A novel method for assessing visual perception of surgical planes

Christopher M. Schlachta, BSc, MDCM
Syed Ali, MD, MSc
Hammood Ahmed, MD
Roy Eagleson, PhD

This paper was presented at the 2013 Canadian Surgery Forum by S. Ali and was recipient of the CAGS Basic Science Award.

Accepted for publication Jul. 30, 2014

Correspondence to:
C.M. Schlachta
Schulich School of Medicine and Dentistry
Western University
Canadian Surgical Technologies and Advanced Robotics
London Health Sciences Centre
339 Windermere Rd
London ON N6A 5A5
christopher.schlachta@lhsc.on.ca

DOI: 10.1503/cjs.007414

Background: Recognition of tissue planes during surgery appears to be a skill acquired with experience. We conducted a pilot study to test this hypothesis using a novel method for evaluating this skill in a simulated environment.

Methods: Twelve surgeons of varying levels of experience were shown 16 captured images from a mesorectal excision. For each image, they were asked to draw the ideal dissection plane with a stylus on a tablet computer. We used a novel metric for comparing agreement between lines to determine the level of precision observed between junior and senior trainees and consultant surgeons and measure the accuracy of junior and senior trainees compared with consultant surgeons.

Results: We observed significant differences in precision for 9 of 16 images; 7 of these followed the predicted stepwise pattern associated with level of experience. Using consultant surgeons as the reference standard, we observed significant differences in accuracy between senior and junior trainees for 11 images, with senior trainees being more accurate in 10 of them. Only 2 images failed to contribute significant findings to our analysis.

Conclusion: The findings of this pilot evaluation of a novel method for measuring a surgeon's ability to recognize tissue planes in a simulated model show that skill improves with experience. Further evaluation of this method will reveal its utility as an assessment tool and possibly as a training instrument.

Contexte : La reconnaissance des différents plans tissulaires durant la chirurgie semble être une compétence qui s'acquiert avec l'expérience. Nous avons procédé à une étude pilote pour vérifier cette hypothèse à l'aide d'une nouvelle méthode d'évaluation de cette compétence dans un environnement simulé.

Méthodes : Nous avons montré 16 images provenant d'une excision mésorectale à 12 chirurgiens de divers degrés d'expérience. Pour chaque image, ils devaient dessiner le plan de dissection idéal avec un stylet sur une tablette électronique. Nous avons utilisé un nouvel outil de mesure pour comparer la concordance entre les lignes et déterminer ainsi le degré de précision observé entre les résidents juniors et seniors et les chirurgiens consultants et comparer la précision des résidents juniors et seniors à celle des chirurgiens consultants.

Résultats : Nous avons observé des différences significatives quant à la précision pour 9 images sur 16; 7 d'entre elles étaient conformes aux séquences prévues compte tenu du degré d'expérience. En utilisant les chirurgiens consultants comme norme de référence, nous avons observé des différences significatives quant à la précision entre les résidents seniors et juniors pour 11 images, les résidents seniors étant plus précis pour 10 de ces images. Seulement 2 images n'ont pas permis d'alimenter de façon significative notre analyse.

Conclusion : Les résultats de cette évaluation pilote d'une nouvelle méthode de mesure de l'aptitude des chirurgiens à reconnaître les plans tissulaires dans un modèle simulé montrent que les habiletés s'améliorent avec l'expérience. Il faudra approfondir l'examen de cette méthode pour en confirmer l'utilité comme outil d'évaluation et instrument potentiel de formation.

It is considered a fundamental principle of good surgical technique to respect tissue planes during surgery. Tissue planes tend to be avascular; therefore, bleeding can be reduced. In addition, there is growing evidence of improved oncologic outcomes associated with adherence to dissection along tissue planes. This has been demonstrated clearly for rectal cancer resections.[1] There is compelling evidence to suggest this is also true for colon cancer surgery[2] and hepatobiliary surgery.[3]

One of the challenges encountered when teaching trainees to operate within tissue planes is how to facilitate the trainee's recognition of the plane. What is intuitively obvious to the expert surgeon is not always obvious to the trainee. There are likely visual clues that allow the expert to see what the novice does not yet appreciate. Currently it is not clear how this visual skill is learned during the course of clinical apprentice-based training.

Our hypothesis is that if the ability to perceive tissue planes is an acquired skill, then it should be possible to develop a visual test of that skill that discriminates between novice and expert surgeons.

Methods

We captured a series of 16 digital images representing progressive stages of a mesorectal excision from digitally recorded video. The chosen procedure was a da Vinci-assisted (Intuitive Surgical) laparoscopic proctectomy, and we obtained consent from the patient for the recording and use of the images for surgical instruction. The use of a da Vinci-assisted procedure, as opposed to a laparoscopic procedure, was of no particular consequence. We chose this particular case simply because the image quality and exposure of the tissue planes captured in the video were of such high quality that they seemed ideally suited for our pilot study. The images themselves were selected from video and chosen on the basis of clear views unobstructed by surgical instruments; they represented multiple viewpoints of the mesorectal envelope, as seen from the posterior, anterior, left and right dissection planes. Images were labelled minimally for orientation only.

Images were transferred to an iPad 2 (Apple Inc.) and presented in Sketchbook Pro software (Autodesk Inc.). Study participants were able to draw on the presented image using a stylus for capacitive touch screens (Slim Stylus, Targus). Each surgeon's line was saved in a separate layer for later analysis.

We recruited 12 surgeons with varying levels of expertise to view the images on the iPad. Since this was a pilot study and we had no preconception as to what kind of results to expect, we did not perform any sample size or power analysis calculations.

The participating surgeons included 4 consultant surgeons experienced in performing a mesorectal excision, 4 senior trainees and 4 junior trainees. For each image, we instructed the surgeons to draw a free line representing the dissection plane on the image as if the stylus were a scalpel or preferred dissecting instrument. The surgeons were instructed to draw the line as precisely as possible and to extend the line for as long as they felt they could comfortably appreciate the dissection plane. They were permitted to erase and redraw the line, if needed.

Statistical analysis

To compare one drawn line with another, we formulated a distance metric (similar to the Hausdorff measure used in computational geometry[4]) based on the Euclidean distance between evenly spaced points along the arc. Using Matlab (MathWorks), each line was iteratively bisected 3 times ($b = 3$), resulting in 8 segments specified by pairs of Cartesian coordinates ($p = 2^b+1$; 9 pairs for the present study; Fig. 1). For each study group (G), this distance metric between any 2 lines (a, b) for a given image (i) was calculated as the summed distance (d), in pixels, between each of the coordinate pairs (x, y) as follows:

$$d(Gab{:}\,i) = \sum_{j=1}^{p} \sqrt{(Gax_j - Gbx_j)^2 + (Gay_j - Gby_j)^2}$$

Within each study group, there were 6 possible pairs of lines that could be compared for each image. Across the set of images, we calculated the means ± standard deviations of these 6 pairs for each of the 3 study groups.

The mean distance metric is inversely related to group precision (i.e., the greater the mean distance metric, the less agreement or consensus there is within that group with respect to the location of the ideal dissection plane). For each image, we compared this measure of precision for all

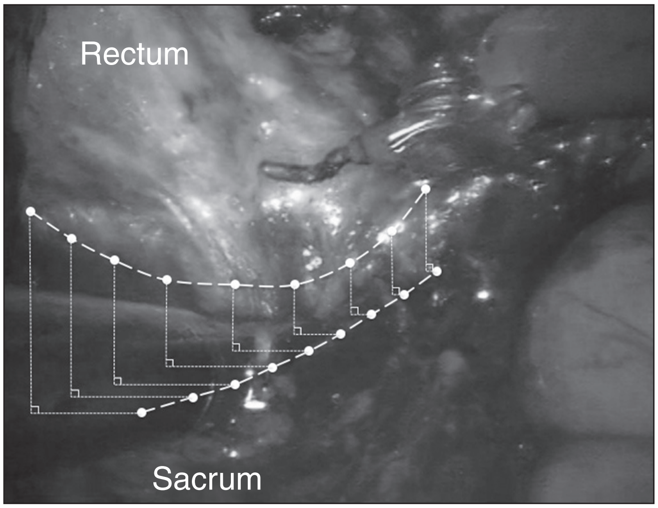

Fig. 1. Distance metric between 2 lines calculated as sum of Euclidean distance between 9 evenly spaced points along each line.

3 study groups using 1-way analysis of variance (ANOVA) with a significance level of $p = 0.05$. When significant differences were found, we performed pairwise comparisons using the Student t test. We used Šidàk's equation to correct the level of significance for multiple comparisons: $\alpha[PT] = 1 - (1 - \alpha[PF])^{1/c}$, where $\alpha[PT]$ is the corrected probability of a type-1 error for an individual test and $\alpha[PF]$ is the probability of a type-1 error for a family of tests. For 3 comparisons and $\alpha[PF] = 0.05$, then $\alpha[PT] = 0.017$.

Since there is no "correct" line representing the tissue plane in each image, and therefore no gold standard with which to compare the accuracy of each line, for the purpose of our analysis we considered the consultant surgeons to be the experts. Accordingly, for each image, we compared each trainee line with each consultant surgeon line in the fashion described above. This produced 16 comparisons (4 × 4) per group per line. The average of these 16 comparisons represents group accuracy, with smaller distances associated with greater accuracy. These were analyzed using the Student t test, with significance set at $p = 0.05$.

Results

For 9 of the 16 images (1–3, 5, 6, 9–11, 16), we observed a stepwise increase in precision according to level of experience from junior trainees to senior trainees to consultant surgeons, although not all were significantly different (Table 1). We observed significant differences in precision in 9 of the 16 images (2, 4–7, 9, 10, 13, 16). Considering these 9 images only, pairwise comparisons revealed significant differences in precision between 10 pairings. In only 1 image (7), significantly greater precision was found in the senior trainee group than in the consultant surgeon group.

In 13 of 16 images, the accuracy of the senior trainee group was greater than that of the junior trainee group (Table 2). These differences were significant for 10 of the images (3–7, 9–11, 14, 16). For image 12, the junior trainees appeared to be significantly more accurate than the senior trainees. We used the results for image 5 as a representative example of these findings. Lines drawn by consultant surgeons (Fig. 2), senior trainees (Fig. 3) and junior trainees (Fig. 4) differed significantly on 1-way ANOVA but not by pairwise comparison (Table 1). For this image, senior trainees were significantly more accurate than junior trainees (Table 2).

Images 8 and 15 contributed no significant findings to any of our analyses.

Discussion

Visual processing of the spatial relations of image properties is known as visual spatial ability. It has been proposed that visual spatial ability can be classified in 5 categories: edge and surface extraction, edge orientation encoding, whole object recognition, mental visualization involving the spatial relations of object parts in 2 dimensions (2D), and mental visualization involving 2D and 3-dimensional (3D) spatial rotations and translations. These represent a hierarchy of ability ranging from low- to high-level ability.[5,6]

Nonverbal, visual–spatial problem solving abilities and the ability to distinguish essential from nonessential detail has been shown to correlate with superior surgical skill in general surgery trainees.[7] This finding was based on correlation between a battery of tests measuring visual–spatial perception, motor sequencing and fine-motor coordination and stress tolerance with a novel rating scale of

Table 1. Mean within-group difference between lines for each surgical image

	Mean distance ± SD (pixels)			1-way ANOVA		Group comparisons ($p < 0.017$)*		
Image	Consultant	Senior	Junior	F	p value	C v. S	C v. J	S v. J
1	1081 ± 327	1496 ± 744	2159 ± 894	3.646	0.05			
2	731 ± 356	1530 ± 480	1973 ± 479	12.158	0.001	Yes	Yes	NS
3	1406 ± 545	2087 ± 760	2917 ± 1888	2.322	0.13			
4	479 ± 336	343 ± 174	1108 ± 373	10.620	0.001	NS	NS	Yes
5	738 ± 429	1468 ± 494	1846 ± 713	6.098	0.012	NS	NS	NS
6	770 ± 318	1161 ± 392	1497 ± 610	3.801	0.046	NS	NS	NS
7	1393 ± 651	693 ± 242	2480 ± 689	15.249	< 0.001	Yes	NS	Yes
8	1619 ± 991	2007 ± 935	1838 ± 691	0.292	0.75			
9	551 ± 237	1050 ± 302	1942 ± 770	12.075	0.001	NS	Yes	NS
10	366 ± 81	801 ± 214	1152 ± 199	30.347	< 0.001	Yes	Yes	Yes
11	1665 ± 974	1869 ± 934	2184 ± 932	0.458	0.64			
12	1511 ± 543	1867 ± 1001	931 ± 370	2.802	0.09			
13	451 ± 102	2066 ± 1130	1257 ± 324	8.430	0.004	NS	Yes	NS
14	1054 ± 371	1733 ± 549	1309 ± 402	3.525	0.06			
15	988 ± 653	2064 ± 1194	1503 ± 641	2.304	0.13			
16	758 ± 274	904 ± 302	1631 ± 631	6.973	0.007	NS	NS	NS

ANOVA = analysis of variance; C = consultant surgeons; J = junior trainees; NS = nonsignificant; S = senior trainees; SD = standard deviation.
*$p = 0.05$, corrected for 3 comparisons.

Table 2. Mean difference from consultant group

Image	Mean distance ± SD (pixels) S v. C	J v. C	*p* value
1	1481 ± 789	1470 ± 988	0.97
2	1051 ± 541	1559 ± 844	0.05
3	1694 ± 917	3183 ± 1365	0.001
4	477 ± 274	1642 ± 667	< 0.001
5	1133 ± 503	1610 ± 526	0.014
6	1019 ± 412	1557 ± 144	< 0.001
7	1245 ± 669	1956 ± 533	0.002
8	1711 ± 1046	1963 ± 712	0.43
9	776 ± 286	1329 ± 327	< 0.001
10	584 ± 248	1271 ± 456	< 0.001
11	1747 ± 782	2605 ± 767	0.004
12	1706 ± 683	731 ± 398	< 0.001
13	1843 ± 1282	1296 ± 462	0.13
14	1384 ± 495	1861 ± 596	0.020
15	1766 ± 1178	2068 ± 780	0.40
16	774 ± 332	1685 ± 843	< 0.001

C = consultant surgeons; J = junior trainees; S = senior trainees; SD = standard deviation.

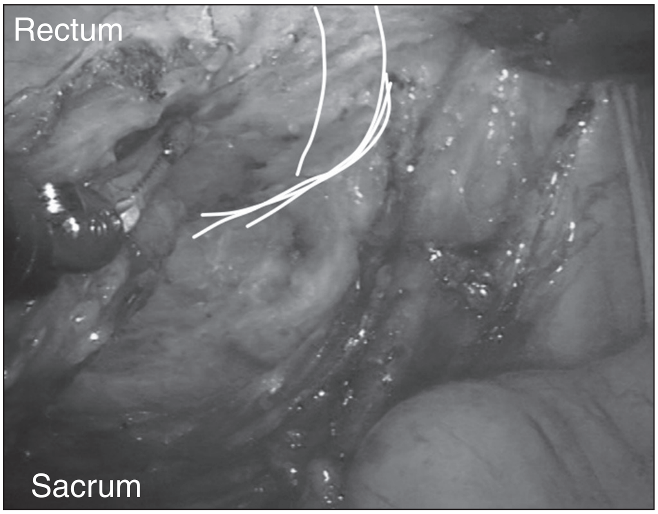

Fig. 2. Consultant surgeons' responses for image 5.

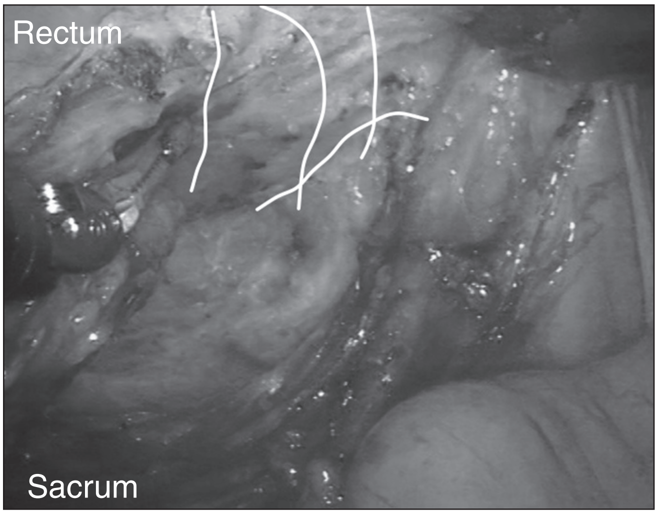

Fig. 3. Senior trainees' responses for image 5.

surgical performance. This rating scale was composed of 12 items assessing mainly technical performance.

Since publication of that report, a growing body of evidence has been produced that demonstrates correlation between performance on various tests of visual–spatial ability and surgical performance.[8] These studies have focused on a range of topics from measures of whole surgical tasks assessed in a laboratory setting to overall assessment of clinical performance. There seems to be stronger correlation between task performance and high-level visual–spatial abilities, such as 3D spatial rotations.

Relevant to the performance of laparoscopic surgery, a unique method for gauging the perceived depth relations of objects presented in a 2D display provided some support for the claim that visual perceptual skills necessary for understanding 3D structure can be improved with practice.[9] However, some neurophysiological studies[10,11] have suggested that in specific perceptual learning, the benefits of perceptual training would be relatively limited.

In the present study, we focused on the initial perception and spatial reasoning skills needed to initiate performing a particular surgical procedure. Based on the presumption that in order to dissect a tissue plane one must first be able to visualize that plane, we developed a novel technique for assessing the ability of surgeons to identify dissection planes on static images acquired from actual surgical procedures. It was the goal of this pilot study to evaluate the tool by determining whether there are measurable performance differences between surgeons and trainees.

Despite our small number of participants, we found significant differences consistent with our a priori hypothesis that recognition of surgical planes is a skill acquired through experience. Considering the within-group variability as an inverse measure of concordance or group precision, we found that we were able to demonstrate a stepwise increase in precision from junior trainees to consultant surgeons.

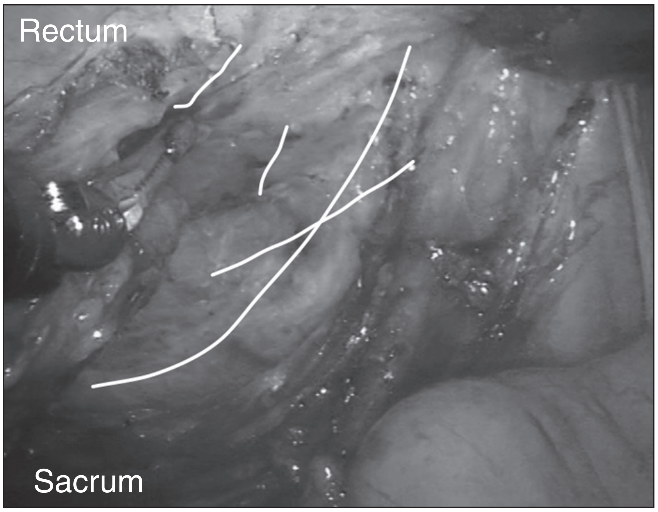

Fig. 4. Junior trainees' responses for image 5.

Limitations

One of the limitations of our methodology is the unavailability of a "correct" plane to use as a reference standard. Therefore, in order to assess accuracy we chose to consider the expert (consultant surgeons) group as the reference standard against which we compared trainee groups. We argue that in many instances the perfect plane of dissection may be a matter of conjecture. Using the experts as a control group, we preserved the variability inherent in multiple opinions. We anticipate that a larger number of experts would narrow this variability or improve the precision of the expert group. Interestingly, for this pilot study our small sample size was still sufficient to demonstrate that senior trainees were predominantly more accurate than junior trainees, which was consistent with our hypothesis.

One of the potential confounders of this study is content validity. It is possible that rather than testing differences in ability to perceive tissue planes, we are simply testing differences in trainees' familiarity with this particular surgical procedure based on their experience and recall. Some of the lines drawn by junior trainees were not only away from the ideal plane, but also clearly demonstrated a lack of recognition of either the anatomy or the next step in the procedure. Future study will need to include an assessment of procedure-specific experience to determine whether this represents an overriding factor.

Another potential weakness is the small number of participants in this pilot study. Despite the ability to appreciate significant differences among the study groups for many of the images, the response of a single individual has the potential to have a large effect on the group. This will be addressed in future studies with larger study groups.

It would be a valid criticism to argue that in practice many factors contribute to the dynamic recognition of surgical planes. The purpose of this pilot was simply to determine if our methodology would be sensitive enough to discriminate between novices and experts based on static images alone. Given that we demonstrated the ability of this methodology to detect expected differences among study groups, we now have the possibility in future endeavours to study the effect of additional clues. This is an area ripe for further investigation, including the use of lead-in video to static images, the contribution of innate psychomotor ability, the difference in visual clues obtained from open compared with laparoscopic surgery or even the role of narrow band imaging. Some of these areas are already being pursued at our institution.

Not all of the images used in this study contributed to significant findings, which may be a result of type-II error owing to the small sample size. It is also possible that some images were less discriminating than others. Future study will require evaluating a larger library of images so that a better understanding of what represents a discriminating image can be developed.

Conclusion

Our pilot study demonstrated a novel method of evaluating visual perception and spatial reasoning of surgical tissue planes. This method shows promise as a tool that possesses the sensitivity to discriminate between levels of experience of surgical trainees. Many questions remain to be addressed through future study and work already in progress. With further development, this methodology has the potential to be developed into both an assessment tool and a training tool to enhance surgical instruction.

Acknowledgements: This research was supported in part by a grant from the Ontario Research Fund, Research Excellence program.

Affiliations: From the Canadian Surgical Technologies and Advanced Robotics, London Health Sciences Centre (Schlachta, Ali, Ahmed, Eagleson); the Department of Surgery, Schulich School of Medicine and Dentistry, Western University (Schlachta, Ali, Ahmed); and the Faculty of Engineering, Western University (Eagleson), London, Ont.

Competing interests: None declared.

Contributors: C. Schlachta and R. Eagleson designed the study. C. Schlachta acquired the data, which all authors analyzed. C. Schlachta and R. Eagleson wrote the article, which all authors reviewed and approved for publication.

References

1. Wibe A, Moller B, Norstein J, et al. A national strategic change in treatment policy for rectal cancer — implementation of total mesorectal excision as routine treatment in Norway. A national audit. *Dis Colon Rectum* 2002;45:857-66.
2. West NP, Morris EJA, Rotina O, et al. Pathology grading of colon cancer surgical resection and its association with survival: a retrospective observational study. *Lancet Oncol* 2008;9:857-65.
3. Imamura H, Seyama Y, Kokudo N, et al. One thousand fifty-six hepatectomies without mortality in 8 years. *Arch Surg* 2003;138:1198-206.
4. Hausdorff F. Dimension und äusseres mass [article in German]. *Math Ann* 1918;79:157-79.
5. Marr D. *Vision: a computational investigation into the human representation and processing of visual information*. New York (NY): Freeman; 1982.
6. Anastakis DJ, Hamstra SJ, Matsumoto ED. Visual–spatial abilities in surgical training. *Am J Surg* 2000;179:469-71.
7. Schueneman AL, Pickleman J, Hesslein R, et al. Neuropsychologic predictors of operative skill among general surgery residents. *Surgery* 1984;96:288-95.
8. Maan ZN, Maan IN, Darzi AW, et al. Systematic review of predictors of surgical performance. *Br J Surg* 2012;99:1610-21.
9. Sidhu RS, Tompa D, Jang RW, et al. Interpretation of three-dimensional structure from two-dimensional endovascular images: implications for educators in vascular surgery. *J Vasc Surg* 2004;39:1305-11.
10. Fahle M. Specificity of learning curvature, orientation, and vernier discriminations. *Vision Res* 1997;37:1885-95.
11. Yotsumoto Y, Watanabe T, Sasaki Y. Different dynamics of performance and brain activation in the time course of perceptual learning. *Neuron* 2008;57:827-33.