2023 International Symposium on Medical Robotics (ISMR)
Atlanta, GA, USA, April 19-21, 2023

Smoothness Constrained Curiosity Driven Multicamera Trajectory Optimization for Robot-Assisted Minimally Invasive Surgery

Divas Subedi, Wenfan Jiang, Ramisa Tahsin Rahman, Heidi Zhang, Kevin Huang, and Yun-Hsuan Su

***Abstract*— This paper presents a novel, curiosity driven camera positioning algorithm for multicamera systems in robot-assisted minimally invasive laparoscopic procedures. The work here extends the authors' prior studies in curiosity driven movement algorithms by introducing a new jerk-minimization term in the extrinsic curiosity reward function. Three basic and three curiosity driven movement baseline algorithms are comparatively evaluated against the novel motion-smoothing approach in both visual and motion metrics - the latter is analyzed with both time and frequency domain scores. All tests were performed on an identical laparoscopic simulation, with dynamic tissue surface, simulated breathing motion, and virtual tool-tissue interactions incorporated. Multicamera systems can be inserted through a single trocar and be magnetically anchored or maneuvered. Such systems can enable enhanced visual feedback and sensing of the surgical cavity, with multiple simultaneous views affording 3D reconstruction. Results of the study presented here are promising, and show that the modified curiosity driven algorithm does indeed reduce camera jerk at a meager cost to visual and reconstructability metrics. It is hypothesized that reduction in overall camera motion jerk, while still maintaining 3D reconstructability quality, is a desirable characteristic in teleoperated laparoscopic procedures.**

***Index Terms*— robot-assisted surgery; multirobot systems; reinforcement learning; 3D reconstruction; path smoothness**

I. Introduction

Robot-assisted minimally invasive surgery (RMIS) presents salient benefits compared to manual laparoscopic procedures, including elimination of inverted motion commands as well as introduction of intelligent software layer to scale motion and reduce tremor at the tool tip. Promising studies have demonstrated that 3D visualization of the surgical scene can reduce operation time [1], yet intraoperative 3D scene reconstruction is still an open challenge. Typically, a surgeon is presented with visual feedback from a single imaging sensor, potentially with narrow baseline stereo-vision. Some approaches seek to expand the capabilities of the inserted vision sensor, for example with structured light methods [2], structure from motion or simultaneous localization and mapping as examples [3]. These methods are constrained by a limited number of inserted sensors. Multicamera systems seek to overcome this restriction, implementing an array of vision sensors that are inserted through a single trocar and subsequently magnetically anchored and maneuvered [4]–[8]. These systems enable multiple, independently controlled viewpoints of the surgical scene. Novel multicamera navigation and reconstruction methods need to be developed to accommodate for the dynamic, deforming and spatially constrained task space in RMIS.

A. Curiosity Learning and Motion Smoothing

To prescribe individual camera motion, the authors' propose a curiosity driven method. Curiosity driven learning is a reinforcement learning technique that drives agent's to explore less predictable states [9], encouraging exploration of areas that are novel or where prediction error is relatively high. In the authors' prior studies, multiple camera trajectories were determined via a curiosity learning method that considered in the extrinsic reward calculation field of view (FOV) coverage of the surgical scene and computational efficiency [10], where 3D reconstructability is determined by points that are viewable by two or more cameras [11]–[13].

In addition to the visual and computational considerations mentioned above, it is of interest to incorporate motion constraints to ensure stable visual feedback quality. Minimizing jerk is desired since excessive jerk leads to blurry and jarring visual feedback, thus interfering with situational awareness. Jerk minimization or cancellation is an active method for video stabilization and motion smoothing. Recent applications include smoothing drone flight plans [14], cinematic camera trajectory paths [15], surgeon motion feedback [16], and smooth robotic manipulation trajectories [17], [18].

B. Contributions

This project seeks to enhance an intelligent, curiosity driven multicamera positioning framework for RMIS. To the best of the authors' knowledge, it is the first to:

i) introduce an extrinsic reward function in a curiosity driven multicamera view adjustment algorithm that simultaneously optimizes for reconstructability coverage and trajectory smoothness;
ii) evaluate the above method, against six other multicamera movement algorithms using both visual and motion scores (time and frequency domain motion metrics);
iii) demonstrate the feasibility of multicamera motion smoothing without sacrificing 3D reconstructability of the surgical scene.

This material is based upon work supported by the National Science Foundation under Grant No. IIS-2101107. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF.

Divas Subedi and Kevin Huang are with Trinity College, Dept. of Engineering, 300 Summit St, Hartford, CT 06106 USA {divas.subedi, kevin.huang}@trincoll.edu

Wenfan Jiang is with University of Michigan, Department of Computer Science, 500 S State St, Ann Arbor, MI 48109 jiangwf@umich.edu.

Ramisa Tahsin Rahman, Heidi Zhang and Yun-Hsuan Su are with Mount Holyoke College, Department of Computer Science, 50 College St, South Hadley, MA 01075 USA {rahma24r, zhang47j, msu}@mtholyoke.edu.

2023 International Symposium on Medical Robotics (ISMR) | 979-8-3503-0162-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISMR57123.2023.10130237

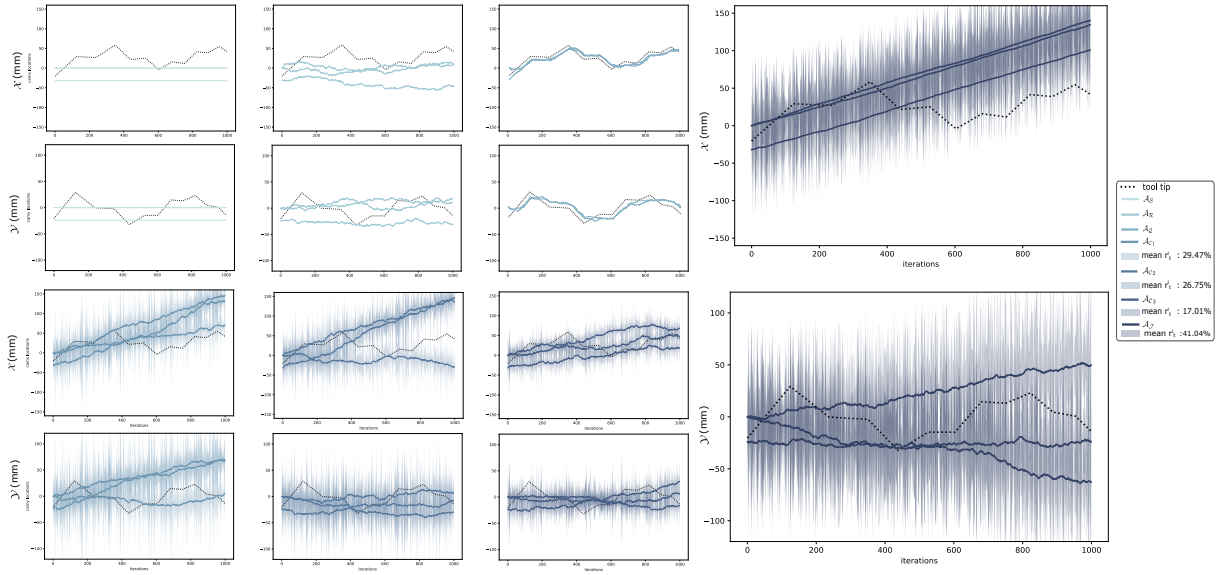

Fig. 1: The camera trajectories in the X and Y direction using the seven algorithms. The tinted regions the bottom two rows and the right column represent the instantaneous intrinsic reward value $r_t^i = ||\phi(\hat{s_{t+1}}) - \phi(s_{t+1})||$, depicted is in percentage (%), given the chosen action a_t.

II. Methods

A. Problem Description

Originally proposed in prior work [10], a curiosity driven reinforcement learning based multicamera viewpoint adjustment framework for abdominal minimally invasive surgery was designed to simultaneously (a) maximize weighted 3D reconstructable coverage of the dynamic surgical scene; and (b) improve runtime efficiency via the dimension reduced 2D state representations. In this study, the authors added a third objective to: (c) encourage smooth camera trajectories. The surgical scene sequences, as in Fig 2, were generated via simulation for easy replication. Specifically, three magnetically anchored virtual cameras (30° FOV, 30 fps) were inserted through a trocar, mounted on the inner abdominal wall, and maneuvered by external magnets. Each camera moves independently according to the chosen action a_t generated by policy π, which is trained using the A3C module [19] from extracted features of the flattened 2D state, s_t, the next state, s_{t+1}, and the two-component reward function, $r_t = r_t^i + r_t^e$. The intrinsic curiosity reward, r_t^i, promotes 'curious exploration' by encoding the discrepancy between the predicted $\hat{s}_{t+1}$ and actual s_{t+1} next state; whereas the custom designed extrinsic environment reward, r_t^e, further optimizes the chosen action toward external objective functions, in this case, maximum 3D reconstructability coverage and minimum camera trajectory jerk.

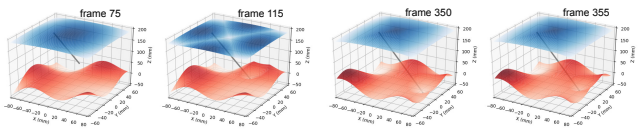

Fig. 2: Sample frames of the surgical scene sequences (total 33.3s captured at 30 Hz). The surgical tool is denoted by the gray line segment. The depicted surfaces represent the deforming abdominal wall (blue) and tissue patch (pink) due to breathing, heartbeat or surgical palpation. Details about the simulation dataset can be found in [10].

B. Reward with Smoothness Constraints

The reward function, $r_t = r_t^i + r_t^e$, is composed of intrinsic reward, r_t^i, and extrinsic reward, r_t^e. The intrinsic reward r_t^i is trained through the ICM framework [20], while the extrinsic reward, r_t^e, is a novel function designed to: (a) reflect the cameras' FOV coverage as a percentage of the entire tissue area, (b) encourage computational efficiency, and (c) encourage smooth camera trajectories. To that end, r_t^e is mathematically represented as

$$\begin{aligned} r_t^e &= k_1 \frac{\mathcal{F}_{\mathcal{S}}\left(\min({}^{R}I, 2k_r)\right)}{2\kappa_p\,\kappa_r} \\ &+ k_2 \frac{\mathcal{F}_{\mathcal{S}}(\alpha)}{255\,\beta} + k_3 \frac{\mathcal{F}_{\mathcal{S}}({}^{B}I \odot \alpha)}{255^2\,\beta} + \gamma \end{aligned} \tag{1}$$

$$\text{, where} \begin{cases} \text{Common Numerator: } \alpha = {}^{G}I \odot \mathcal{F}_{\mathcal{T}}({}^{R}I, 2\kappa_r) \\ \text{Common Denominator: } \beta = \mathcal{F}_{\mathcal{S}}\left(\mathcal{F}_{\mathcal{T}}({}^{R}I, \kappa_r)\right) \\ \text{Motion Penalty: } \gamma = \dfrac{\sum_{i=1}^{3}\sqrt{x_i'''(t) + y_i'''(t)}}{3\,\kappa_m} \\ \text{Balancing Coeff: } \displaystyle\sum_{i=1}^{3} k_i = 1, \forall k_i \in (0.0, 1.0) \end{cases}$$

Here, ${}^{R,G,B}I \in \mathbb{R}^{120\times160}$ store pixel values of each color channel in the flattened state, $s_t \in \mathbb{R}^{120\times160\times3}$; $x_i'''(t), y_i'''(t)$ are axial jerk of the i^{th} camera smoothed by a moving average sliding window of four time frames; $\kappa_p = 120 \times 160 = 19200$ is size of the tissue point cloud; $\kappa_r = 50$ and $\kappa_m = 0.2$ are heuristically tuned hyperparameters that determine the intensity step size; and motion penalty severity respectively. $\mathcal{F}_{\mathcal{S}}: \mathbb{R}^{120\times160} \rightarrow \mathbb{R}$ calculates the grand sum of a matrix; $\mathcal{F}_{\mathcal{T}}: \left(\mathbb{R}^{120\times160} \times \mathbb{R}\right) \rightarrow \mathbb{B}^{120\times160}$ is an element-wise binary threshold operator; and $\odot$ denotes element-wise multiplication. The motion penalty γ discourages motion jerk, and is a novel term introduced to the r_t^e derivation from prior work [10].

III. Experimental Procedure

This study introduces a novel, smoothness-constrained curiosity driven multicamera viewpoint adjustment algorithm, $\mathcal{A}_{\mathcal{J}}$. The method is comparatively evaluated against six baseline approaches, with each algorithm tested in simulation, for repeatability, on the same surgical sequences (Fig. 2), each containing 1000 time steps.

A. Baseline Approaches

The six baseline approaches - $\mathcal{A}_{\mathcal{S}}$, $\mathcal{A}_{\mathcal{R}}$, $\mathcal{A}_{\mathcal{Q}}$, $\mathcal{A}_{\mathcal{C}1}$, $\mathcal{A}_{\mathcal{C}2}$, and $\mathcal{A}_{\mathcal{C}3}$ - are described in the authors' prior work [10], [21]. These methods were implemented here to evaluate against the proposed algorithm, $\mathcal{A}_{\mathcal{J}}$, and are briefly described in the following paragraphs (the reader is directed to [10], [21] for additional details).

1) **Static Cameras ($\mathcal{A}_{\mathcal{S}}$):** this algorithm considers the strategy of all cameras remaining fixed at their starting configurations with respect to the abdominal wall. The camera orientation remains normal to the inner abdomen surface, even with breathing motion. This baseline algorithm assesses the need for dynamically moving cameras in the case where multiple cameras are present.

2) **Random Camera Motion ($\mathcal{A}_{\mathcal{R}}$):** in this algorithm, cameras move randomly along the abdominal wall surface. The camera orientation is again constrained to be normal to the abdomen surface. This baseline algorithm is used analyze the effectiveness of the curiosity driven motion planning [9], [20] in the context of multicamera viewpoint adjustment.

3) **Three dimensional states ($\mathcal{A}_{\mathcal{Q}}$):** initially proposed in [21], the 3D locations of key points in the surgical scene are concatenated into a single vector. The vector is then directly treated as state S_t. This baseline is executed to analyze the state dimension reduction proposed in [10].

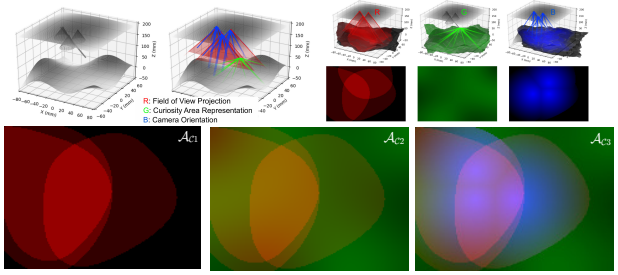

Fig. 3: (top) The 3D surgical scene is flattened to 2D state representation though feature extraction. The RGB color channels respectively encode the field of view extraction (R), curiosity area representation (G) and camera orientation (B). (bottom) Sample state representations using baseline algorithms $\mathcal{A}_{\mathcal{C}1}$, $\mathcal{A}_{\mathcal{C}2}$ and $\mathcal{A}_{\mathcal{C}3}$. Mathematical details are provided in [10].

4) **Curiosity Driven Variations ($\mathcal{A}_{\mathcal{C}1}$, $\mathcal{A}_{\mathcal{C}2}$, $\mathcal{A}_{\mathcal{C}3}$):** three curiosity based algorithms share the same reinforcement learning module, with the only difference being the state representation, S_t. S_t in algorithm $\mathcal{A}_{\mathcal{C}1}$ encodes information only in the R color channel. $\mathcal{A}_{\mathcal{C}2}$ and $\mathcal{A}_{\mathcal{C}3}$ respectively consider two (R+G) and three (R+G+B) color channels. Figure 3 shows the three variations of state representation given a sample 3D simulated surgical scene. Each color channel from red to blue is designed to encapsulate critical information that is sorted by importance and affects different aspects of the calculated reward. Mathematical details are provided in [10].

B. Evaluation Metrics

1) Visual Scores: The seven algorithms were comparatively evaluated using four visual metrics [10], defined as:

- $\mathbf{M_V}$: **Visibility** denotes the average percentage of surgical scene points visible by at least one camera.
- $\mathbf{M_R}$: **Reconstructability** is the average percentage of surgical scene points visible by at least two cameras.
- $\mathbf{M_{T-R}}$: **Tool-weighted Reconstructability** records the mean weighted reconstructability [%] whose weights are determined by ${}^{B}I$.
- $\mathbf{M_{A-R}}$: **Accumulative Reconstructability** is the percentage of reconstructable points out of all points already recognized by the model (i.e. observed by at least two camera within the most recent five time steps).

2) Motion Scores: Two motion metrics were used to determine the camera trajectory smoothness, one in the time domain ($\mathbf{M_J}$) and one in the frequency domain ($\mathbf{M_S}$).

- $\mathbf{M_J}$**: Absolute Jerk** records the motion jerk magnitude, disregarding the motion axis. This time domain measure of jerk [22], commonly understood as the derivative of acceleration, is represented in mathematical terms as

$$\mathbf{M_J} = \left(\int_{t_2}^{t_1} [x'''(t)]^2 dt \right) \frac{D^3}{{v_M}^2} \tag{2}$$

, where t is time, $x(t)$ is the movement function, $D = t_2 - t_1$ is the movement duration, and v_M is the maximum speed along the entire movement procedure. The $\mathbf{M_J}$ score is expressed in $^m/_{s^3}$ (the SI units of jerk) throughout the article.

- $\mathbf{M_S}$**: Spectral Arc Length**, on the other hand, captures the motion smoothness in the frequency domain. This is a modified version of the continuous spectral arc length measure [23] specifically accounting for discrete movements [24], mathematically represented as:

$$\begin{aligned} \mathbf{M_S} &= \int_0^{\omega_c} \left[\left(\frac{1}{\omega_c} \right)^2 + \left(\frac{d\hat{V}(\omega)}{d\omega} \right)^2 \right]^{\frac{1}{2}} d\omega \qquad (3) \\ &= \int_0^{\omega_c} \left[\left(\frac{1}{\omega_c} \right)^2 + \frac{1}{[V(0)]^2} \left(\frac{dV(\omega)}{d\omega} \right)^2 \right]^{\frac{1}{2}} d\omega \end{aligned}$$

, where $V(\omega)$ is the Fourier magnitude spectrum of the trajectory, $V(0)$ is DC magnitude, $\hat{V}(\omega) = \frac{V(\omega)}{V(0)}$ is the normalized magnitude spectrum, and ω_c is a coefficient term set as:

$$\omega_c \triangleq \min \left(\max\, \omega_c, \min \left(\omega, \hat{V}(r) < \bar{V}\ ,\ \forall r > \omega \right) \right) \tag{4}$$

In implementation of the time domain approach, the camera motion is updated at 33 Hz, which translates to one camera position recording every 0.0303 seconds. Therefore, the movement duration, D, when calculating $\mathbf{M_J}$ is the $4 \times 0.0303 = 0.1212$ seconds, the smallest interval to approximate $x'''(t)$ in discrete sampling.

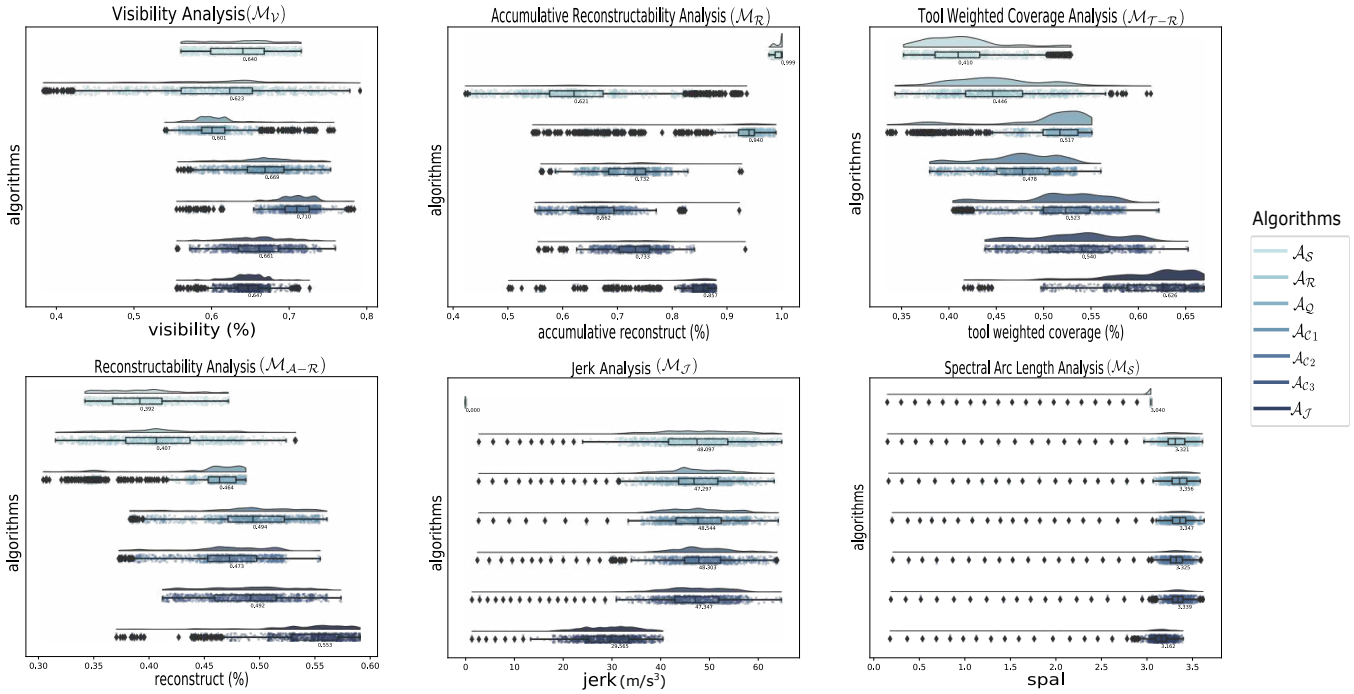

Fig. 4: Histograms of the four visual metrics ($\mathbf{M_V}$, $\mathbf{M_R}$, $\mathbf{M_{T-R}}$, $\mathbf{M_{A-R}}$) and the two motion metrics ($\mathbf{M_J}$, $\mathbf{M_S}$) using $\mathcal{A}_{\mathcal{J}}$ and the six baseline algorithms at the testing phase. Mean scores are annotated on the graphs.

For the spectral arc length, the above mentioned ω_c provides a better sensitivity and reliability through constantly updating ω. This method is designed for movements that are only a few seconds long. Therefore, ω_c was heuristically tuned to 40π when calculating $\mathbf{M_S}$.

IV. Results and Discussion

Table I showcases the mean metric scores across all time frames for each algorithm and performance difference (the gray rows) against the algorithm of interest, $\mathcal{A}_{\mathcal{J}}$. The proposed smoothness constrained algorithm, $\mathcal{A}_{\mathcal{J}}$, shows superior metric scores in $\mathbf{M_R}$, $\mathbf{M_{T-R}}$, $\mathbf{M_J}$, and $\mathbf{M_S}$ compared to all six baselines.

Beyond numeric results, different facets of the results were also visualized. Figures 1, 4, 6, and 7 respectively demonstrate the camera trajectories, metric score histograms, chosen actions, and extrinsic rewards of the testing surgical sequence.

Finally, the authors examined path smoothness of the multicamera system under three assumptions regarding operator experience: (a) surgeons prefer to observe the surgical scene from a camera viewpoint that has minimal to no movement; (b) frequent changes to the viewpoint (camera) that corresponds to the least jerk can be distracting to surgeons; and (c) accurate (tool-weighted) 3D reconstruction of the dynamic surgical scene is beneficial for making critical decisions during surgery. To compare the algorithms against the three hypotheses, the motion smoothness of each individual camera across time frames is illustrated in Fig. 5. Figure 8 then shows the instantaneous correlation between reconstructability ($\mathbf{M_R}$) and motion smoothness ($\mathbf{M_J}$, $\mathbf{M_S}$).

A. Observable Effects of Smoothness Constraints

The positive influences and side effects of the smoothness constraints in $\mathcal{A}_{\mathcal{J}}$ are perceivable in multiple figures.

Algorithm \ Metric	$\mathbf{M_V}$ ↑ preferred	$\mathbf{M_R}$ ↑ preferred	$\mathbf{M_{T-R}}$ ↑ preferred	$\mathbf{M_{A-R}}$ ↑ preferred	mean $\mathbf{M_J}$ (Camera1, Camera2, Camera3, **All Cameras**) ↓ preferred	mean $\mathbf{M_S}$ (Camera1, Camera2, Camera3, **All Cameras**) ↓ preferred
$\mathcal{A}_{\mathcal{S}}$	0.640 1.08% ↓	0.392 29.11% ↓	0.410 34.50% ↓	0.999 16.56% ↑	(48.537, 49.745, 47.684, **48.655**) 61.73% ↑, 72.69% ↑, 59.60% ↑, **64.58**% ↑	(3.357, 3.309, 3.363, **3.343**) 6.44% ↑, 4.88% ↑, 5.95% ↑, **5.76**% ↑
$\mathcal{A}_{\mathcal{R}}$	0.623 3.71% ↓	0.407 26.40% ↓	0.446 28.75% ↓	0.621 27.54% ↓	(46.627, 47.566, 50.097, **48.097**) 55.37% ↑, 65.13% ↑, 67.68% ↑, **62.69**% ↑	(3.316, 3.357, 3.287, **3.320**) 5.14% ↑, 6.37% ↑, 3.56% ↑, **5.03**% ↑
$\mathcal{A}_{\mathcal{Q}}$	0.601 7.11% ↓	0.464 16.09% ↓	0.517 17.41% ↓	0.940 9.68% ↑	(47.948, 44.861, 49.080, **47.296**) 59.77% ↑, 55.74% ↑, 64.27% ↑, **59.98**% ↑	(3.345, 3.345, 3.366, **3.355**) 6.06% ↑, 5.99% ↑, 6.05% ↑, **6.14**% ↑
$\mathcal{A}_{\mathcal{C}1}$	0.669 3.47% ↑	0.494 10.67% ↓	0.478 23.64% ↓	0.732 14.59% ↓	(48.465, 47.413, 49.752, **48.543**) 61.49% ↑, 64.60% ↑, 66.52% ↑, **64.10**% ↑	(3.349, 3.359, 3.332, **3.347**) 6.18% ↑, 6.43% ↑, 4.98% ↑, **5.88**% ↑
$\mathcal{A}_{\mathcal{C}2}$	0.710 9.74% ↑	0.473 14.47% ↓	0.523 16.45% ↓	0.662 22.75% ↓	(50.169, 47.573, 47.165, **48.302**) 67.17% ↑, 65.16% ↑, 57.86% ↑, **63.38**% ↑	(3.319, 3.298, 3.356, **3.324**) 5.23% ↑, 4.50% ↑, 5.73% ↑, **5.16**% ↑
$\mathcal{A}_{\mathcal{C}3}$	0.661 2.16% ↑	0.492 11.03% ↓	0.540 13.74% ↓	0.734 14.35% ↓	(47.879, 47.794, 46.367, **47.347**) 59.54% ↑, 65.92% ↑, 55.19% ↑, **60.15**% ↑	(3.340, 3.343, 3.333, **3.338**) 5.90% ↑, 5.93% ↑, 5.01% ↑, **5.60**% ↑
$\mathcal{A}_{\mathcal{J}}$	0.647 0%	0.553 0%	0.626 0%	0.857 0%	(30.011, 28.805, 29.877, **29.564**) (0%, 0%, 0%, 0%)	(3.154, 3.156, 3.174, **3.161**) (0%, 0%, 0%, 0%)

TABLE I: Mean metric scores and percentage difference (red and green text) of six baseline algorithms compared against the proposed $\mathcal{A}_{\mathcal{J}}$. The best performing algorithm under each evaluation metric is highlighted in blue.

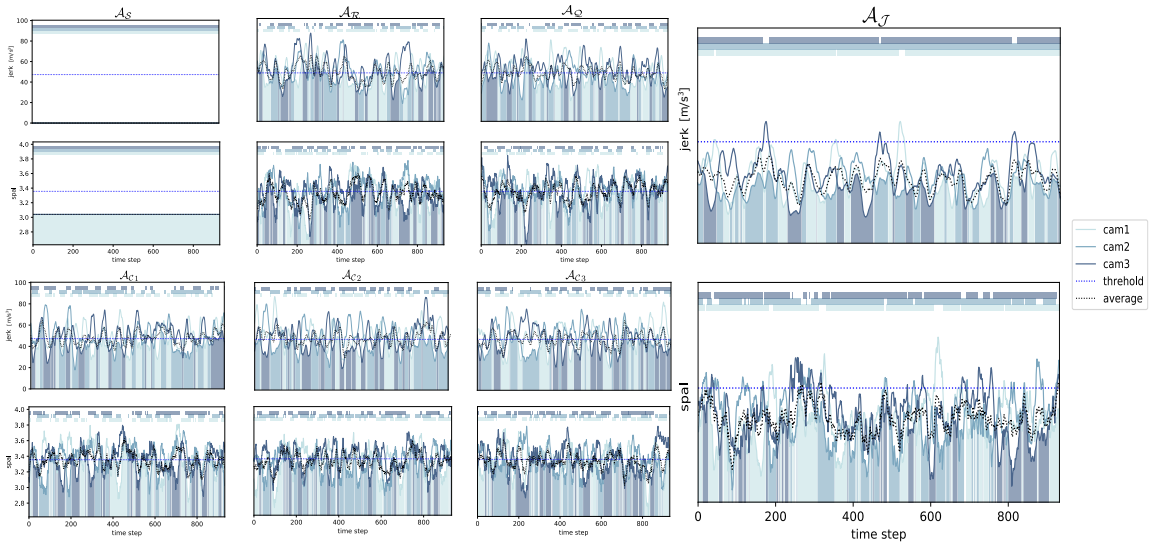

Fig. 5: The instantaneous smoothness metric scores over the 1000 time frames during the test surgical sequence are analyzed. Results from the baseline algorithms are displayed on the left 3 columns, while the right most column for $\mathcal{A}_\mathcal{J}$. In every subgraph, $\mathbf{M_J}$ (top half) and $\mathbf{M_S}$ (bottom half) scores for each camera are illustrated as the solid curves in three different shades of blue. The black dotted curve denotes the instantaneous average scores across the three cameras, whereas the blue dotted line represents a constant comfort threshold that is heuristically determined by the mean metric score using $\mathcal{A}_\mathcal{Q}$. The top of each subgraph is color-coded to indicate the time periods where the smoothness score of a certain camera falls below the threshold. Finally, the bottom of each subgraph is also filled with a shade of blue to reflect the camera whose instantaneous smoothness score is the lowest.

- From the camera trajectory log in Fig. 1, $\mathcal{A}_\mathcal{J}$ shows little motion direction changes along the x axis, an axis where the tool trajectory has a more consistently ascending trend, for all three cameras. Although the average intrinsic reward, r_t^i, (color filled area centering the trajectory curves) corresponding to the chosen action, a_t, reduces as more color channels are incorporated in the state representation, $\mathcal{A}_\mathcal{J}$ reflects a higher r_t^i compared to any algorithm in the $\mathcal{A}_\mathcal{C}$ series. This is because the added smoothness constraints can explicitly alter the chosen action and consequently the actual next state, which subsequently increases r_t^i, a quantity calculated as the error between the predicted next states, $\hat{s_{t+1}}$, and the actual next state, s_{t+1}.

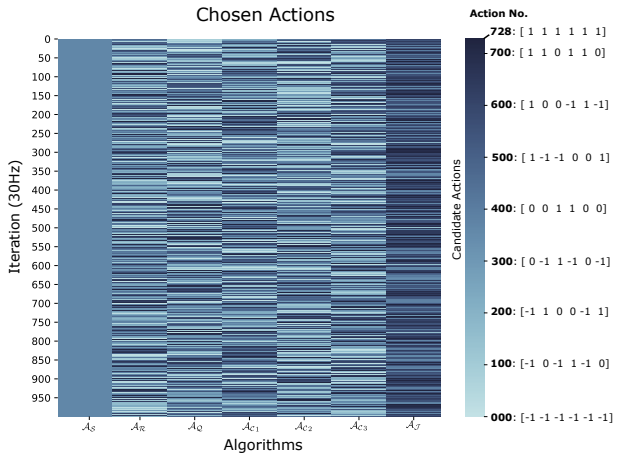

Fig. 6: Chosen actions a_t across 1000 testing time frames. The selected multicamera movement action $a_t = [\Delta x_k, \Delta y_k, \forall k \in [1,3]] \in \mathbb{R}^6$ at time t encodes the motion of all 3 independently moving cameras. It is a length 6 vector where each element contains an axial movement choice (either along x or y) by -1, 0 or 1 mm for a particular camera.

- Figure 6 demonstrates a chosen action sequence for $\mathcal{A}_\mathcal{J}$ that shares a more similar shade of blue compared to other baseline algorithms (except the static camera baseline, $\mathcal{A}_\mathcal{S}$). This means that $\mathcal{A}_\mathcal{J}$ is not only selecting actions based on scene reconstructability, tool location, and exploration of unseen region, but also takes into account differences with previous actions, eventually picking ones that more resemble past ones.
- In Fig. 7, the extrinsic reward (r_t^e) curve for $\mathcal{A}_\mathcal{J}$ shares the same oscillating frequencies as the baselines, which are believed to be a result of periodic breathing motions of the abdominal wall. However, it also exhibits several sudden drops (from the smoothness constraints) on an overall superior r_t^e curve. This means that other than the occasional penalties due to excess jerk, the external optimization objectives are well met.

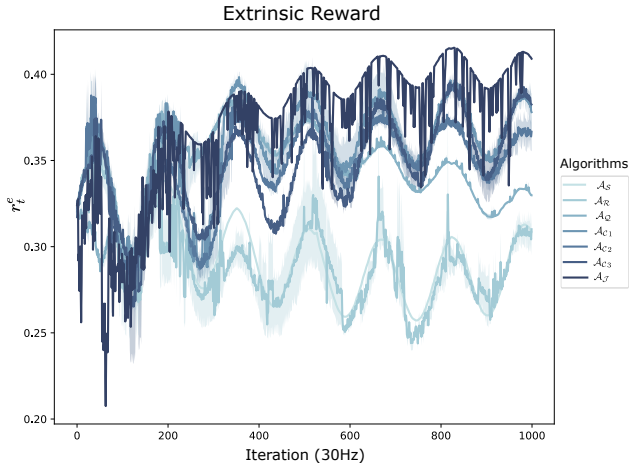

Fig. 7: The instantaneous extrinsic reward r_t^e in the test sequence.

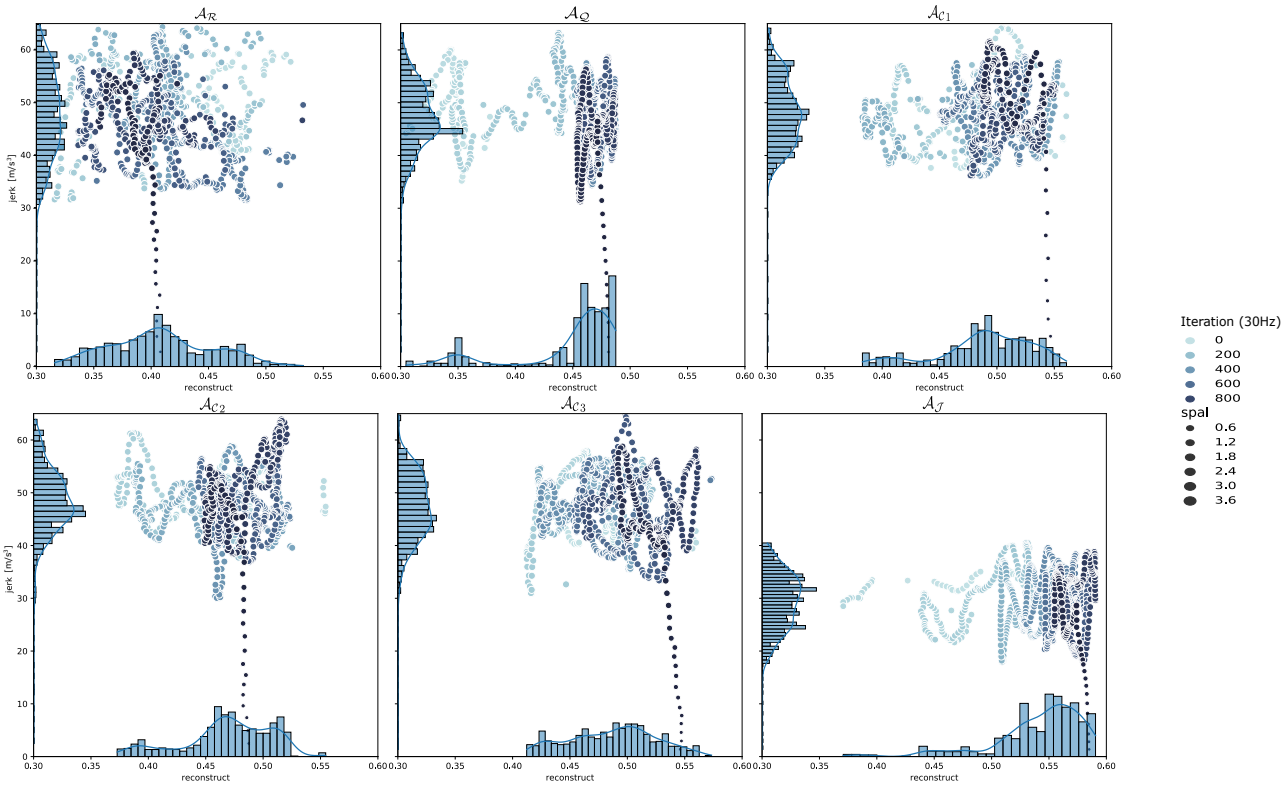

Fig. 8: The relational plots demonstrate the relationship between reconstructability $\mathbf{M_R}$ and jerkiness $\mathbf{M_J}$ across time frames t (color of dots) and different spectral arc length values $\mathbf{M_S}$ (size of dots). Along the two axes of each subgraph are histograms of $\mathbf{M_R}$ and $\mathbf{M_J}$ respectively. Results from the static baseline $\mathcal{A}_\mathcal{S}$ is not shown since it has a constant jerk value of 0.

B. Reconstructability Under Smooth Motion

The proposed smoothness constrained algorithm $\mathcal{A}_\mathcal{J}$ has shown significant success in reducing motion jerk for all three cameras while maintaining decent visual coverage and outstanding reconstructability of the dynamic surgical scene. This is evident in Table I, where $\mathcal{A}_\mathcal{J}$ was the top algorithm in four out of six metrics, the second runner-up for $\mathbf{M_{A-R}}$ after algorithms $\mathcal{A}_\mathcal{S}$ and $\mathcal{A}_\mathcal{Q}$, and the fourth in rank for $\mathbf{M_V}$ just below the $\mathcal{A}_\mathcal{C}$ series. The 'curiosity-driven' nature of the $\mathcal{A}_\mathcal{C}$ series motivates the cameras to explore unseen (or highly dynamic) areas, which can be a trade-off against repeated visits to known regions in the surgical scene. With that said, the higher $\mathbf{M_V}$ and lower $\mathbf{M_{A-R}}$ scores are to be expected for the three $\mathcal{A}_\mathcal{C}$ baselines. Demonstrating a more balanced exploration and exploitation trade-off, the added smoothness constraint in $\mathcal{A}_\mathcal{J}$ is proven to improve $\mathbf{M_{A-R}}$ compared to its algorithmically closest baseline competitor, $\mathcal{A}_{\mathcal{C}_\ni}$.

In addition to numeric analysis, the correlation between $\mathbf{M_R}$ and $\mathbf{M_J}$ across time frames using each algorithm is captured in Fig. 8. Below are a few observations:

- The 'tail' that shrinks in size and points downward in each subgraph comes from a common decrease in $\mathbf{M_J}$ and $\mathbf{M_S}$ scores toward the end of the surgical sequence.
- Other than the random baseline $\mathcal{A}_\mathcal{R}$, all other algorithms show an improving tendency in $\mathbf{M_R}$ as time progresses, as concluded from the progressively darker blue dots from the left to the right in each subgraph.
- While $\mathcal{A}_\mathcal{Q}$ was the fastest to reach the algorithm specific best $\mathbf{M_R}$ score, the reconstructability oscillations as show in the $\mathcal{A}_\mathcal{C}$ series (by the darker dots in the low $\mathbf{M_R}$ region) are evidence of 'curious exploration' of unseen areas in the surgical scene.
- Even with the smallest $\mathbf{M_J}$ throughout the surgical sequence, the 'tail' for $\mathcal{A}_\mathcal{J}$ lands at the highest $\mathbf{M_R}$ score compared to the baselines. This shows the outstanding capability of $\mathcal{A}_\mathcal{J}$ to reconstruct the dynamic surgical scene despite its camera motion smoothness constraints.

Together, the positive results speak directly to the operator comfort hypotheses (a)-(c) defined in Section IV.

C. Minimal Viewpoint Switching

Apart from the promising performances in both camera trajectory smoothness and scene reconstructability, $\mathcal{A}_\mathcal{J}$ also requires minimal viewpoint switching to maintain a comfortable view of the surgical site. Assuming a viewpoint is deemed comfortable to stare at when its smoothness score falls below the comfort threshold, Figure 5 confirms the user comfort hypothesis (b) in Section IV by showing the fewest white gaps in the top color bars of the $\mathcal{A}_\mathcal{J}$ subgraphs.

V. Conclusion

This work demonstrated the feasibility of motion smoothness constraints in a deep-reinforcement, curiosity learning framework for multicamera positioning in RMIS. The results show trajectory jerk reduction of at least 59.98% in $\mathbf{M_J}$, and 5.03% in $\mathbf{M_S}$, while maintaining quality reconstructability scores as observed through $\mathbf{M_R}$ and $\mathbf{M_{T-R}}$. Development towards this end is a step towards sim-to-real translation and clinical adoption of intelligent multicamera systems.

REFERENCES

[1] Y. Itatani, K. Obama, T. Nishigori, R. Ganeko, S. Tsunoda, H. Hosogi, S. Hisamori, K. Hashimoto, and Y. Sakai, "Three-dimensional stereoscopic visualization shortens operative time in laparoscopic gastrectomy for gastric cancer," *Scientific reports*, vol. 9, no. 1, pp. 1–7, 2019.

[2] C. Sui, J. Wu, Z. Wang, G. Ma, and Y.-H. Liu, "A real-time 3d laparoscopic imaging system: design, method, and validation," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 9, pp. 2683–2695, 2020.

[3] Y.-H. Su, K. Lindgren, K. Huang, and B. Hannaford, "A comparison of surgical cavity 3d reconstruction methods," in *2020 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2020, pp. 329–336.

[4] M. Silvestri, T. Ranzani, A. Argiolas, M. Vatteroni, and A. Menciassi, "A multi-point of view 3d camera system for minimally invasive surgery," *Sensors and Actuators A: Physical*, vol. 202, pp. 204–210, 2013.

[5] G. Yin, W. K. Han, S. Faddegon, Y. K. Tan, Z.-W. Liu, E. O. Olweny, D. J. Scott, and J. A. Cadeddu, "Laparoendoscopic single site (less) in vivo suturing using a magnetic anchoring and guidance system (mags) camera in a porcine model: impact on ergonomics and workload," *Urology*, vol. 81, no. 1, pp. 80–84, 2013.

[6] M. Morgan, E. O. Olweny, and J. A. Cadeddu, "Less and notes instrumentation: future," *Current opinion in urology*, vol. 24, no. 1, pp. 58–65, 2014.

[7] P. Valdastri, C. Quaglia, E. Buselli, A. Arezzo, N. Di Lorenzo, M. Morino, A. Menciassi, P. Dario, *et al.*, "A magnetic internal mechanism for precise orientation of the camera in wireless endoluminal applications," *Endoscopy*, vol. 42, no. 6, p. 481, 2010.

[8] N. Di Lorenzo, L. Cenci, M. Simi, C. Arcudi, V. Tognoni, A. L. Gaspari, and P. Valdastri, "A magnetic levitation robotic camera for minimally invasive surgery: Useful for notes?" *Surgical endoscopy*, vol. 31, no. 6, pp. 2529–2533, 2017.

[9] Y. Burda, H. Edwards, D. Pathak, A. Storkey, T. Darrell, and A. A. Efros, "Large-scale study of curiosity-driven learning," *arXiv preprint arXiv:1808.04355*, 2018.

[10] Y.-H. Su, H. Zhang, W. Jiang, K. Ngo, and K. Huang, "Deep curiosity driven multicamera 3d viewpoint adjustment for robot-assisted minimally invasive surgery," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 7724–7730.

[11] Y. H. Su, K. Huang, and B. Hannaford, "Multicamera 3d reconstruction of dynamic surgical cavities: Camera grouping and pair sequencing," in *Medical Robotics (ISMR), 2019 International Symposium on*. IEEE, 2019.

[12] ——, "Multicamera 3d reconstruction of dynamic surgical cavities: Autonomous optimal camera viewpoint adjustment," in *Medical Robotics (ISMR), 2020 International Symposium on*. IEEE, 2020.

[13] ——, "Multicamera 3d reconstruction of dynamic surgical cavities: Non-rigid registration and point classification," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 7911–7918.

[14] G. Rousseau, C. S. Maniu, S. Tebbani, M. Babel, and N. Martin, "Quadcopter-performed cinematographic flight plans using minimum jerk trajectories and predictive camera control," in *2018 European Control Conference (ECC)*. IEEE, 2018, pp. 2897–2903.

[15] M. Grundmann, V. Kwatra, and I. Essa, "Auto-directed video stabilization with robust l1 optimal camera paths," in *CVPR 2011*. IEEE, 2011, pp. 225–232.

[16] S. Pandey, M. D. Byrne, W. H. Jantscher, M. K. O'Malley, and P. Agarwal, "Toward training surgeons with motion-based feedback: initial validation of smoothness as a measure of motor learning," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 61, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2017, pp. 1531–1535.

[17] D. Chen, S. Li, W. Li, and Q. Wu, "A multi-level simultaneous minimization scheme applied to jerk-bounded redundant robot manipulators," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 1, pp. 463–474, 2019.

[18] K. Huang, D. Chitrakar, F. Rydén, and H. J. Chizeck, "Evaluation of haptic guidance virtual fixtures and 3d visualization methods in telemanipulation—a user study," *Intelligent Service Robotics*, vol. 12, no. 4, pp. 289–301, 2019.

[19] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *International conference on machine learning*. PMLR, 2016, pp. 1928–1937.

[20] D. Pathak, P. Agrawal, A. A. Efros, and T. Darrell, "Curiosity-driven exploration by self-supervised prediction," in *International conference on machine learning*. PMLR, 2017, pp. 2778–2787.

[21] Y.-H. Su, K. Huang, and B. Hannaford, "Multicamera 3d viewpoint adjustment for robotic surgery via deep reinforcement learning," *Journal of Medical Robotics Research*, p. 2140003, 2021.

[22] N. Hogan and D. Sternad, "Sensitivity of smoothness measures to movement duration, amplitude, and arrests," *Journal of Motor Behavior*, vol. 41, no. 6, pp. 529–534, 2009.

[23] S. Balasubramanian, A. Melendez-Calderon, and E. Burdet, "A robust and sensitive metric for quantifying movement smoothness," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2126–2136, 2012.

[24] S. Balasubramanian, A. Melendez-Calderon, and A. e. a. Roby-Brami, "On the analysis of movement smoothness," *Journal of NeuroEngineering and Rehabilitation*, vol. 12, no. 112, 2012.