



OPEN 3D vs. 2D simulated fetoscopy for spina bifida repair: a quantitative motion analysis

Mirza Awais Ahmad^{1,2,6}✉, Yolán Weiler^{1,6}, Luc Joyeux^{2,6}, Elisenda Eixarch^{3,4}, Tom Vercauteren⁵, Sebastien Ourselin⁵, Jan Deprest² & Emmanuel Vander Poorten¹

3D imaging technology is becoming more prominent every day. However, more validation is needed to understand the actual benefit of 3D versus conventional 2D vision. This work quantitatively investigates whether experts benefit from 3D vision during minimally invasive fetoscopic spina bifida (fSB) repair. A superiority study was designed involving one expert team (> 18 procedures prior) who performed six 2D and six 3D fSB repair simulations in a high-fidelity animal training model, using 3-port access. The 6D motion of the instruments was recorded. Among the motion metrics are total path length, smoothness, maximum speed, the modified Spectral Arc Length (SPARC), and Log Dimensionless Jerk (LDLJ). The primary clinical outcome is operation time (power 90%, 5% significance) using Sealed Envelope Ltd. 2012. Secondary clinical outcomes are water tightness of the repair, CO₂ insufflation volume, and OSATS score. Findings show that total path length and LDLJ are considerably different. Operation time during 3D vision was found to be significantly shorter compared to 2D vision (113 ± 9 vs. 149 ± 19 min; p = 0.026). These results suggest enhanced performance with 3D vision during interrupted suturing in fetoscopic SBA repair. To confirm these results, a larger-scale follow-up study involving multiple experts and novice surgeons is recommended.

Stereoscopy is an imaging technology enabling three-dimensional (3D) vision. It uses two cameras that capture the same scene from two different viewpoints, similar to humans, where images from both eyes, situated at an offset, perceive depth. If the left/right images from the stereo camera are “displayed” exclusively to the corresponding left/right eye, the human visual cortex can similarly interpret the different viewpoints and perceive depth in the stereo images.

This study investigates whether 3D imaging is useful for minimally invasive fetoscopic surgery (MIS). Spina bifida repair serves as a case study. Related literature investigating the benefits of 3D visualization reported that passive polarised glasses produce more conclusive results (in favor of 3D) in laparoscopic interventions¹. In 3D, complex tasks can be executed more easily and with less stress^{2–4}, and operation time and error reduced⁵. Better surgical performance and lower stress are reported⁶, whereas reduced tool path length and lower stress measured via NASA’s Task Load Index⁷. Many studies comparing 3D and 2D endoscopy showed significant differences by using these outcome metrics, however, *motion metrics* could provide an additional and possibly more objective assessment. This work analyses the surgeon’s motion in fetoscopic spina bifida repair allowing for a truly objective assessment of performance and skill.

Spina bifida (SB) aperta is a congenital neural malformation that represents about 10% of all spinal deformities causing physical and mental impairment which require intense lifelong medical follow-up⁸. It is caused by incomplete closure of the neural tube, leading to exposure of neural elements to the amniotic fluid. This causes trauma and damage to the spinal cord and nerves, resulting in motor function loss of the limbs⁹. Treatment options are either postnatal or prenatal repair. Prenatal repair halts and prevents ongoing neurodegeneration and has been shown to improve the outcome, as demonstrated in the management of myelomeningocele study (MOMS) trial¹⁰. In that study, repair was done through laparotomy and hysterotomy and layered anatomical repair, referred to as open fetal surgery (OFS). To avoid complications from hysterotomy, like uterine rupture

¹Department of Mechanical Engineering Sciences, Catholic University of Leuven, 3000 Leuven, Belgium. ²Obstetrics and Gynaecology, University Hospital of Leuven, 3000 Leuven, Belgium. ³BCNatal Fetal Medicine Research Center, Hospital Clinic, Hospital Sant Joan de Déu, Universitat de Barcelona, Barcelona, Spain. ⁴Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centre for Biomedical Research on Rare Diseases (CIBERER), Barcelona, Spain. ⁵Department of Imaging and Biomedical Engineering, Kings College, London WC2R 2LS, UK. ⁶These authors contributed equally: Mirza Awais Ahmad, Yolán Weiler and Luc Joyeux. ✉email: mirzaawais.ahmad@kuleuven.be

in subsequent pregnancies, significant efforts have been made to make this operation minimally invasive (MIS) using fetoscopic instrumentation¹¹.

Prenatal MIS fetoscopic repair may reduce maternal and fetal risks, however, an MIS approach is more challenging than its open counterpart. The vulnerable surgical environment and difficult closure techniques are combined with common disadvantages of endoscopic approaches: the fulcrum effect that inverts motion, the reduction of motion to only four degrees of freedom, poor tactile feedback, loss of depth perception, and difficult hand-eye coordination together leading to increased operation times. 3D imaging techniques could help the surgeon regain the sense of depth and improve efficiency.

This work exploits quantitative motion metrics to objectively assess whether 3D fetoscopy actually benefits expert surgeons in performing prenatal MIS for fSB-repair. Our hypothesis is that the skills of expert surgeons improve when performing fetoscopic repair with 3D vision rather than 2D vision and that these improvements are measurable using quantitative motion metrics.

Methods

Study design

A superiority study was designed involving one expert laparoscopic fetal surgeon (JD) and assistant (LJ) performing a simulated fetoscopic SB repair in a high-fidelity animal training model using 3-port access¹². The primary outcome of this experiment was total operation time (TOT) which includes the preparation of the animal model until the end of the last step of the surgery. Based on estimated TOT, six surgeries per group (2D vs. 3D) were required ($\beta = 90\%$, $\alpha = 5\%$). Secondary clinical outcomes were water tightness at the end of the repair, total volume CO₂ used, and OSATS score. To document the performance of the surgeon in terms of instrument use, 39 motion metrics were computed including path length, smoothness, maximum speed, the modified Spectral Arc Length (SPARC), Log Dimensionless Jerk (LDLJ), Number of Speed Peeks (NSP) and Number of Acceleration Peeks (NAP)^{13–20}. These metrics are calculated for the surgeon's left hand (LH), right hand (RH) and the assistant's hand holding the fetoscope (FH).

Since clinically, today's fetoscopic surgeons would typically transit from 2D to 3D vision^{21,22}, the surgeon first operated in 2D after which all 3D experiments were performed. Prior to the first operation herein, the surgeon conducted 18 interventions on the model. An earlier study using competency cumulative sum (C-CUSUM) analysis¹² demonstrated competency after six consecutive procedures.

Training model

The experiments used a high-fidelity training model as earlier described by Joyeux et al.¹². In an adult male New Zealand rabbit animal model all operative steps and surgical conditions present in current clinical multi-layered fetoscopic spina bifida repair, can be simulated. The rabbit's abdominal cavity mimics the available intra-uterine workspace (15 × 10 × 5 cm) when insufflated with 3 l, of CO₂ at a pressure of 5 mmHg²³. The back of the fetus (22–24 weeks of gestation) is represented by the rabbit stomach²⁴. In this model, laparoscopic gastric Nissen fundoplication²⁵ the ten surgical steps required for fetoscopic spina bifida repair, including gastric wall patch suturing, can be mirrored, and necessitate delicate tissue handling, thorough dissection, hemostasis, and suturing.

Surgical procedure

We used adult New-Zealand rabbits and standard pediatric laparoscopic instruments and a 3D endoscope (Karl Storz™ 4 mm scope, TIPCAM1 S, 7240 BA3D). Four main motion sequences were identified containing similar movements to fetoscopic repair. These sequences are used for comparison of 3D vs. 2D motion-based skill:

- I. Dissection movements: the surgeon performs a series of explorative, dissection and grasping movements to free up the esophagogastric junction (EGJ) in preparation of the fundoplication.
- II. High-force movements: the surgeon creates an anti-reflux valve by pulling the upper part of the stomach (called the fundus) behind the EGJ.
- III. Interrupted suturing: the surgeon performs the fundoplication with a series of interrupted sutures.
- IV. Running sutures: the surgeon sutures a patch on the large anterior curvature of the rabbit stomach with two semi-circular running sutures.

The secondary outcomes included the individual operation time for motion sequences I, II, III, and IV, and their sum which equals the fetal operation time (FOT) and the time spent on operative steps, while the TOT includes the preparation and conclusion steps as well.

Experimental setup

Figure 1 shows the experimental setup. The Aurora (NDI, Ontario, Canada) electromagnetic field generator (FG) henceforth referred to as the Aurora tracker, measures the 3D position and 3D orientation of the LH tool, the RH tool, the fetoscope, and the fetoscope port. All instruments had a dedicated electromagnetic tracking (EMT) sensor attached. Because the Aurora tracker has only four input ports, only four sensors can be tracked simultaneously. Therefore, during some instrument changes sensors needed to be dis- and reconnected. Upon instrument switch or when a surgical step or phase was completed, Aurora recordings were stopped. This results in each trial being segmented in multiple motion recordings. Notes were made during the surgery to keep track of the recorded segments or unexpected events. When non-connected instruments were used briefly, the

connections were not changed. This results inevitably in motion recordings having partially invisible (sensor out of range) segments.

The endoscope used in this study was a Karl Storz™ 4 mm scope (TIPCAM1 S, 30° angle, 18 cm length, 7240 BA3D) which allows to easily switch from 2D to 3D vision²⁶. To perform 3D surgery, we used the same HD monitor and operators were wearing the Karl Storz™ 3D glasses. Unlike autostereoscopy²⁷, the HD monitor allows multiple users to have a stereo view as long as they wear the 3D glasses. During surgery, one surgical assistant manipulated the fetoscope while the expert surgeon performed the operation. Seven laparoscopic instruments were used. Video recordings of the full surgical scene and fetoscope were made.

Motion metrics and clinical outcomes

Motion can be analyzed to assess the surgeon’s performance under different circumstances¹⁴. Table 1 lists the different motion metrics that were employed alongside the equations to calculate them. The metrics were computed for the LH, RH and fetoscope, and port motions. The SPARC, LDLJ, path length and maximum, mean, and consistency of the speed and acceleration are time-independent motion metrics while the others are time-dependent. Time independence is an important characteristic for a metric to truly represent skill.

The EM sensors may have been only visible for a duration of time that is relevant for a specific segment of the procedure. For this, the visibility ratios v_i are calculated to provide a metric for the reliability of the motion data. Here, the motion data is considered “visible” even if one of the three (RH, LH, and fetoscope) instrument’s EM sensor measurements are available for a relevant surgical step.

Motion data processing

Both the motion and video data are used in the process to calculate the motion metrics. Figure 2 shows the process with three main steps: pre-processing, processing, and debugging. These steps are computed with MATLAB®. The goal of pre-processing is to transform the raw motion data captured from the experiment into motion data that is suited for calculating the motion metrics. Pre-processing can be further split into five separate steps:

Metric	Equation
Total time	$T_i \triangleq \int_{t_{i,start}}^{t_{i,end}} dt = t_{i,end} - t_{i,start}$
Path length	$P_i \triangleq \int_0^{T_i} v_i(t) dt$
Volume	$V_i = \text{convhull}(p_i(t))$
Maximum speed	$S_{i,max} \triangleq \max(v_i(t))$
Mean speed	$\bar{S}_i \triangleq \text{mean}(v_i(t))$
Speed consistency	$\sigma_{S_i} \triangleq \text{std}(v_i(t))$
Number of speed peaks	$NSP_{i,\alpha} \triangleq \{v_i(t) \mid \frac{dv_i(t)}{dt} = 0, \frac{d^2v_i(t)}{dt^2} < 0, v_i(t) > \alpha \bar{S}_i\} $
Speed peak rate	$SPR_{i,\alpha} = \frac{NSP_{i,\alpha}}{T_i}$
Movement arrest period ratio	$MAPR_{i,\alpha} \triangleq \frac{T_{i,v_i(t) > (\alpha/100)S_{i,max}}}{T_i}$
Maximum acceleration	$A_{i,max} \triangleq \max(a_i(t))$
Mean acceleration	$\bar{A}_i \triangleq \text{mean}(a_i(t))$
Acceleration consistency	$\sigma_{A_i} \triangleq \text{std}(a_i(t))$
Number of acceleration Peaks	$NAP_{i,\alpha} \triangleq \{a_i(t) \mid \frac{da_i(t)}{dt} = 0, \frac{d^2a_i(t)}{dt^2} < 0, a_i(t) > \alpha \bar{A}_i\} $
Acceleration peak rate	$APR_{i,\alpha} = \frac{NAP_{i,\alpha}}{T_i}$
Integral of acceleration vector	$IAV_i \triangleq \int_0^{T_i} a_i(t) dt$
Smoothness	$Sm_i \triangleq \frac{1}{T_i} \sqrt{\frac{1}{2} \int_0^{T_i} \left[\left(\frac{d^3x}{dt^3} \right)^2 + \left(\frac{d^3y}{dt^3} \right)^2 + \left(\frac{d^3z}{dt^3} \right)^2 \right] dt}$
Spectral arc length	$SAL_i \triangleq - \int_0^{\omega_c} \sqrt{\left(\frac{1}{\omega_c} \right)^2 + \left(\frac{d\hat{V}_i(\omega)}{d\omega} \right)^2} d\omega, \hat{V}_i(\omega) = \frac{\hat{V}_i(\omega)}{\hat{V}_i(0)}$
Modified spectral arc length	$SPARC_i \triangleq - \int_0^{\omega_c} \sqrt{\left(\frac{1}{\omega_c} \right)^2 + \left(\frac{d\hat{V}_i(\omega)}{d\omega} \right)^2} d\omega, \hat{V}_i(\omega) = \frac{\hat{V}_i(\omega)}{\hat{V}_i(0)}, \omega_c \triangleq \min\{\omega_c^{max} \mid \min\{\omega \mid \hat{V}_i(r) < \bar{V}_i, \forall r > \omega\}\}$
Dimensionless jerk	$DLJ_i \triangleq \frac{(T_i)^3}{v_{i,peak}^2} \int_0^{T_i} j_i(t)^2 dt$
Log dimensionless jerk	$LDLJ_i \triangleq \ln DLJ_i $
Visibility ratio	$v_i \triangleq \frac{T_{motion,i}}{T_i}$

Table 1. Metrics used for quantitative motion analysis. The $|\cdot|$ denotes the cardinal of a set and i denotes the metric corresponding to the i th segment. The function `convhull()` is a built-in MATLAB® (Mathworks, Massachusetts, US) function.

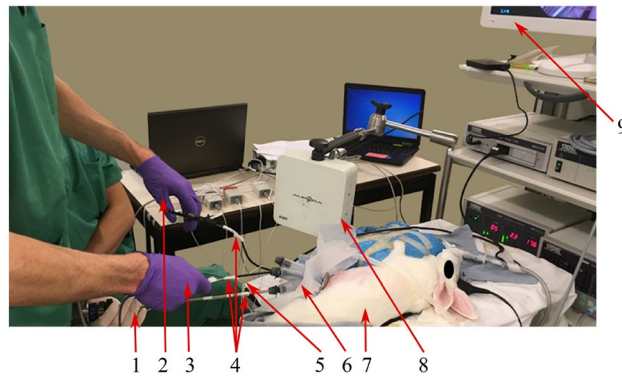


Figure 1. Set-up of experiment. (1) Assistant's hand holding the fetoscope. (2) Left hand (LH) of expert surgeon holding a tool. (3) Right hand (RH) of expert surgeon holding a tool. (4) Aurora sensors taped to LH tool, RH tool, fetoscope. (5) Operating table corner used as a reference for Aurora tracker. (6) The fetoscope port with aurora sensor attached. (7) The male adult New Zealand rabbit. (8) Aurora tracker magnetic field generator (FG). (9) The display showing images from the fetoscope.

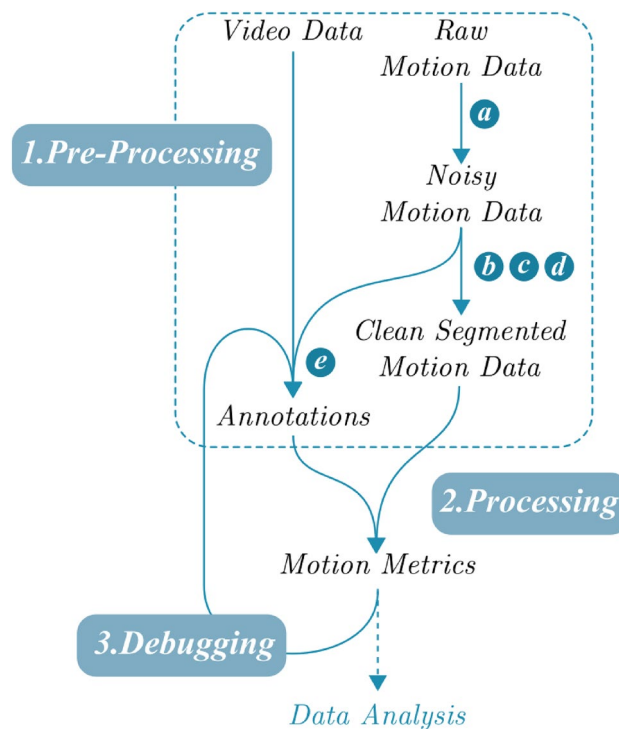


Figure 2. Data processing schematic. (a) Formatting, (b) transformations, (c) visibility, (d) smoothing and (e) annotations. The pre-processing, processing and debugging steps are performed using MATLAB[®].

- Formatting (a) data in the right structure. The stored data files (encrypted ROS bag files) from the Aurora tracker is transformed into a readable database.
- Transforming (b) the data to the correct reference frame. A corner of the operating table (number 5 in Fig. 1) is used as the base reference frame for all sensors.
- Pruning (c) invisible data. Motion data that was invisible for more than 1 second is removed from the data to not affect the metrics. Segments that were invisible for less than a second are filled using quadratic spline interpolation based on the visible data points before and after the invisible segment.
- Smoothing (d) the data. A butterworth filter ($f_c = 10$ Hz, $n = 5$) and an exponential filter ($\alpha = 0.05$) are applied to the motion data.
- Annotating (e) the data. Starting and ending timestamps of relevant surgical step segments in motion data are noted manually based on the fetoscope's video.

Processing consists of calculating the required motion metrics for each of the four global motion sequences mentioned above. In order to do this, it is necessary to aggregate the results of the smaller motion segments i generated during step c. Depending on the motion metric, aggregation is done differently. For example, the path length P_i of each segment is summed together to represent the total path length of the sequence, whereas the mean speed \bar{S}_i of each segment is averaged to find the overall mean speed. The maximum speed $S_{i,max}$ is aggregated by taking the maximum speed of all segments. The last step, named debugging ensures that the quadratic interpolation adopted to process the invisibility is performed accurately: If the interpolation generated a maximum acceleration above 6 ms^{-2} , which is outside the range of human-produced accelerations, this part of motion data is not included in the analysis.

Statistical data analysis

The null hypothesis states that the means of the motion metrics obtained in the 3D group are equal to those in the 2D group. This requires an unpaired t-test or, in the case the metrics were not normally distributed or when the variances σ^2 were not equal, a parametric Wilcoxon rank-sum test. To test for normality and equal variances we used the Saphiro-Wilk test and the F-test respectively. The software GraphPad Prism 8.4.2. (Dotmatics, Boston, MA, USA) was used for processing these tests.

Ethical approval statement

This experiment was approved by the Animal Ethics Committee of the Group Biomedical Sciences of the KU Leuven (P093-2016). The experiments were done following the ARRIVE guidelines for animal research^{28,29}, and the guidelines of the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs).

Results

Clinical outcomes

For each clinical outcome the mean for both the 2D (μ_{2D}) and 3D (μ_{3D}) groups are shown in Table 2, as well as their difference $\mu_{(2D-3D)}$. A 95% confidence interval (CI) is reported as well. All execution times are lower when using 3D and TOT, as well as FOT, are significantly reduced in the 3D group with an average reduction of respectively 35 and 24 minutes. Taking into account the different steps in the surgery, the operation time for the dissection movements (I) and interrupted suturing (III) show an even more obvious difference. For the dissections, a 40% reduction is appearing, and the time for interrupted suturing is reduced from an average time of 21.7 minutes in the 2D group to 13.2 minutes in the 3D group, which is more than a 30% reduction. No differences are found for the OSATS score nor for the insufflation volume.

Quantitative motion outcomes

Visibility

The visibility ratios of the motion recordings are shown in Table 3. Sequences III and IV are well recorded with an average visibility of respectively 91% and 94%. Sequences I and II are less well recorded due to missing recordings and lower visibility. The bold values show visibility below 80%, suggesting the motion data is less reliable.

Recordings for trial 3 for the 2D group as well as a segment in trial 5 for both 2D and 3D groups are missing due to technical issues. The asterisk (*) marks a potential mismatch between motion signals and surgical movements due to the difficulties with the annotations. In these cases the motion segment was chosen to be slightly longer than the actual movement on the video, to make sure that the motions typical for the surgical procedure are also included in the final sequence. The downside of this approach is that motions not representative of relevant surgical step might get introduced as well, reducing the data quality. Instead of a visibility ratio greater than 1, the visibility is set to be equal to the average visibility of that sequence.

Another way to check the reliability of the motion data is to compare the total motion time from Figs. 3 and 4 with the clinical fetal operation time from Table 2. Indeed, the previously reported clinical operation time is showing a considerable difference for the dissection motions (I) and interrupted suturing motions (III). For the latter, the motion-based total time is also demonstrating a considerable difference, confirming the clinically

Outcome	μ_{2D}	[95% CI]	μ_{3D}	[95% CI]	μ_{2D-3D}	[95% CI]	P value
OSATS (/25)	18.5	[15.8, 21.2]	19.3	[15.8504, 22.7]	-.833	[-5.83, 4.16]	0.718
CO ₂ Vol. (l)	255	[165, 345]	184	[118, 250]	71.2	[-55.2, 198]	0.238
OT I: Dissection motions	30.2	[16.8, 43.6]	17.8	[14, 21.6]	12.3	[7.15, 40.8]	0.001**
OT II: High-force motions	10.5	[6.44, 14.6]	8.83	[6.05, 11.6]	1.67	[10.8, 59.9]	0.521
OT III: Interrupted suturing	21.7	[12.88, 3.52]	13.2	[1.6912, 15.7]	8.5	[5.04, 12]	<0.001***
OT IV: Running suturing	56.2	[52.2, 6.24]	54.7	[45, 59.4]	1.5	[-9.89, 12.9]	0.775
OT I-IV: Foetal	119	[117, 12.7]	94.5	[88.2, 10.8]	24	[-3.92, 7.25]	0.011*
Total OT incl. model preparation	149	[130, 168.5]	113	[104, 122]	35.3	[6.04, 18.6]	0.026*

Table 2. Unpaired t-test for clinical outcomes. $\alpha = 0.05$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, bold: Wilcoxon rank sum test instead of unpaired t-test. OT operation time.

Metric	μ_{2D}	[95% CI]	μ_{3D}	[95% CI]	μ_{2D-3D}	[95% CI]	P-value
Total Time	960	[786, 1130]	432	[316, 548]	528	[357, 699]	<.001***
Path Length	18.4	[12.4, 24.4]	8.98	[6.56, 11.4]	9.38	[4.37, 14.4]	.004**
Smoothness	0.78	[0.106, 1.45]	0.624	[0.43, 0.818]	0.156	[-0.374, 0.686]	.523
SAL	-22	[-28.9, -15.1]	-16.8	[-21.9, -11.7]	-5.25	[-12.3, 1.84]	.247
SPARC	-7.78	[-8.82, -6.74]	-7.38	[-8.9, -5.86]	-0.397	[-2.06, 1.27]	.602
LDLJ	-29.6	[-30.3, -29]	-28.4	[-29, -27.7]	-1.29	[-2.11, -0.459]	.007**
Max Speed	0.343	[0.188, 0.498]	0.206	[0.158, 0.255]	0.137	[0.0131, 0.26]	.03*
NSP0	8090	[5280, 10900]	3160	[2240, 4080]	4930	[2680, 7180]	<.001***
NSP1	3770	[2130, 5420]	1440	[985, 1900]	2330	[1040, 3620]	.003**
NSP2	1430	[723, 2140]	515	[295, 734]	917	[354, 1480]	.005**
NSP5	108	[65.4, 151]	29.3	[12.9, 45.7]	79.1	[43.6, 115]	<.001***
NSP10	8.8	[-2.28, 19.9]	0.833	[-1.31, 2.98]	7.97	[-0.448, 16.4]	.013*
SPRate0	8.32	[6.77, 9.86]	7.3	[6.78, 7.81]	1.02	[-0.221, 2.26]	.096
SPRate1	3.84	[2.82, 4.87]	3.31	[2.94, 3.68]	0.534	[-0.304, 1.37]	.183
SPRate2	1.45	[0.98, 1.92]	1.17	[0.853, 1.49]	0.279	[-0.185, 0.744]	.207
SPRate5	0.114	[0.0665, 0.162]	0.0667	[0.0331, 0.1]	0.0474	[-0.000546, 0.0954]	.052
SPRate10	0.0101	[-0.00468, 0.0248]	0.00203	[-0.0032, 0.00726]	0.00802	[-0.00395, 0.02]	.043*
Mean Speed	0.019	[0.0156, 0.0223]	0.0208	[0.0183, 0.0234]	-0.00188	[-0.00537, 0.00161]	.254
Speed Consist.	0.0199	[0.015, 0.0247]	0.019	[0.0157, 0.0222]	0.000928	[-0.00384, 0.00569]	.67
MAPR10	0.315	[0.188, 0.443]	0.451	[0.376, 0.527]	-0.136	[-0.255, -0.0166]	.03*
MAPR25	0.0726	[0.0308, 0.114]	0.12	[0.0797, 0.16]	-0.0471	[-0.0968, 0.00246]	.06
MAPR50	0.0101	[0.00493, 0.0153]	0.0158	[0.0088, 0.0227]	-0.00563	[-0.0134, 0.00215]	.136
MAPR75	0.00207	[0.00146, 0.00268]	0.00273	[0.00157, 0.0039]	-0.000661	[-0.00188, 0.000561]	.429
MAPR90	0.0007	[0.000435, 0.000964]	0.00083	[0.000662, 0.000998]	-0.00013	[-0.000384, 0.000124]	.276
Max Accel	6.23	[3.1, 9.35]	3.3	[2.41, 4.19]	2.93	[0.473, 5.38]	.024*
NAP0	11600	[8220, 15000]	4720	[3370, 6070]	6880	[4070, 9690]	<.001***
NAP1	6280	[4430, 8140]	2640	[1870, 3420]	3640	[2080, 5200]	<.001***
NAP2	2540	[1580, 3500]	958	[557, 1360]	1580	[774, 2390]	.002**
NAP5	176	[98.9, 252]	63	[33, 93]	113	[49.1, 176]	.003**
NAP10	16.6	[2.27, 30.9]	3.33	[-1.53, 8.2]	13.3	[1.71, 24.8]	.026*
APRate0	12	[10.4, 13.5]	10.9	[10.3, 11.5]	1.08	[-0.176, 2.34]	.084
APRate1	6.48	[5.54, 7.43]	6.1	[5.7, 6.5]	0.386	[-0.409, 1.18]	.429
APRate2	2.6	[2.04, 3.16]	2.18	[1.58, 2.78]	0.423	[-0.29, 1.14]	.213
APRate5	0.187	[0.085, 0.29]	0.147	[0.0944, 0.199]	0.0407	[-0.0501, 0.132]	.337
APRate10	0.0185	[-0.00108, 0.0382]	0.00779	[-0.00387, 0.0194]	0.0108	[-0.0076, 0.0291]	.078
Mean Accel	0.319	[0.167, 0.47]	0.279	[0.234, 0.324]	0.0393	[-0.0803, 0.159]	.476
Accel Consist.	0.318	[0.178, 0.458]	0.258	[0.193, 0.323]	0.0601	[-0.061, 0.181]	.291
IAV	316	[115, 517]	121	[83.1, 159]	195	[42.6, 347]	.018*
Volume	0.00155	[-0.000547, 0.00364]	0.000619	[8.52e-05, 0.00115]	0.000929	[-0.000697, 0.00255]	.082

$\alpha = 0.05$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$. **bold**: Wilcoxon rank sum test, Spectral Arc Length (SAL), Modified Spectral Arc Length (SPARC), Log Dimensionless Jerk (LDLJ), Number of Speed Peaks with threshold equal to x times the average speed (NSPx), Speed Peak Rate (SPRx), Movement Arrest Period Ratio (MAPRx), Number of Acceleration Peaks (NAPx), Acceleration Peak Rate (APRx) and Integral of Acceleration Vector (IAV).

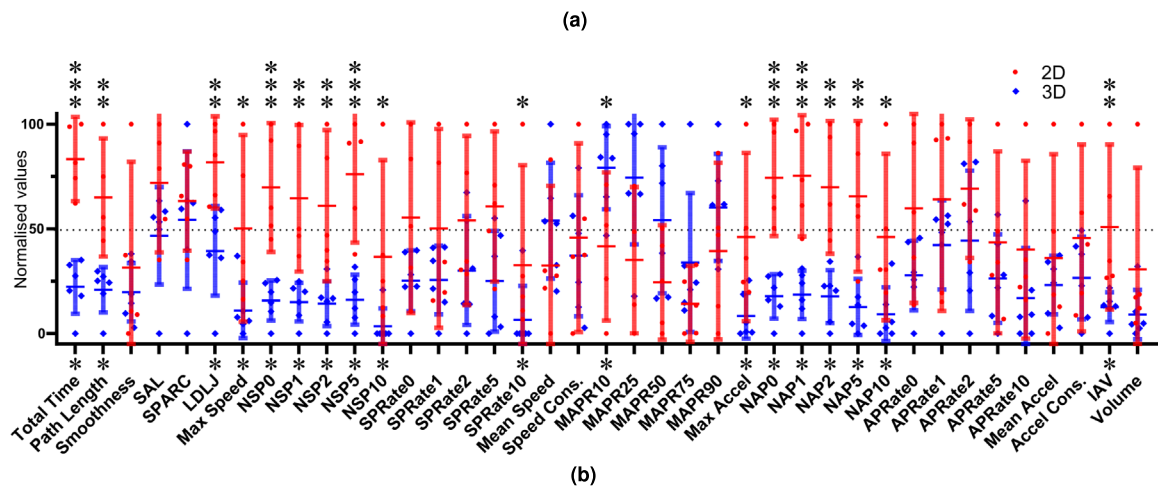


Figure 3. Right hand motions during interrupted suturing comparing 3D (blue) and 2D (red) vision. $n_{2D} = 5, n_{3D} = 6$.

Metric	μ_{2D}	[95% CI]	μ_{3D}	[95% CI]	μ_{2D-3D}	[95% CI]	P-value
Total Time	974	[778, 1170]	432	[316, 548]	542	[359, 726]	<.001***
Path Length	15.6	[11.6, 19.5]	7.39	[5.21, 9.56]	8.19	[4.61, 11.8]	<.001***
Smoothness	0.527	[0.21, 0.845]	0.476	[0.358, 0.594]	0.0512	[-0.209, 0.311]	.667
SAL	-23.5	[-34.2, -12.7]	-17	[-23.5, -10.4]	-6.49	[-16.7, 3.67]	.182
SPARC	-8.5	[-11.4, -5.58]	-7.74	[-8.87, -6.61]	-0.755	[-3.16, 1.66]	.497
LDLJ	-29.6	[-30.7, -28.4]	-28.1	[-29.1, -27.2]	-1.41	[-2.64, -0.177]	.029*
Max Speed	0.29	[0.0531, 0.527]	0.196	[0.127, 0.265]	0.0941	[-0.0926, 0.281]	.284
NSP0	8270	[5840, 10700]	3150	[2200, 4110]	5110	[3100, 7120]	<.001***
NSP1	3510	[2020, 5000]	1220	[818, 1630]	2290	[1120, 3450]	.002**
NSP2	1300	[903, 1700]	510	[347, 672]	794	[459, 1130]	<.001***
NSP5	86.8	[46.9, 127]	36.5	[18.2, 54.8]	50.3	[16, 84.6]	.009**
NSP10	2.8	[0.579, 5.02]	0.333	[-0.209, 0.875]	2.47	[0.75, 4.18]	.043*
SPRate0	8.43	[7.21, 9.65]	7.26	[6.76, 7.77]	1.17	[0.146, 2.19]	.029*
SPRate1	3.54	[2.6, 4.49]	2.8	[2.45, 3.15]	0.745	[-0.0306, 1.52]	.058
SPRate2	1.33	[1.08, 1.58]	1.17	[1.08, 1.26]	0.161	[-0.0397, 0.361]	.103
SPRate5	0.0915	[0.0475, 0.136]	0.0832	[0.0501, 0.116]	0.00832	[-0.0373, 0.054]	.69
SPRate10	0.00313	[0.000533, 0.00573]	0.00742	[-0.000482, 0.00197]	0.00239	[0.000137, 0.00464]	.043*
Mean Speed	0.016	[0.0141, 0.0179]	0.017	[0.015, 0.019]	-0.00104	[-0.00343, 0.00134]	.349
Speed Consist.	0.0168	[0.0129, 0.0207]	0.0178	[0.016, 0.0196]	-0.001	[-0.00435, 0.00235]	.247
MAPR10	0.376	[0.245, 0.508]	0.391	[0.337, 0.444]	-0.0144	[-0.124, 0.0956]	.773
MAPR25	0.109	[0.064, 0.153]	0.13	[0.095, 0.164]	-0.021	[-0.0681, 0.0262]	.341
MAPR50	0.0151	[0.00648, 0.0237]	0.0225	[0.00952, 0.0354]	-0.00738	[-0.0215, 0.0067]	.266
MAPR75	0.00236	[0.00114, 0.00357]	0.00275	[0.0012, 0.00431]	-0.000394	[-0.00215, 0.00136]	.624
MAPR90	0.000649	[0.000339, 0.000959]	0.000624	[0.000474, 0.000773]	2.54e-05	[-0.000245, 0.000296]	.837
Max Accel	3.91	[2.52, 5.3]	3.16	[1.71, 4.61]	0.747	[-0.993, 2.49]	.177
NAP0	11800	[8560, 15100]	4620	[3260, 5980]	7220	[4470, 9970]	<.001***
NAP1	6850	[4650, 9060]	2480	[1630, 3330]	4370	[2550, 6200]	<.001***
NAP2	2460	[1930, 2990]	946	[628, 1260]	1520	[1020, 2010]	<.001***
NAP5	105	[84.3, 126]	58.8	[36.6, 81.1]	46.6	[19.9, 73.2]	.003**
NAP10	5.2	[2.81, 7.59]	1.83	[0.801, 2.87]	3.37	[1.34, 5.39]	.013*
APRate0	12.1	[10.7, 13.5]	10.6	[10.1, 11.2]	1.43	[0.259, 2.6]	.022*
APRate1	6.97	[5.91, 8.02]	5.68	[4.82, 6.53]	1.29	[0.149, 2.43]	.031*
APRate2	2.53	[2.29, 2.77]	2.17	[1.93, 2.41]	0.362	[0.0683, 0.655]	.021*
APRate5	0.111	[0.0744, 0.148]	0.136	[0.098, 0.175]	-0.0253	[-0.0715, 0.0208]	.246
APRate10	0.00564	[0.00258, 0.0087]	0.0045	[0.00189, 0.00711]	0.00114	[-0.00225, 0.00453]	.467
Mean Accel	0.236	[0.177, 0.296]	0.223	[0.187, 0.258]	0.0137	[-0.0419, 0.0692]	.591
Accel Consist.	0.212	[0.171, 0.253]	0.21	[0.179, 0.241]	0.00208	[-0.0404, 0.0445]	.914
IAV	232	[141, 324]	97.3	[65.5, 129]	135	[60.9, 209]	.003**
Volume	0.000773	[0.000412, 0.00113]	0.000372	[0.000218, 0.000526]	0.000401	[9.62e-05, 0.000705]	.016*

$\alpha = 0.05$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, bold: Wilcoxon rank sum test, Spectral Arc Length (SAL), Modified Spectral Arc Length (SPARC), Log Dimensionless Jerk (LDLJ), Number of Speed Peaks with threshold equal to x times the average speed (NSPx), Speed Peak Rate (SPRx), Movement Arrest Period Ratio (MAPRx), Number of Acceleration Peaks (NAPx), Acceleration Peak Rate (APRx) and Integral of Acceleration Vector (IAV).

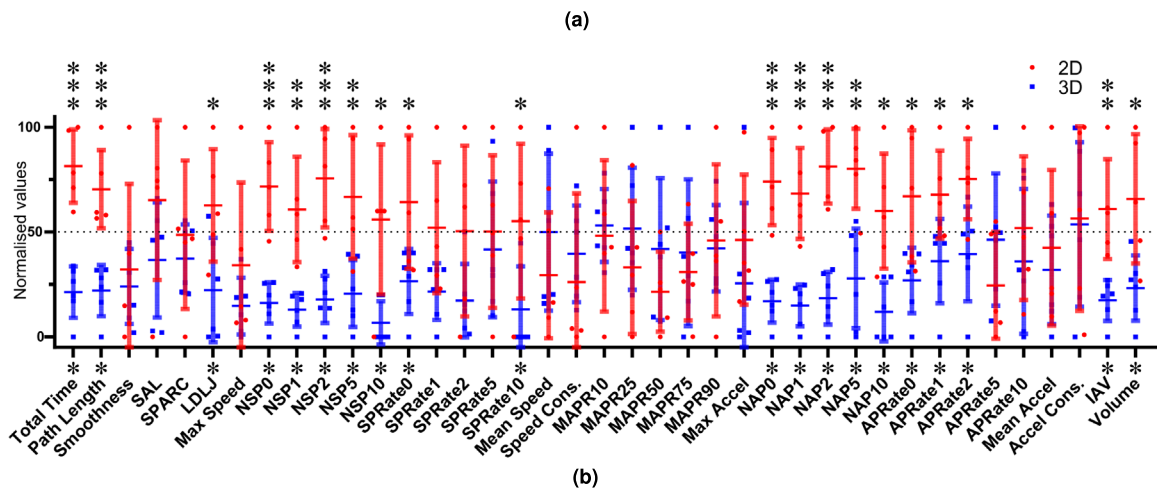


Figure 4. Left hand motions during interrupted suturing comparing 3D (blue) and 2D (red) vision. $n_{2D} = 5, n_{3D} = 6$.

Trial	1		2		3		4		5		6	
	2D	3D	2D	3D	2D	3D	2D	3D	2D	3D	2D	3D
I: Dissections	82.1	69.5	81.5	98.6	0	99.6	74.1	88.1	86.6	0	68.2	90.2
II: High force motions	87.1	98.4	86.5	88.5*	0	88.5*	91.4	79.6	0	76.9	81.3	99.1
III: Interrupted suturing	91.3	90.8	89.9	91.4*	0	86.7	96.1	99.1	81.4	98.9	94.4	81.5
IV: Running suturing	97.5	99.6	94.0	94.7	0	90.2	91.1	91.9	85.9	99.9	95.6	91.5

Table 3. Visibility (%) for all trials. *Visibility calculated differently than explained in chapter 3. Bold: visibility below 80%.

measured operation time. In the case of the dissection motions, the motion-based total time is showing a considerable difference for the right hand but not for the left hand, even though the left and right hand are equally used. This contrasts with the clinically measured operation time and therefore suggests that the motion data of this sequence is not totally reliable. The visibility analysis confirms these findings. Therefore only detailed results are shown in the following for the interrupted suturing motions.

Motion metrics

The results of the statistical analysis comparing the motion metrics from the 3D and 2D groups are reported in Table 4. More detailed data for the right hand and left hand for the interrupted suturing motions are reported in Figs. 3 and 4. The surgeon's left hand has the Acceleration Peak Rate (APR) as an additional significantly different metric. Note that the total time shown in these tables is calculated using the motion data of the corresponding instruments rather than the clinically measured operation time.

There are no significant differences in the running suturing motions (IV) while multiple considerable differences are found when comparing interrupted suturing motions (III). The significant metrics for both surgeon's hands and the assistant's hand include path length, Log Dimensionless Jerk (LDLJ), Number of Speed Peaks (NSP), and Number of Acceleration Peaks (NAP). For the NSP and NAP metrics, the P-value is increasing with an increasing threshold α . At a higher threshold, the motion metrics are often not deemed normally distributed. Therefore a non-parametrical Wilcoxon-ranksum test instead of a student t-test is used. These metrics are marked in bold in the tables. The SPARC, Smoothness, MAPR, SPR and APR are not showing significant differences.

Discussion

Principal findings

Results indicate significant differences in total operation time, fundoplication time, and dissection time between 3D and 2D fetoscopy, in line with available literature. This suggests that 3D vision makes fundoplication and dissections easier. A significant difference was found for several quantitative motion metrics such as path length, Log dimensionless jerk, number of speed peaks, and number of acceleration peaks when comparing interrupted suturing. No significant differences were detected when running suturing motions were compared. Highly standardized tasks like running sutures may over time become automatic for experienced surgeons, negating immediate benefits of 3D fetoscopy³⁰.

Results in existing knowledge context

Several studies have looked at the differences in skill under 3D and 2D vision between novices or experts. In these, differences in operation time were found for running suturing tasks^{14–17}. In our case, running suturing motions (IV) *does not* show quantitative differences. Motion analysis has been subject to a high degree of variability from study to study. Patel et al.³¹ compared 2D and 3D fetoscopy's impact on spina bifida repair using a low fidelity simulator. Surgical performance was gauged subjectively through skin cutting, dural patch placement, and suturing, using the NASA-TLX questionnaire³². In 16 participants there were no significant differences except for 3D's shorter total operation time. In a study by Wilhelm et al.³³, experienced laparoscopists among 48 study participants, excelled with a 3D endoscope. Operation time and instrument path length improved with 3D, but there was no difference in NASA-TLX scores. Nomura et al.³⁴ compared imaging modes for endoscopic submucosal dissection in animals, finding comparable operation times. 3D reduced time and negatively impacted eyestrain symptoms (excluding blurred vision), as measured on a 100 mm visual analog scale³⁵.

The SPARC, as proposed in¹⁸, is showing a non-significant improvement towards the 3D group. Although the metric is known to be time-independent and consistent such as in the detection of Parkinson's disease¹⁹, it is sensitive to noise and therefore requires the right tuning in order not to be affected too much. We assume that the interrupted suturing sequence that we have been using for the comparison includes too many different motions for this metric to appear decisive in our comparison. We expect that it would be necessary to further segment the four motion sequences into smaller more distinguishable movements to leverage this metric to allow for stronger conclusions. The LDLJ, another advanced and recurring motion metric in literature, does show a noticeable difference, although small. This suggests a detectable motion difference between 3D and 2D vision during interrupted suturing. As for the SPARC, our whole analysis could possibly be improved as well by further splitting the motion data.

The threshold α is slightly affecting the motion metrics NSP, NAP, and their corresponding SPRate and APRate. It appears that the metrics with a higher threshold are not normally distributed and become less suitable

Metric	I: Dissection movements			II: High-force movements			III: Interrupted Suturing			IV: Running suturing		
	RH	LH	Scope	RH	LH	Scope	RH	LH	Scope	RH	LH	Scope
Total time	0.004**	0.523	0.094	0.345	0.114	0.345	<0.001***	<0.001***	<0.001***	0.792	0.849	0.858
Path length	0.019*	0.29	0.077	0.29	0.065	0.245	0.004**	<0.001***	0.016*	0.641	0.613	0.422
Smoothness	0.87	0.786	0.97	0.257	0.999	0.773	0.523	0.667	0.18	0.063	0.086	0.073
SAL	0.234	0.968	0.151	0.983	0.999	0.708	0.247	0.182	0.222	0.222	0.077	0.173
SPARC	0.876	0.346	0.246	0.882	0.545	0.322	0.602	0.497	0.996	0.764	0.054	0.935
LDLJ	0.187	0.238	0.395	0.812	0.506	0.766	0.007**	0.029*	0.011*	0.081	0.076	0.228
Max speed	0.145	0.045*	0.557	0.718	0.25	0.352	0.03*	0.284	0.429	0.247	0.21	0.199
NSP0	0.016*	0.674	0.074	0.125	0.066	0.17	<0.001***	<0.001***	<0.001***	0.46	0.999	0.261
NSP1	0.056	0.86	0.135	0.113	0.046*	0.224	0.003**	0.002**	0.002**	0.976	0.682	0.255
NSP2	0.011*	0.524	0.095	0.127	0.071	0.352	0.005**	<0.001***	0.001**	0.616	0.537	0.515
NSP5	0.023*	0.229	0.063	0.111	0.484	0.24	<0.001***	0.009**	0.004**	0.129	0.555	0.785
NSP10	0.016*	0.754	0.5	0.81	0.429	0.038*	0.013*	0.043*	0.035*	0.061	0.515	0.932
SPRate0	0.095	0.477	0.387	0.005**	0.22	0.119	0.096	0.029*	0.061	0.239	0.529	0.06
SPRate1	0.421	0.396	0.73	0.033*	0.233	0.914	0.183	0.058	0.077	0.972	0.59	0.084
SPRate2	0.151	0.941	0.652	0.048*	0.651	0.968	0.207	0.103	0.502	0.543	0.662	0.409
SPRate5	0.114	0.123	0.159	0.644	0.083	0.778	0.052	0.69	0.328	0.118	0.526	0.792
SPRate10	0.032*	0.743	0.34	0.81	0.786	0.105	0.043*	0.043*	0.238	0.069	0.537	0.896
Mean speed	0.054	0.606	0.759	0.565	0.637	0.826	0.254	0.349	0.447	0.368	0.383	0.662
Speed consist.	0.448	0.058	0.999	0.61	0.571	0.476	0.67	0.247	0.429	0.033*	0.971	0.418
MAPR10	0.159	0.853	0.492	0.725	0.78	0.551	0.03*	0.773	0.617	0.224	0.481	0.201
MAPR25	0.31	0.905	0.345	0.672	0.676	0.441	0.06	0.341	0.76	0.796	0.39	0.126
MAPR50	0.31	0.483	0.239	0.462	0.373	0.995	0.136	0.266	0.926	0.431	0.284	0.126
MAPR75	0.192	0.465	0.421	0.447	0.463	0.803	0.429	0.624	0.11	0.108	0.11	0.247
MAPR90	0.236	0.309	0.565	0.476	0.852	0.707	0.276	0.837	0.931	0.046*	0.093	0.329
Max accel	0.01*	0.583	0.481	0.422	0.088	0.886	0.024*	0.177	0.537	0.128	0.931	0.792
NAP0	0.01*	0.641	0.06	0.17	0.081	0.199	<0.001***	<0.001***	<0.001***	0.546	0.699	0.348
NAP1	0.029*	0.736	0.072	0.174	0.076	0.224	<0.001***	<0.001***	0.002**	0.441	0.561	0.573
NAP2	0.015*	0.429	0.087	0.17	0.102	0.61	0.002**	<0.001***	0.004**	0.921	0.429	0.999
NAP5	0.019*	0.269	0.539	0.417	0.668	0.701	0.003**	0.003**	0.054	0.319	0.36	0.975
NAP10	0.024*	0.177	0.493	0.895	0.21	0.905	0.026*	0.013*	0.17	0.2	0.634	0.429
APRate0	0.095	0.372	0.285	0.005**	0.323	0.081	0.084	0.022*	0.059	0.261	0.479	0.072
APRate1	0.55	0.439	0.28	0.024*	0.036*	0.362	0.429	0.031*	0.107	0.426	0.408	0.471
APRate2	0.461	0.595	0.876	0.082	0.969	0.466	0.213	0.021*	0.926	0.909	0.429	0.883
APRate5	0.171	0.332	0.468	0.999	0.09	0.969	0.337	0.246	0.406	0.357	0.411	0.979
APRate10	0.208	0.209	0.289	0.784	0.786	0.714	0.078	0.467	0.623	0.429	0.656	0.429
Mean accel	0.548	0.905	0.628	0.354	0.163	0.374	0.476	0.591	0.141	0.295	0.329	0.177
Accel consist.	0.378	0.504	0.844	0.171	0.586	0.712	0.291	0.914	0.203	0.208	0.891	0.199
IAV	0.035*	0.419	0.151	0.074	0.024*	0.138	0.018*	0.003**	0.009**	0.388	0.662	0.25
Volume	0.051	0.511	0.117	0.61	0.044*	0.352	0.082	0.016*	0.177	0.104	0.537	0.926
P count	14	1	0	5	4	1	18	20	12	2	0	0

Table 4. P values for unpaired t-tests comparing the motion metrics. $\alpha = 0.05$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, bold: Wilcoxon rank sum test. SAL spectral arc length, SPARC modified spectral arc length, LDLJ log dimensionless jerk, NSPx number of speed peaks with threshold equal to x times the average speed, SPRx speed peak rate, MAPRx movement arrest period ratio, NAPx number of acceleration peaks, APRx acceleration peak rate, IAV integral of acceleration vector.

for comparison. Although this group of metrics is showing very significant differences for both the right and left hand ($P_{NSP0, NAP0} = < 0.001$), due to their nature they are highly dependent on the total motion time. Therefore they do not provide much additional information in addition to the motion time.

Clinical and research implications

Given the progressive nature of spina bifida, ensuring a watertight closure of the dura and fascia layers over the neural placode is critical to prevent amniotic fluid-related damage³⁶. Various closure methods, like suturing, can significantly influence repair quality. Through subtask-level motion analysis, this study directly gauged the impact of stereo fetoscopy on these tasks. This approach could potentially unveil alternative closure techniques for better outcomes or optimize existing methods by leveraging specific imaging technology.

During interrupted suturing motions (III), many significant differences are found. During this surgical step, four interrupted sutures are required in order to perform the fundoplication. In contrary to the running suture technique, a knot is required for every suture. The surgeon has to perform more manipulations with the suture. Moreover, the dynamic range of movements also increases; instead of a continuous motion making the semi-circular suture, the surgeon has to move the fundus to perform the suture at a totally different location. Because the manipulations take place in the full range of the surgical field, it appears that improved depth perception is needed for successful completion. A similar interpretation could be made for the dissection motions (I), where the surgeon has to manipulate a series of soft tissues laying caudal in the rabbit. The surgeon has to extend the tools far in the rabbit to reach the tissues, also requiring more depth perception. Such surgical tasks provide an interesting opportunity to be used as a training or evaluation exercise to study the effects of 3D vision vs. 2D vision on the skill level of a surgeon.

Robot-assisted surgery's growing popularity prompts exploration of its application in spina bifida repair. Many available surgical robots feature console stereo displays. Falk et al.³ evaluated 3D visualization's impact using the Da Vinci telerobotics system. Metrics included time, task errors, and kinematic analysis. Results demonstrated the superiority of 3D over 2D at the same/higher resolution, enabling faster, more precise performance. Given the delicate nature of fetal spina bifida repair, robotic solutions would offer stable instruments and high resolution 3D vision. Supported by this study, this makes them an encouraging consideration for surgical teams worldwide.

Strengths and limitations

In this study, we have underscored the significance of pre-processing raw data and introduced methods for that. Also, this study has undertaken a more detailed analysis of surgical steps by breaking them down into specific actions and maneuvers at sub-task level, subsequently annotating the data accordingly. Additionally, we have conducted a comprehensive exploration of 39 motion metrics extracted from existing literature to determine which metrics are most conducive to providing meaningful comparisons in future investigations.

An important limitation of this research is the low number of simulations performed (sample size $n_{2D} = n_{3D} = 6$) while comparing the impact on just one expert endoscopic-fetal surgeon. This study however was powered to detect differences in operation time (primary outcome) and not in motion metrics. Furthermore, technical challenges, such as sensor breakage, workspace being moved out of range during procedure, recording system crash resulted in a further reduction of the motion tracking data, making the data for dissection (sequence I) and high-force movements (sequence II) unreliable. In case of technical errors in the motion tracking system and loss of signals, the trial was not interrupted in order not to increase operation time and animal suffering.

Interested surgeons may face some challenges in replicating this study due to ethical, logistical and financial hurdles linked to animal trials. To address this concern and further improve the external validity by increasing trial numbers and surgeons, one could adopt high or low fidelity surgical simulators, including for spina bifida repair^{31,37,38}.

Conclusion

In the high-fidelity rabbit model for fetoscopic SBA repair, 3D vision via 3-port access substantially shortens operation time for an expert fetal surgeon during interrupted suturing. A substantial difference was found for several quantitative motion metrics such as path length, Log dimensionless jerk, number of speed peaks, and number of acceleration peaks. No substantial differences were detected when running suturing motions were compared.

Three takeaways with respect to motion analysis for future research are summarised here. Firstly, this research shows that one can draw interesting conclusions from comparing interrupted and running suturing motions. Secondly, the importance of time-independent motion metrics, like SPARC, has been pointed out. It is indeed important to have different motion metrics based on multiple aspects. The SPARC and LDLJ are two metrics that have been consistently performing in other literature under the right conditions which is why the recommendation is made to define the sample size based on this metric. The challenge will be to segment the motion data up to a sufficiently small resolution. Finally, it is extremely important to record the full surgical scene and to provide cues in the video when the motion recording is started or stopped. This provides major help in annotating the motion recordings such that accurate and precise motion segments can be used for comparison.

This research opens the door for follow-up studies in the field of 3D vs 2D motion analyses. The results should be interpreted from a guidance perspective. It is a pilot study that demonstrates the urge and potential for research on this topic and has proposed key learnings to help future research.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 19 February 2023; Accepted: 14 November 2023

Published online: 28 November 2023

References

- Held, R. T. & Hui, T. T. A guide to stereoscopic 3D displays in medicine. *Acad. Radiol.* **18**, 1035–1048 (2011).
- Fergo, C., Burcharth, J., Pommegaard, H.-C., Kildebro, N. & Rosenberg, J. Three-dimensional laparoscopy vs 2-dimensional laparoscopy with high-definition technology for abdominal surgery: A systematic review. *Am. J. Surg.* **213**, 159–170 (2017).
- Falk, V., Mintz, D., Grunenfelder, J., Fann, J. & Burdon, T. Influence of three-dimensional vision on surgical telemanipulator performance. *Surg. Endosc.* **15**, 1282–1288 (2001).

4. Silvestri, M. *et al.* Autostereoscopic three-dimensional viewer evaluation through comparison with conventional interfaces in laparoscopic surgery. *Surg. Innov.* **18**, 223–230 (2011).
5. Sørensen, S. M. D., Savran, M. M., Konge, L. & Bjerrum, F. Three-dimensional versus two-dimensional vision in laparoscopy: A systematic review. *Surg. Endosc.* **30**, 11–23 (2016).
6. Nguyen, D. H., Nguyen, B. H., Van Nong, H. & Tran, T. H. Three-dimensional laparoscopy in urology: Initial experience after 100 cases. *Asian J. Surg.* **42**, 303–306 (2019).
7. Smith, R. *et al.* Effect of passive polarizing three-dimensional displays on surgical performance for experienced laparoscopic surgeons. *J. Br. Surg.* **101**, 1453–1459 (2014).
8. Winn, H. R. *Youmans and Winn Neurological Surgery* (Elsevier Ltd, 2016).
9. Copp, A. J. *et al.* Spina bifida. *Nat. Rev. Dis. Primers* **1**, 1–18 (2015).
10. Farmer, D. L. *et al.* The management of myelomeningocele study: Full cohort 30-month pediatric outcomes. *Am. J. Obstet. Gynecol.* **218**(256), e1–256.e13. <https://doi.org/10.1016/j.ajog.2017.12.001> (2018).
11. Joyeux, L. *et al.* Learning curves of open and endoscopic fetal spina bifida closure: Systematic review and meta-analysis. *Ultrasound Obstet. Gynecol.* **55**, 730–739 (2020).
12. Joyeux, L. *et al.* Validation of a high-fidelity training model for fetoscopic spina bifida surgery. *Sci. Rep.* <https://doi.org/10.1038/s41598-021-85607-6> (2021).
13. Zundel, S., Lehnick, D., Heyne-Pietschmann, M., Trück, M. & Szavay, P. A suggestion on how to compare 2D and 3D laparoscopy: A qualitative analysis of the literature and randomized pilot study. *J. Laparoendosc. Adv. Surg. Tech.* **29**, 114–120 (2019).
14. Cotin, S. *et al.* Metrics for laparoscopic skills trainers: The weakest link! In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 35–43. https://doi.org/10.1007/3-540-45786-0_5 (Springer, 2002).
15. Trejos, A. L., Patel, R. V., Malthaner, R. A. & Schlachta, C. M. Development of force-based metrics for skills assessment in minimally invasive surgery. *Surg. Endosc.* **28**, 2106–2119. <https://doi.org/10.1007/s00464-014-3442-9> (2014).
16. Cavallo, F., Megali, G., Sinigaglia, S., Tonet, O. & Dario, P. A biomechanical analysis of surgeon's gesture in a laparoscopic virtual scenario. *Stud. Health Technol. Inf.* **119**, 79–84 (2006).
17. Fard, M. J. *Computational Modeling Approaches for Task Analysis in Robotic-assisted Surgery*. Ph.D. Thesis (Wayne State University, 2016).
18. Balasubramanian, S., Melendez-Calderon, A., Roby-Brami, A. & Burdet, E. On the analysis of movement smoothness. *J. Neuroeng. Rehabil.* **12**, 1–11. <https://doi.org/10.1186/s12984-015-0090-9> (2015).
19. Beck, Y. *et al.* Sparc: A new approach to quantifying gait smoothness in patients with Parkinson's disease. *J. Neuroeng. Rehabil.* **15**, 1–9. <https://doi.org/10.1186/s12984-018-0398-3> (2018).
20. Gulde, P. & Hermsdörfer, J. Smoothness metrics in complex movement tasks. *Front. Neurol.* **9**, 615. <https://doi.org/10.3389/fneur.2018.00615> (2018).
21. Sakata, S., Watson, M. O., Grove, P. M. & Stevenson, A. R. L. The conflicting evidence of three-dimensional displays in laparoscopy. *Ann. Surg.* **263**, 234–239. <https://doi.org/10.1097/sla.0000000000001504> (2016).
22. Zwart, M. J. *et al.* Added value of 3D-vision during laparoscopic biotissue pancreatico- and hepaticojejunostomy (LAELAPS 3d2d): An international randomized cross-over trial. *HPB* **21**, 1087–1094. <https://doi.org/10.1016/j.hpb.2019.04.012> (2019).
23. Quintero, R. A. *et al.* Hydrolaparoscopy in the rabbit: A fine model for the development of operative fetoscopy. *Am. J. Obstet. Gynecol.* **171**, 1139–1142 (1994).
24. Chitty, L. S., Altman, D. G., Henderson, A. & Campbell, S. Charts of fetal size: 3. Abdominal measurements. *BJOG Int. J. Obstet. Gynaecol.* **101**, 125–131 (1994).
25. Jamieson, G. G. & Britten-Jones, R. Laparoscopic Nissen fundoplication. In *Surgery of the Upper Gastrointestinal Tract*. 571–574 (Springer, 1994).
26. Endoskope, K. S. *Karl Storz Endoscopes*.
27. Perlin, K., Paxia, S. & Kollin, J. S. An autostereoscopic display. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. 319–326 (2000).
28. NC3Rs. *National Centre for the Replacement, Refinement and Reduction of Animals in Research*.
29. Kilkeny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: The arrive guidelines for reporting animal research. *PLoS Biol.* **8**, e1000412 (2010).
30. Chan, A. *et al.* Comparison of two-dimensional vs three-dimensional camera systems in laparoscopic surgery. *Surg. Endosc.* **11**, 438–440 (1997).
31. Patel, S. K. *et al.* Comparison of two- and three-dimensional endoscopic visualization for fetal myelomeningocele repair: A pilot study using a fetoscopic surgical simulator. *Child's Nervous Syst.* **37**, 1613–1621 (2021).
32. Hart, S. G. & Staveland, L. E. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in Psychology*. Vol. 52. 139–183 (Elsevier, 1988).
33. Wilhelm, D. *et al.* Comparative evaluation of HD 2D/3D laparoscopic monitors and benchmarking to a theoretically ideal 3D pseudodisplay: Even well-experienced laparoscopists perform better with 3D. *Surg. Endosc.* **28**, 2387–2397 (2014).
34. Nomura, K. *et al.* Comparison of 3D endoscopy and conventional 2D endoscopy in gastric endoscopic submucosal dissection: An ex vivo animal study. *Surg. Endosc.* **33**, 4164–4170 (2019).
35. Crichton, N. Visual analogue scale (VAS). *J. Clin. Nurs.* **10**, 706–6 (2001).
36. Joyeux, L. *et al.* Neuroprotection is improved by water tightness of fetal spina bifida repair in the sheep model. *Ultrasound Obstet. Gynecol.* **61**, 81–92 (2023).
37. Spoor, J. K. *et al.* Development of a simulator for training of fetoscopic myelomeningocele surgery. *Prenatal Diagn.* **43**, 355–358 (2023).
38. da Costa, M. D. S. *et al.* Training model for the fetal myelomeningocele correction with multiportal endoscopic technique. *Child's Nervous Syst.* 1–6 (2023).

Acknowledgements

M.A. Ahmad is supported by Erasmus Mundus FetalMed PhD Program. This work received funding from the FLEXFET project (3E200384), Wellcome [WT101957; 203148/Z/16/Z] and EPSRC [NS/A000027/1; NS/A000049/1]. T. Vercauteren is supported by a Medtronic / Royal Academy of Engineering Research Chair [RCSR1819/7/34].

Author contributions

L.J. and J.D. performed animal surgeries and designed the quantitative methodology with the help of E.V.P., M.A.A., and Y.W. M.A.A. collected the motion data during the surgeries. Y.W. processed the data, analyzed the results, and wrote the first version of the manuscript. M.A.A. adapted the manuscript to the final version. E.V.P., T.V., E.E., S.O., and J.D. provided guidance and supervision throughout and reviewed the manuscript. All co-authors critically revised and approved the final submitted manuscript.

Competing interests

The authors declare that they have no conflict of interest. All the instrumentation from the company KARL STORZ was part of a research loan (agreement 1705230948). Above mentioned support of KARL STORZ is in no way affiliated with any other service or procurement decisions on the part of the contractual parties.

Additional information

Correspondence and requests for materials should be addressed to M.A.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023