SurgSync** overview: synchronized sensing, hardware integration, processing, and dataset generation.

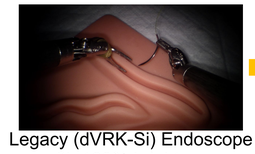
## 3. Data Collection Framework

**Two recording modes**

- **Online-matching**
  - Strict synchronization during capture
  - Discard pairs outside the time tolerance
  - Ready to use
- **Offline-matching**
  - Capture first, synchronize later
  - Higher throughput for dataset building

**Imaging Result**

- Image sharpness improved by **>30x**
- Laplacian variance
  - CSR: 529.48 ± 23.77
  - dVRK-Si: 16.93 ± 2.47

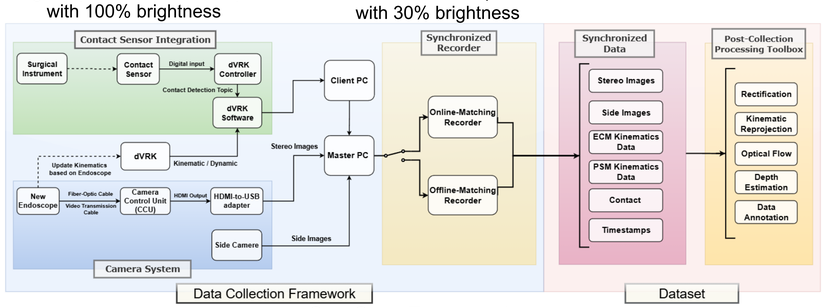


Legacy (dVRK-Si) Endoscope with 100% brightness

CSR Endoscope with 30% brightness

Laplacian Variance on Suturing Scenes

| Endo scope | Mean | STD |
|---|---|---|
| CSR | 529.45 | 23.77 |
| dVRK-Si | 16.93 | 2.47 |




SurgSync Framework

## 4. Post-collection Toolbox

**Features**

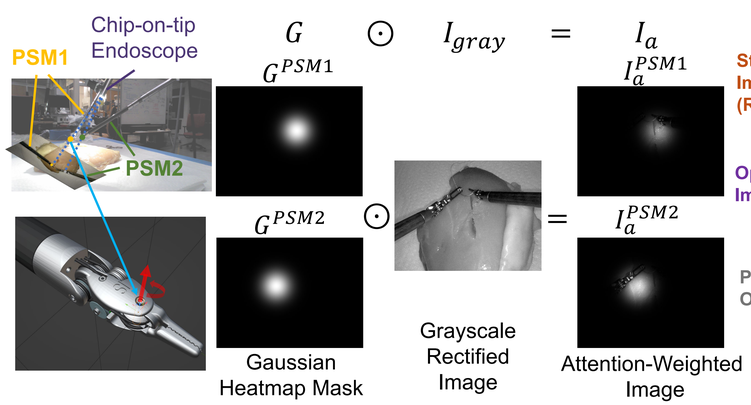
- Stereo rectification
- Depth/disparity estimation (FoundationStereo) *[available on website]*
- Optical flow (RAFT) *[available on website]*
- Kinematic reprojection
- GUI-based annotation

**Kinematic reprojection**

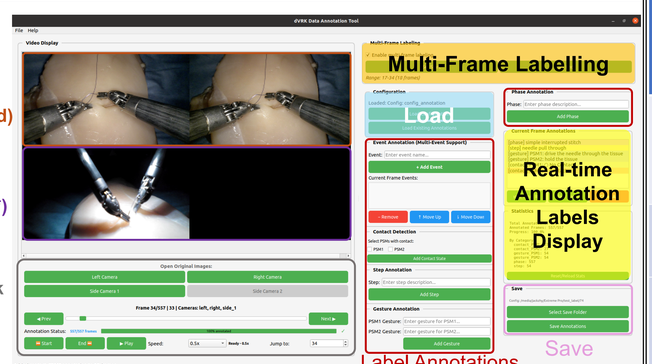
- Project robot kinematics (3D) into image space (2D)
- Use Gaussian heatmap mask
- Generate attention-guided visual features

**Annotation**

- Video playback
- Multi-frame annotation
- Contact Labels
- Hierarchical task workflow
  - ❑ Phase
  - ❑ Step
  - ❑ Gesture (arm-level)
- Special Events




Kinematic reprojection with Gaussian heatmap masks
**robot hand-eye calibration included**




(*) can show side-view image / disparity map ….

- Phase, Step, Event : hierarchical workflow descriptions
- Gesture: Arm-level task descriptions
- Contact Detection (Boolean)

GUI-based annotator

## 5. Dataset and Synchronization Results

User-study dataset

- 214 validated instances
- 13 human subjects
- Incorporated to NVIDIA open-h-embodiment (2500+ trajectories)

Tasks include

- Peg transfer
- Suturing
- Tissue manipulation
- Cold-cut dissection

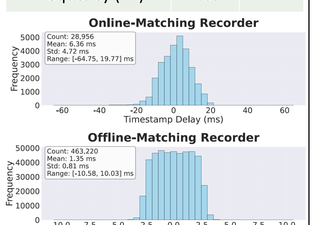
**Dataset Distribution**

P: professionals (surgeons or residents)
E: experienced (researchers)
N: novice (students)

| Training Task | User Group | Number of Instances | | Total |
|---|---|---|---|---|
| | | Online | Offline | |
| Suturing and Knot Tying | N | 13 | 2 | 104 |
| | E | 36 | 12 | |
| | P | 2 | 39 | |
| Peg Transfer | N | 7 | - | 18 |
| | E | 11 | | |
| Tissue Manipulation | N | 9 | - | 21 |
| | E | 12 | | |
| Cold-Cut Dissection | N | 6 | - | 71 |
| | E | 15 | 9 | |
| | P | 1 | 40 | |

**Time Synchronization Performance**

| Attribute | Online | Offline |
|---|---|---|
| Cross-modal latency mean (ms) | 6.36 | **1.35** |
| Cross-modal latency STD (ms) | 4.72 | **0.81** |
| Cross-modal latency median (ms) | 5.58 | 1.33 |
| Recording Frequency (Hz) | 4.04 ± 1.69 | **10** |




## 6. Conclusion and Future work

- Develop scalable data infrastructure including a C++ multi-threaded data collection framework and a Python-based data processing toolbox using AI-driven models.
- Transform heterogenous surgical robot kinematics, vision and sensor streams into learning-ready dataset for downstream
  - Perception
  - State Estimation (e.g. contact detection)
  - Skill assessment
  - Scene understanding
  - Policy-development workflows

Future work includes additional data collection, analysis, and further recorder upgrade (e.g. using GPU acceleration).


## Acknowledgement

- This work was supported in part by NSF AccelNet award OISE-1927275 and OISE-1927354.
- Thanks to Intuitive Surgical Inc. for their hardware support of dVRK-Si.
- Thanks to Cornerstone Robotics Ltd. for the chip-on-tip endoscope.