Journal of Digital Imaging

Computer Input Devices: Neutral Party or Source of Significant Error in Manual Lesion Segmentation?

James Y. Chen,[1,2,3] F. Jacob Seagull,[2] Paul Nagy,[2] Paras Lakhani,[1,2] Elias R. Melhem,[1] Eliot L. Siegel,[2,4] and Nabile M. Safdar[2]

Lesion segmentation involves outlining the contour of an abnormality on an image to distinguish boundaries between normal and abnormal tissue and is essential to track malignant and benign disease in medical imaging for clinical, research, and treatment purposes. A laser optical mouse and a graphics tablet were used by radiologists to segment 12 simulated reference lesions per subject in two groups (one group comprised three lesion morphologies in two sizes, one for each input device for each device two sets of six, composed of three morphologies in two sizes each). Time for segmentation was recorded. Subjects completed an opinion survey following segmentation. Error in contour segmentation was calculated using root mean square error. Error in area of segmentation was calculated compared to the reference lesion. 11 radiologists segmented a total of 132 simulated lesions. Overall error in contour segmentation was less with the graphics tablet than with the mouse ($P<0.0001$). Error in area of segmentation was not significantly different between the tablet and the mouse ($P = 0.62$). Time for segmentation was less with the tablet than the mouse ($P = 0.011$). All subjects preferred the graphics tablet for future segmentation ($P = 0.011$) and felt subjectively that the tablet was faster, easier, and more accurate ($P = 0.0005$). For purposes in which accuracy in contour of lesion segmentation is of the greater importance, the graphics tablet is superior to the mouse in accuracy with a small speed benefit. For purposes in which accuracy of area of lesion segmentation is of greater importance, the graphics tablet and mouse are equally accurate.

KEY WORDS: Image segmentation, user-computer interface, computer assisted detection, computer hardware, data collection, human computer interaction, evaluation research, segmentation

INTRODUCTION

Lesion segmentation involves outlining the contour of an abnormality on an image to distinguish boundaries between normal and abnormal tissue. Segmentation of lesion volume (or area on 2D images slices) is essential to track malignant and benign disease in medical imaging for clinical, research, and treatment purposes.

Clinically, large trials such as the ACRIN National Lung Screening Trial as well as individual clinical cases currently rely on accurate and repeatable methods of lesion measurement and segmentation.[1–5] Automated tumor segmentation for radiation therapy can improve treatment planning accuracy[6] and highly localized radiotherapies, such as proton beam,[7] require accurate pretreatment targeting.[8,9]

For research, segmentation can be used for objective comparison of new imaging sequences and modalities or for the creation of automated segmentation tools.[10–13] The validation of automated segmentation tools commonly relies on testing for lesion contour and size against manual segmentation.[14–19] Manual lesion segmentation, however, may vary in accuracy, depending on the input method used for measurement. To our

[1]*From the University of Pennsylvania, Philadelphia, USA.*
[2]*From the University of Maryland, College Park, USA.*
[3]*From the San Diego Veterans Administration Hospital, University of California San Diego, 3350 La Jolla Village Dr, San Diego, CA, 92161, USA.*
[4]*From the Baltimore Veterans Administration Hospital, Baltimore, USA.*

Correspondence to: James Y. Chen, San Diego Veterans Administration Hospital, University of California San Diego, 3350 La Jolla Village Dr, San Diego, CA, 92161, USA; e-mail: James_Chen@alumni.brown.edu

Copyright © 2009 by Society for Imaging Informatics in Medicine

Online publication 5 January 2010
doi: 10.1007/s10278-009-9258-9

knowledge, nearly all clinical and research users use the mouse as the input device for manual lesion segmentation.

The QWERTY keyboard and mouse are the de facto standard configuration in computer input devices. Although these work well for standard user interface interactions, many graphic designers have chosen to replace or augment these devices with graphics tablets that more closely mimic conventional pen and paper. Applications that require manual lesion segmentation bear similarities to the tasks of graphic designers, chiefly a need for finer motor control. Most people performing segmentation, however, still rely mainly on the mouse interface. The mouse interface can be variably accurate and used with either finer finger and wrist movements or larger arm and wrist movements; whereas, the accuracy of pen strokes is typically limited to fine movements such as wrist flexion.

To our knowledge, at the time of the writing of this manuscript, there has been a single published medical study comparing the accuracy of the computer mouse with alternative input devices including the graphics tablet and touch-sensitive screen.[20]

We hypothesize that the pen-and-tablet interface should be empirically and subjectively at least as accurate, easy, and fast as an optical mouse for lesion segmentation in regards to both lesion contour and size.

MATERIALS AND METHODS

This is an IRB-approved prospective study comparing two different computer input devices for manual segmentation of simulated lesions, including a post-experiment survey of all participants.

Simulated reference lesions were created and then combined into one set of two groups of six images. One group comprised of three lesion morphologies in two sizes, one for each input device (Fig. 1). Lesion contours were created to simulate clinical lesions, including ovoid, lobulated, and spiculated forms. To account for differences in lesion size/zooming, three lesion shapes were resized 50% in each dimension using bicubic interpolation to create a second group of three smaller size lesions. Each image contained only a single lesion with high-contrast, black-on-white backgrounds, and hard edges to minimize the cognitive task and time required to identify the lesion and its borders.

A wireless, 800 dpi, laser mouse (Logitech International S.A.; Switzerland) using the default software driver (Microsoft Corporation, version 5.1.2600.5512) was set to default movement parameters including variable gain/acceleration. The batteries were used for less than 1 week during the experiment to minimize performance degradation from battery exhaustion, and the same non-reflective mouse surface was used for all subjects. The graphics tablet (Wacom Corporation; Japan) was also set to default parameters within

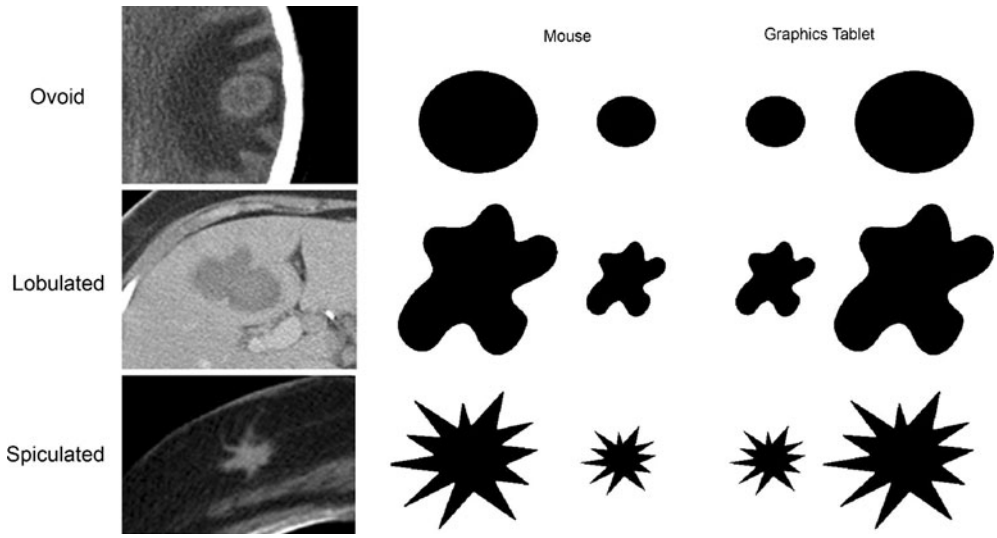

Fig 1. Three lesion shapes (*left*) were used as templates to create high-contrast, hard-edged images in two sizes each. Study participants segmented one set of six images with the mouse and the other set of six images with the graphics tablet.

the driver software (Wacom Corporation, version 6.00-5) and matched the aspect ratio of the viewing screen. Subjects were encouraged to position both devices and display for their comfort.

Radiologists performed manual lesion segmentation with each input device for each group of six images (12 lesion segmentations per subject) using commercially available image editing software (Adobe Systems, USA). Simulated lesions and segmentation device order were randomized for each subject. Participants were allowed a single image on which to practice segmentation up to two times with each input device prior to segmenting each experimental set. Segmentation time was recorded manually for each individual lesion by a single observer throughout the study.

To evaluate error in contour of segmented lesions, undersegmented and oversegmented areas were combined to assess the mis-sampled area. Undersegmented areas were defined as areas of reference lesion that was not included in manual segmentation (false exclusion). Oversegmented areas were defined as areas of segmented lesion that did not correlate to the reference lesion (false inclusion). The segmented lesions were subtracted from the reference lesion to obtain areas of undersegmentation; the reference lesion was then subtracted from the segmented lesion to obtain the area of oversegmentation. The root mean square error was then calculated from the undersegmented and oversegmented areas. Statistical differences were compared with a Wilcoxon test.

To evaluate error in area of segmented lesions, total area of segmentation was recorded for each lesion and the difference in area from the reference lesion was calculated. These differences were compared with a Wilcoxon test.

After segmentation of both lesion sets, each individual completed a force-choice survey between the graphics tablet and mouse for perceived ease, speed, and accuracy of segmentation; and device preference for segmenting lesions in the future. Statistical differences in preferences were compared with a binomial test. Subject demographics were recorded during the survey including number of years of experience with a mouse or tablet.

RESULTS

Eleven radiologists (nine male, two female) participated in the study, segmenting a total of 132 lesions.

Average lesion area was 16,041 pixels with large lesions averaging 22,203 pixels, and small lesions averaging 5,508 pixels. Ovoid lesions averaged 12,562 pixels, lobulated lesions 18,569 pixels, and spiculated lesions 10,437 pixels.

Contour segmentation with the tablet was more accurate than with a mouse (Table 1) with average RMS error of 690 (standard deviation of 530, overall RMS of 4.3% of average lesion area) versus 992 (standard deviation of 1,033, overall RMS of 6.2% of average lesion area, an increase of 44% in error; $P<0.0001$). For the large lesions, the tablet demonstrated significantly less error (RMS error of 1,003, standard deviation of 617, overall RMS of 4.5% of large lesion area) compared to the mouse (RMS error of 1,489, standard deviation of 1,288, overall RMS of 6.7% of average large lesion area, corresponding to a 48% increase in error; $P<0.0001$). For small lesions, the tablet also demonstrated significantly less error (RMS error of 377, standard deviation of 246, overall RMS of 6.8% of small lesion area) compared to the mouse (RMS error of 496, standard deviation of 390, overall RMS of 9.0% of small lesion area, corresponding to a 32% increase in error; $P=0.0121$). Of the 11 partic-

Table 1. Root Mean Square Error in Contour Error as a Percent of Reference Lesion Area

Contour segmentation error between graphics tablet and mouse

Lesion set	Tablet	Mouse	Increase from tablet	*P*
All lesions	4.3%	6.2%	44%	<0.0001
Large lesions	4.5%	6.7%	48%	<0.0001
Small lesions	6.8%	9.0%	32%	0.0121
Ovoid lesions	2.9%	4.6%	59%	0.0008
Lobulated lesions	3.8%	5.1%	33%	0.0127
Spiculated lesions	9.3%	14%	46%	0.0156

There is significantly greater contour error with a mouse across all lesions and lesion subgroups

ipants, ten (91%) were more accurate with the tablet compared to one (9%) with the mouse.

When lesions were subgrouped into ovoid, lobulated, and spiculated morphologies, the tablet again demonstrated less contour error than the mouse, except with the lobulated lesions. For lobulated lesions, the tablet demonstrated significantly less error (P=0.0127). Tablet error (RMS of 714, standard deviation of 498, overall RMS 3.8% of lobulated lesion area) compared to the mouse (RMS of 950, standard deviation of 898, overall RMS of 5.1% of lobulated lesion area, corresponding to 33% increase in error of 33%). With ovoid lesions, the tablet demonstrated significantly less error (RMS of 364, standard deviation of 294, overall RMS 2.9% of ovoid lesion area) compared to the mouse (RMS of 580, standard deviation of 574, overall RMS 4.6% of ovoid lesion area, corresponding to a 59% increase in error; P=0.0008). With spiculated lesions, the tablet demonstrated significantly less error (RMS of 975, standard deviation of 682, overall 9.3% of spiculated lesion area) compared to the mouse (RMS of 1,424, standard deviation of 1,431, overall 13.6% of spiculated lesion area corresponding to a 46% increase in error; P=0.0156).

Error in pixel area of segmentation with the tablet and mouse was not significantly different (Table 2). Across all lesions, the tablet demonstrated a mean error in area of 678 pixels (4.9% of lesion area and standard deviation of 920 pixels) compared to 617 pixels (4.5% of lesion area and a standard deviation of 1,119 pixels) for the mouse (P=0.6194). For large lesions, the tablet demonstrated a mean error in area of 1,052 pixels (4.7% of lesion area and standard deviation of 1,114 pixels) compared to 954 pixels for the mouse (4.3% of lesion area and standard deviation of 1,402 pixels; P=0.4658). For small lesions, the tablet demonstrated a mean error in area of 318 pixels (5.8% of lesion area and standard deviation of 458 pixels) compared to 322 pixels for the mouse (5.8% of lesion area and standard deviation of 628 pixels; P=0.8442). For ovoid lesions, the tablet demonstrated a mean error in area of 364 pixels (2.9% of lesion area and standard deviation of 555 pixels), compared to −17 pixels for the mouse (−0.1% of lesion area and standard deviation of 998 pixels; P=0.2479). For lobulated lesions, the tablet demonstrated a mean error in area of 362 pixels (1.9% of lesion area and standard deviation of 875 pixels) compared to 499 pixels for the mouse (2.7% of lesion area and standard deviation of 593 pixels; P=0.8486). For spiculated lesions, the tablet demonstrated a mean error in area of 1,309 pixels (12.5% of lesion area and standard deviation of 958 pixels) compared to 1,421 pixels for the mouse (13.6% of lesion area and standard deviation of 1,184 pixels; P=0.8382). Of the 11 participants, seven were more overall more accurate with the tablet (64%) than with the four with the mouse (36%).

Table 2. Pixel Area Error as a Percent of Reference Lesion Pixel Area

Pixel area segmentation error between graphics tablet and mouse			
Lesion set	Tablet	Mouse	*P*
All lesions	4.9%	4.5%	0.619
Large	4.7%	4.3%	0.466
Small	5.8%	5.8%	0.844
Ovoid	2.9%	−0.1%	0.248
Lobulated	1.9%	2.7%	0.849
Spiculated	13%	14%	0.838

There is no significant difference in pixel area error between the graphics tablet and mouse

Overall lesion segmentation time was statistically significantly less with the tablet, averaging 27 s versus 29 s with the mouse (P=0.031, R^2= 0.96). For the small lesions, the tablet (23 s) was significantly faster than the mouse (25 s; P= 0.011). There was no significant difference between the tablet and mouse when other subsets were analyzed. For the large lesions, there was no significant difference between the tablet (32 s) and mouse (34 s; P=0.24). There was no difference in segmentation time for ovoid lesions between the tablet (15 s) and mouse (16 S; P=0.41), for lobulated lesions between the tablet (24 s) and mouse (27 s; P=0.22), or for spiculated lesions (42 s) versus (45 s) (P=0.13).

All radiologists reported more experience using a mouse than a tablet, with 100% having >5 years of mouse experience. Only three subjects (27%) had previously used a graphics tablet, two (18%) reported up to 1 year of experience, and one (9%) reported 3–5 years of experience.

Nine (82%) reported the perception that segmentation with tablet was faster, while one (9%) perceived the mouse to be faster, and one (9%) abstained from the question (P=0.011).

All (100%) participants reported segmentation with a mouse to be more difficult as well as less accurate. All (100%) participants also reported a

preference for using a graphic tablet in the future when segmenting lesions (P=0.0005).

DISCUSSION

A significant decrease in contour error was seen when manually segmenting lesions with a graphics tablet compared with the same task performed with a mouse, but no significant difference in pixel area error. Despite a general lack of experience with a graphics tablet, there was a small, but statistically significant decrease in time to segment lesions between the graphics tablet and mouse. All operators reported the graphics tablet to be subjectively more accurate and easier, and all preferred the tablet for future task performance. This suggests that with the same amount or decreased time expenditure, lesions can be segmented more accurately in contour or equally accurately in area with less effort than with a mouse, a finding that could have significant implications for research or clinical work involving large numbers of lesions. These results mirror findings in the computing literature that the tablet input device is more accurate for precision movements.[21,22]

While the contour segmentation error as a percentage of total lesion area was relatively small, ranging from an average of 5–14% for the mouse and 2.9–9% for the tablet, this was for a single image in 2D space. Actual pathology occurs in 3D space and accurate assessment of lesion volume would require segmenting a stack of 2D images and adding the areas to calculate 3D volume. Even these small errors accumulated over a stack of images can potentially result in significant cumulative error. In clinical trials where relatively small differences are expected of treatment efficacy, even these small amounts of error can have potentially large effects on study data and results. For highly targeted radiotherapeutic treatments, small errors in targeting can result in under-treatment of a lesion or damage to non-pathologic structures adjacent to the intended target.

When error in the total area of segmentation was compared, there was no statistically significant difference between the graphics tablet and the mouse. For purposes in which total lesion size is the subject of interest over lesion contour, such as in clinical evaluation of lesion size for treatment response or dosing of medication, the mouse and tablet can be considered equivalent.

The perceived increased ease of lesion segmentation may magnify differences between the mouse and graphics tablet when large volumes of lesions are segmented within and across patients. One set of investigators have found that a pen tablet system creates less overall muscular load than a mouse,[23] while another found improved productivity in general cursor control.[24] Operator fatigue and frustration with the input method during segmentation can lead to less accurate segmentation when the task is performed repeatedly. If segmentation with the tablet is truly easier, the graphics tablet could potentially decrease the amount of and effects of fatigue. This could result in creating more accurate datasets with large numbers of segmented lesions against which to evaluate automated methods of segmentation and new imaging sequences or modalities as they are developed.

We speculate that additional experience with the tablet beyond that held by the participants in this study might increase the speed of lesion segmentation but only minimally affect accuracy—an area that remains for future evaluation.

LIMITATIONS

Manual timing of lesion segmentation time introduces the possibility that the differences in segmentation time were at least partially, if not, wholly related to variability in relying on a human observer. As a single human observer performed timing measurements for all data points, variability was limited to that single observer instead of across multiple observers. Another study found that the tablet is faster than the mouse for accurate clinical contouring.[25]

The study design was not conducive to a meaningful analysis of intra-observer variability, as no participant performed the same task more than once. Any analysis of intraobserver variability would therefore be confounded by additional factors, such as size and/or shape of the segmented lesion, and the device used. This design was chosen to limit the total time required by the participants.

Study participants were not allowed to correct their lesion segmentations post hoc. In clinical use, radiologists have the opportunity to correct their

lesions segmentations regardless of the input device used, but the frequency and total amount of correction are not known. An input device which introduces greater error would also increase the time and labor burden of correcting the segmentation potentially decreasing the incentive for error correction. This could lead to the creation of segmented data sets with greater total error.

Our generalizability to true clinical images and lesion segmentation is currently unknown, but is a potential area for future investigation. Lesions seen in clinical imaging studies can be less conspicuous and may have less well-defined margins. This can complicate the cognitive task of defining a lesion's presence and borders, but we expect that this would affect the use of both the graphics tablet and mouse equally.

Two of our subjects expressed a dislike for the specific mouse and mouse-surface used in this study. The mouse and surface chosen, however, were popularly available, commercial products with conventional designs, which are expected to be similar to the default devices provided for clinical use. It is unknown whether this is related to the specific devices or related to a preference of different driver settings. In this study, default settings were chosen, as these were felt to be the most commonly used in clinical workstations. The effect of the specific mouse, mouse surface, and driver settings may be a topic for future investigation.

CONCLUSION

The choice of input device for manual lesion segmentation was found to significantly affect accuracy of segmentation of lesion contour, but not total lesion size. These findings suggest that for specific purposes in which lesion contour is of greater importance, such as in comparing different imaging sequences and modalities or for targeting highly selective radiotherapy, there is significantly greater error with a mouse than with a graphics tablet at a slight speed penalty. In purposes in which total lesion size is of greater importance, the mouse and tablet are equally accurate for lesion segmentation.

REFERENCES

1. Saini S. Radiologic measurement of tumor size in clinical trials. AJR 176:333–334, 2001

2. Gavrielides MA, Kinnard LM, Myers KJ, et al. Noncalcified lung nodules: volumetric assessment with thoracic CT. Radiology 251:26–37, 2009

3. Hopper KD, Kasales CJ, Van Slyke MA, et al. Analysis of interobserver and intraobserver variability in CT tumor measurements. AJR 167:851–854, 1996

4. Schwartz LH, Ginsberg MS, DeCorato D, et al. Evaluation of tumor measurement in oncology: use of film-based and electronic techniques. J Clin Oncol 18:2179–2184, 2000

5. Zijdenbos AP, Forghani R, Evans AC. Automatic "pipeline" analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. IEEE Trans Med Imaging 21:1280–1291, 2002

6. Yu H, Caldwell C, Mah K, et al. Automated radiation targeting in head-and-neck cancer using region-based texture analysis of PET and CT images. Int J Radiat Oncol Biol Phys 75(2):618–625, 2009

7. Suit H. The Gray Lecture 2001: coming technical advances in radiation oncology. Int J Radiat Oncol Biol Phys 53(4):798–809, 2002

8. De Xivry JO, Janssens G, Bosmans G, et al. Tumor delineation and cumulative dose computation in radiotherapy based on deformable registration of respiratory correlated CT images of lung cancer patients. Radiother Oncol 85:232–238, 2007

9. Hamilton CS, Ebert MA. Volumetric uncertainty in radiotherapy. Clin Oncol 17:456–464, 2005

10. Cao L, Li X, Zhan J, Chen W. Automated lung segmentation algorithm for CAD system of thoracic CT. Journal of Medical Colleges of PLA 23:215–222, 2008

11. Lao Z, Shen D, Liu D, et al. Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machine. Acad Radiol 15:300–313, 2008

12. Kostopoulos S, Glotsos D, Kagadis GC, et al. A hybrid pixel-based classification method for blood vessel segmentation and aneurysm detection on CTA. Comput Graph 31:493–500, 2007

13. Goodman LR, Gulsun M, Washington L, et al. Inherent variability of CT lung nodule measurements in vivo using semi-automated volumetric measurements. AJR 186:989–994, 1996

14. Alfano B, Brunetti A, Larabina M, et al. Automated segmentation and measurement of global white matter lesion volume in patients with multiple sclerosis. J Magn Reson Imaging 12:799–807, 2000

15. Vannier MW, Pilgram TK, Speidel CM, et al. Validation of magnetic resonance imaging (MRI) multispectral tissue classification. Comput Med Imaging graph 15:217–223, 1991

16. Zijdenbos AP, Dawant BM, Margolin RA. Measurement reliability and reproducibility in manual and semi-automatic MRI segmentation. Proceedings of the 15th IEEE-Engineering in Medicine and Biology Society 15:162–163, 1993

17. Street E, Hadjiiski L, Sahiner B, et al. Automated volume analysis of head and neck lesions on CT scans using 3D level set segmentation. Med Phys 24:4399–4408, 2007

18. Anbeek P, Vincken KL, van Bochove GS, et al. Probabilistic segmentation of brain tissue in MR imaging. Neuroimage 27:795–804, 2005

19. Solloway S, Hutchinson CE, Waterton JC, et al. The use of active shape models for making thickness measurements of articular cartilage from MR images. Magn Reson Med 37:943–952, 1997

20. McWalter EJ, Wirth W, Siebert M, et al. Use of novel interactive input devices for segmentation of articular cartilage from magnetic resonance images. Osteoarthr Cartil 13:48–53, 2005

21. Kravits JF. Of mice and pen: effects of input device on different age groups performing goal-oriented tasks. Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting 2007: 45–49

22. MacKenzie IS, Sellen A, Buxton W. A comparison of input devices in elemental pointing and dragging tasks. Proceedings of ACM CHI'91 Conference on Human Factors in Computing Systems 161–166, 1991

23. Kotani K, Horii K. An analysis of muscular load and performance in using a pen tablet system. J Physiol Anthropol Appl Hum Sci 22:89–95, 2003

24. Bertuca DJ. Letting go of the mouse: using alternative computer input devices to improve productivity and reduce injury. OCLC Syst Serv 17(2):79–83, 2001

25. Larsson SN, Stapleton S, Larsson L. A comparison of speed and accuracy of contouring using mouse versus graphics tablet. Clin Oncol 19:S36, 2007