Compare Display Schemes for Lung Nodule CT Screening

Xiao Hui Wang,¹ Janet E. Durick,² Amy Lu,³ David L. Herbert,² Carl R. Fuhrman,² Joan M. Lacomis,² Cynthia A. Britton,² Diane C. Strollo,² Sherry S. Shang,⁴ Saraswathi K. Golla,² and Walter F. Good¹

This study investigated the relative efficiencies of a stereographic display and two monoscopic display schemes for detecting lung nodules in chest computed tomography (CT). The ultimate goal was to determine whether stereoscopic display provides advantages for visualization and interpretation of three-dimensional (3D) medical image datasets. A retrospective study that compared lung nodule detection performances achieved using three different schemes for displaying 3D CT data was conducted. The display modes included slice-by-slice, orthogonal maximum intensity projection (MIP), and stereoscopic display. One hundred lung-cancer screening CT examinations containing 647 nodules were interpreted by eight radiologists, in each of the display modes. Reading times and displayed slab thickness versus time were recorded, as well as the probability, location, and size for each detected nodule. Nodule detection performance was analyzed using the receiver operating characteristic method. The stereo display mode provided higher detection performance with a shorter interpretation time, as compared to the other display modes tested in the study, although the difference was not statistically significant. The analysis also showed that there was no difference in the patterns of displayed slab thickness versus time between the stereo and MIP display modes. Most radiologists preferred reading the 3D data at a slab thickness that corresponded to five CT slices. Our results indicate that stereo display has the potential to improve radiologists' performance for detecting lung nodules in CT datasets. The experience gained in conducting the study also strongly suggests that further benefits can be achieved through providing readers with additional functionality.

KEY WORDS: Volume visualization, lung neoplasm, image display, image interpretation

INTRODUCTION

Lung cancer affects more than 100,000 Americans each year in all ages but particularly in the smoking population more than 50 years of age.¹⁻³ Most types of lung cancer can be effec-

tively treated if they are detected early.⁴ Radiographic imaging plays an important role in the early detection and diagnosis of lung cancer. The primary radiographic imaging tool for lung nodule screening and early detection has shifted from film-based projection radiography to computed tomography (CT).

Although CT generates volumetric datasets that avoid artifacts due to tissue superposition, it also requires that radiologists interpret a much larger volume of image data. A CT lung examination, for example, often contains more than 200 slices, which is in contrast to the two projection radiographs that were traditionally acquired for a lung study. However, despite the increased effort required, the acquisition of volumetric datasets offers many opportunities for improving performance over that achieved on projection images. However, radiologists need to continue to adopt new reading strategies to exploit the potential

¹From the Image Research Group, University of Pittsburgh, School of Medicine, 3362 Fifth Ave, Pittsburgh, PA 15213, USA.

²From the UPMC Radiology, University of Pittsburgh, School of Medicine, 200 Lothrop St, Pittsburgh, PA 15213, USA.

³From the UPMC Radiology, University of Pittsburgh, School of Medicine, 300 Halket St, Pittsburgh, PA 15213, USA.

⁴From the Department of Radiology, Hahnemann University Hospital, Drexel University College of Medicine, Philadelphia, PA 19102, USA.

Correspondence to: Xiao Hui Wang, Image Research Group, University of Pittsburgh, School of Medicine, 3362 Fifth Ave, Pittsburgh, PA 15213, USA; tel: +1-412-6412561; fax: +1-412-6412582; e-mail: xwang@mail.magee.edu

Copyright © 2010 by Society for Imaging Informatics in Medicine

Online publication 13 April 2010

doi: 10.1007/s10278-010-9293-6

benefits of volumetric imaging while reducing the impact of the increased number of images.

Computed tomography images are typically viewed in a slice-by-slice mode, that is, radiologists view images sequentially, one at a time, and quickly move through the three-dimensional (3D) volume. Slice-by-slice viewing requires radiologists to mentally reconstruct the local 3D structure to differentiate between 3D features of interest. Although this is the reading method for which radiologists have been trained, it is not likely optimal for reading 3D datasets, especially with the increasing number of acquired images per examination and the decreasing information on individual slices, as the axial resolution of CT increases. To assist radiologists in reading and interpreting CT images, a variety of display algorithms have been developed to integrate information between slices and provide a 3D representation of imaged objects.⁵⁻⁷

The most common volumetric display method used for the task of lung nodule detection is maximum intensity projection (MIP), which allows thicker slabs comprised of multiple slices to be viewed. MIP projects the slab by taking the highest voxel value along a projection path as a final pixel value for display. The use of thicker slabs improves differentiation between nodules and linear structures, such as vessels, but also introduces the possibility of artifacts due to tissue superposition. In addition, MIP does not always correctly portray geometric structure and therefore may lead radiologists to misinterpret rendered images, especially images with small structures. Recently, several groups including us have begun to investigate the application of stereoscopic displays for radiological imaging.⁸⁻¹⁰ Unlike other volumetric display methods, stereo display exploits visual stereopsis to avoid ambiguities associated with structure superimposition. One study has shown that stereo mammographic display improved detection and characterization performance compared to projection mammograms.⁹

Chest CT for lung nodule detection is particularly well suited for stereo display. The lung geometry, at the inspiration level maintained during screening examinations, makes it particularly easy to create unambiguous projections along the axial direction. In addition, by rendering the air transparent, the rather sparsely distributed structures of interest can be visually isolated in the 3D

space. Displaying chest CT in stereoscopic 3D could potentially increase the efficiency of radiologists in reading these examinations, and by enabling radiologists to derive more information from the 3D datasets, it may increase their performance for detecting and classifying lung nodules. However, the potential benefit that can ultimately be realized from stereo displays is unknown.

Despite much effort devoted towards developing 3D display methods for volumetric imaging in radiology, there have been few studies focused on comparing the relative performances of these methods. To investigate the practical impact of stereo and other more conventional displays on the radiologists' performance, we have developed a stereographic display workstation for lung CT images.^{10,11} This study uses that display to compare three display modes for lung nodule detection, namely, stereoscopic, MIP, and slice-by-slice. The relative performances of these modes, as measured by receiver operating characteristic (ROC) analysis, and other issues are addressed below.

MATERIALS AND METHODS

The study described herein followed a fully crossed ROC paradigm in which eight radiologists interpreted chest CT examinations using each of three display modes. The primary statistical issue being addressed is whether different display modes resulted in different levels of performance for the detection of lung nodules.

Dataset

In this study, 100 chest CT examinations were randomly selected from a screening study being conducted in our institution of smokers between the ages of 50 and 79. The CT images were acquired on a LightSpeed, 4-slice CT scanner (GE Medical Systems, Milwaukee, WI) following a low-dose protocol. Each slice was 512×512 pixels, where the pixel size ranged from $0.69 \times 0.69 \text{ mm}^2$ to $0.94 \times 0.94 \text{ mm}^2$ in the axial plane. The slice thickness was 2.5 mm for all examinations. GE in-house software was used to post-process the images using a lung kernel mask. Because the screening is performed in a high-risk population,

nodule prevalence is relatively much higher than that in the general population. Basic statistical data of the dataset are listed in Table 1, and the size distribution of nodules is shown in Fig. 1.

To establish a gold standard for individual nodules in the dataset, we combined findings from a previous expert panel that had reviewed these cases with all findings from all readers and all display modes in this study. The location of each finding was represented by a 3D coordinate in the volumetric data, and we assumed that if two findings were within 15 pixels of each other, they corresponded to the same finding. All findings that were reported in this study, but had not been previously identified by the expert panel, were reviewed by experts who did not otherwise participate in this study to determine whether they were actual nodules or false-positives. In cases where there was unresolved ambiguity, the findings were excluded from our analysis. This process provided the finalized truth file that served as a gold standard in evaluating performance. Although this method of establishing a gold standard is imperfect, at the present time, given that performing a biopsy of all suspected lung nodules is infeasible, this is the most accurate method of establishing truth that is available for this type of study.

Display Modes

Slice-by-slice Display This display mode was equivalent to the conventional method of viewing CT images in clinical practice. Specifically, individual slices were displayed in sequence, under the control of viewers. Viewers were also able to adjust window and level settings, but other than that, the slices were not modified. All images were displayed on a single monitor.

Stereo Display The general methods we use for stereo rendering from CT slices, and our stereo display system, have been introduced elsewhere.¹⁰ A brief description of the process is given here.

Fig 1. Distribution of nodule sizes in the dataset.

A stack of sequential CT slices represents an anisotropically sampled tissue volume (i.e., resolution in the axial direction is less than in the transverse direction). To create a pair of image projections from such a volume, suitable for stereographic viewing, the raw CT slices are first conceptually positioned in their correct spatial relationship. This geometrically correct volume is then projected from two different viewpoints to generate left-eye and right-eye views. Each pixel in one of these views is calculated as a weighted integral of values along the oblique ray between the pixel and the corresponding viewpoint. At each of a fixed number of periodically spaced positions along the ray, values are calculated for the position on the ray by trilinear interpolation of the eight CT voxels comprising the vertices of the rectangular parallelepiped encompassing the position. This interpolation process has the potential to slightly blur sharp edges in the original data. In practice, given the degree of filtration applied during the original CT reconstruction and the fact that these images contain relatively few sharp edges, the edge blurring effect appears to be minimal. Of more concern is the artifact attributed to inadequate resolution of CT in the axial dimension, which is present in all reading modes.

To calculate the projections, we assumed that the interocular distance was 6.5 cm, and viewers would be viewing a 25×25-cm square region of the display from a distance of 50 cm. The geometric aspects of projection are well understood and widely used in computer graphics.

Table 1. Basic Characteristics of the Dataset Used for the Study

Case			Nodule		
Total number	Number of zero nodules	Number of at least one nodule	Total number	Average number/case	Average size
100	5	95	647	6.47	4.4 mm

The main differences between projection methods relate to the methods by which a projected pixel value is calculated from the voxel values along rays of the projection. For this study, we used a depth weighing of the maximum voxel value along each ray, which has the advantage, similar to MIP, of providing high contrast between nodules and the background. Depth weighting was based on a model of optical transmission and was included to provide an additional depth cue.

Stereo display is achieved by displaying left-eye and right-eye views on separate monitors and combining the images with a mirror to form a superimposed cross-polarized image. When observers view the superimposed image through glasses having cross-polarized lenses, their left eye sees only the left-eye view and their right eye sees only the right-eye view, giving the impression that they are viewing a 3D object.

Observers were permitted to dynamically adjust slab thickness and axial position at will. The thinnest slab for stereo display is one CT slice, which is equivalent to the slice-by-slice display mode described above. Observers were also allowed to adjust window and level settings as desired.

Maximum Intensity Projection Display The MIP for each pixel on a displayed image is obtained by choosing the highest voxel value through each ray of voxels perpendicular to the plane of the slices. Rendered MIP images were displayed on a single monitor.

As in the case of stereo display, viewers were given dynamic control over slab thickness (i.e., the number of slices) and slab position in the axial direction. Again, the minimum slab thickness was one slice, so that the slice-by-slice mode was encompassed by the MIP mode.

Display Implementation

A display system was configured for this project to provide real-time execution of all functionality available to observers. The computer used for display and the study was a 2.8-GHz AMD Athlon 64 personal computer with three hard disk drives connected via RAID technology to create a disk capacity of 600 GB. The graphics card installed in the computer, an NVIDIA Quadro FX1100, is stereo-capable in that it provides multiple display

buffers and OpenGL/DirectX support necessary for stereographic presentation.

The actual display was a Planar SD2020 (Planar Systems, Inc., Beaverton, OR) comprised of two 20-in. monitors and a special mirror for superimposing images from the two monitors. The function of the stereo display is to present a left-eye image to the left eye and a separate right-eye image to the right eye. In the case of the Planar system,¹² which is depicted in Fig. 2, the two images are displayed on individual active matrix liquid crystal displays that are configured so that the planes of polarization of the two images are perpendicular. The cross-polarized images are combined with a mirror having 50% transmission, 50% reflectance, and an antireflective coating on its second surface to minimize secondary image reflection and prevent birefringence. When the superimposed images are viewed through crossed-polarizing glasses, each eye sees only its corresponding image.

A customized key pad was programmed to provide all the functionality required for the study, including window/level adjustment, slice position selection, and slab thickness adjustment in the case of stereo and MIP displays.

Images rendered in stereo and MIP were precalculated and prestaged for a range of viewing thickness (number of CT slices in a slab). The maximum slab thickness permitted in the stereo and MIP modes was 15 CT slices. The CT images were processed and viewed in axial plane for all three display schemes. The computer's performance and stereographic capability were tested to

Fig 2. Picture depicting the stereo display and keypad in use.

insure that they met the level of performance required in this study.

Data Collection and Interpretation

Eight radiologists, each having more than 5 years of experience, participated in this study. All cases were read in each of the three display modes by each radiologist. The specific task was to detect and report the location of individual lung nodules.

A randomized display mode was established for each reading session. During a reading session, cases were randomly picked with the constraint that the same case could not be repeated (in a different mode) unless a period of at least 2 weeks had elapsed to avoid any bias caused by readers remembering previously seen cases. There was no time limit for a reading session, but the average time was approximately 1 hour. Readers reported their findings on an electronic scoring form, which could be completed by using a computer mouse. Prior to the study, the radiologists were given time to become familiar with and practice the reading protocols for each display mode. All activity of readers, (e.g., changing slab thickness or axial position) was recorder by time-of-occurrence to enable subsequent comparisons between display modes.

Analysis

Receiver operating characteristic methodology has been widely used in evaluating performance of imaging systems, computer-assisted detection schemes, and many other diagnostic task-related medical devices. Traditionally, the classification of disease states with an ROC study is case-based regardless of the number of lesions in a case. As we mentioned, our lung CT examinations were drawn from a high-risk population; the average number of nodules per case was 6+; and very few cases were nodule free. Therefore, applying case-based ROC analysis to our dataset would not have sufficient statistical power and sensitivity to differentiate classification performance between the three display modes. To more properly use the dataset to reveal the nature of the three display schemes, we treated each finding as one detection instance and used finding-based instances for ROC curve analysis. The area under ROC curves, A_z , was used as the figure-of-merit to characterize the

performance of each display mode for finding-based results.

This method of analysis cannot provide absolute measures of performance because any nodules that were not detected by either the readers in the study or the expert panel were not counted as false-negatives. Nevertheless, these undetected nodules have no impact in comparing the relative performances between display modalities.

Radiologists' reading time was recorded every 0.25 second for each examination and each display mode. The total reading time for each case was obtained by adding up the segments, whereas times spent on any particular position of slice/slab exceeding 1 minute were excluded on the assumption that these excessive long times were results of activities other than examination interpretation. Reading time was analyzed for each reader and each display mode. The Wilcoxon–Mann–Whitney test was used to compare the amount of time between display modes.

RESULTS

The mean time for reading each case varied between readers for all cases and all display modes, ranging from 0.88 ± 0.63 to 5.50 ± 2.01 minutes. The mean time over all readers and cases were 2.70 ± 2.01 , 2.78 ± 1.89 , and 2.75 ± 1.92 minutes for stereo, MIP, and slice-by-slice, respectively. There is no significant statistical difference of interpretation times between the three display modes.

For the stereo and MIP display modes, the time spent on reading as a function of slab thickness was recorded and expressed as a percentage of the total reading time. Figure 3 shows the results of averaged times from the eight radiologists for each slab thickness. The results indicate that most of the radiologists preferred reading 3D images with a slab thickness equivalent to about five slices, and there was no difference in the time distribution pattern between the stereo and MIP modes.

Averaged areas under ROC curves (A_z) for the performance on stereo, MIP, and slice-by-slice displays are 0.67 ± 0.06 , 0.65 ± 0.06 , 0.65 ± 0.04 , respectively. Although the A_z for stereo is higher than for MIP or slice-by-slice, there were no statistically significant differences between performances of the three display modes.

Fig 3. Average time, expressed as percent of total time, spent viewing different slab thickness for stereo display (a) and MIP display (b). At each slab thickness, eight bars are plotted, each representing the time spent by one of the eight radiologists.

DISCUSSION

Radiology is rapidly shifting from film-based projection imaging to imaging systems that acquire 3D datasets for many applications, and this has dramatically affected the process of reading and interpretation in radiological practice. By far, the most common method used for viewing inherently 3D data is reading 2D images sequentially from the 3D dataset in a slice-by-slice mode—a laborious task that can be error prone.

A dilemma arises with volumetric imaging because to increase axial resolution and to make voxels more isotropic, it is necessary to acquire thinner slices. This increase in resolution is desirable for many reasons but is necessary to depict smaller features and to avoid stair-step artifacts, which are a characteristic of certain projections of oblique vessels. However, the acquisition of thinner slices can dramatically increase the number of slices that must be reviewed by a radiologist. At the same time, for a fixed total exposure, the effective exposure per slice is reduced, which results in a corresponding deterioration in the signal-to-noise ratio of individual slices. This loss of signal-to-noise ratio

makes the tasks of detecting small objects much more difficult. Given these considerations, this study was based on the hypothesis that some form of volumetric display will be needed to maximize performance and efficiency of the interpretation process.

In this study, the overall performance of all three display modes was relatively low. The main reason for this was inter- and intrareader variability. When examining the results, we found that about 40% of the total findings were reported by only one radiologist in one of the three display modes (Fig. 4). Among those single-observation findings, about 37% are true-positive findings, and these are equally distributed within the three display modes, which implies that for the three display modes tested, the diagnostic performance for lung nodule detection is equivocal if the detection of single nodules is of importance. However, case-based performance measures would be higher because most positive cases contained multiple nodules.

At the axial resolution used throughout this study, each case consisted of roughly 100 slices. This number of slices was too few for us to be able to demonstrate unequivocal improvements in efficiency, particularly given confounding effects such as the observer's lack of experience with the relatively novel stereo display and the fact that they spent some time in the MIP and stereo modes viewing data as very thin slabs (Fig. 3). As the number of slices in a case increases, as happens when the axial resolution is increased, the efficiency of the stereo display relative to the slice-by-slice mode should increase.

Fig 4. Distribution of all findings (positives and negatives) is plotted against the number of observations. The maximum number of observations for each finding is 24 (the candidate nodule was observed by all eight radiologists in all three display modes) and the minimum number of observation for each finding is 1 (the candidate nodule was found by only one radiologist in only one of the three display modes).

In our preliminary study¹⁰ we reported that some of the radiologists spent most of their time viewing slabs comprised of a single slice in the stereographic display mode. The most likely explanation for this is that they were relatively more familiar with slice-by-slice display than with stereo display. In this study, we reinvestigated this issue by examining the amount of time radiologists spent viewing different slab thicknesses in the MIP and stereo display modes. Figure 3 shows that, with either MIP display or stereo display, radiologists spent most of their time viewing slabs equivalent to approximately five slices. We attribute this change to the increased experience our observers have in interpreting 3D datasets, which suggests that with increasing exposure, training, and practicing in 3D imaging and 3D displays, radiologists can quickly adopt new imaging technologies.

Advanced medical imaging technology has dramatically improved image quality and dose efficiency for disease screening and diagnosis. However, as the 3D datasets acquired in radiology grow in size and complexity, there must be a corresponding adaptation of display and interpretation paradigms if a commensurate improvement in observer performance and efficiency is to be achieved. There are a variety of compelling reasons to believe that stereographic display methods will provide considerable benefits for the display of 3D datasets, though the efficacy of these methods needs to be tested empirically.

Any meaningful comparison of stereographic display technology to traditional display methods will require that observers gain sufficient experience with the new technology to discover optimal methods for its utilization. The comparison problem is compounded by the increasing number of display features and image processing options that must be considered in connection with stereographic display. Nevertheless, the results of this study, although not significant, suggest that the stereographic display of large 3D datasets may have advantages in efficiency and detection performance over more traditional displays, although

further work will be required to establish this conclusively.

ACKNOWLEDGEMENT

This work is sponsored in part by the US Army Medical Research Acquisition Center, 820 Chandler Street, Fort Detrick, MD 21702-5014, under Contract PR043488. The content of the contained information does not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred.

REFERENCES

1. Data & Statistics. Available at: External link <http://www.lungusa.org/site/c.dvLUK9O0E/b.33347/>. Accessed April 7, 2009.
2. SEER Stat Fact Sheets. Available at: External link <http://seer.cancer.gov/statfacts/html/lungb.html>. Accessed April 7, 2009.
3. Kabat GC: Recent developments in the epidemiology of lung cancer. *Semin Surg Oncol* 9:73–79, 1993
4. Molina JR, Yang P, Cassivi SD, et al: Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc* 83:584–594, 2008
5. Beigelman-Aubry C: Post-processing and display in multislice CT of the chest. *JBR-BTR* 90:85–88, 2007
6. Vick 3rd, GW: Three- and four-dimensional visualization of magnetic resonance imaging data sets in pediatric cardiology. *Pediatr Cardiol* 21:27–36, 2000
7. Uttecht S, Thulborn KR: Software for efficient visualization and analysis of multiple, large, multi-dimensional data sets from magnetic resonance imaging. *Comput Med Imaging Graph* 26:73–89, 2002
8. Goodsitt MM, Chan HP, Darner KL, et al: The effects of stereo shift angle, geometric magnification and display zoom on depth measurements in digital stereomammography. *Med Phys* 29:2725–2734, 2002
9. Chan HP, Goodsitt MM, Helvie MA, et al: ROC study of the effect of stereoscopic imaging on assessment of breast lesions. *Med Phys* 32:1001–1009, 2005
10. Wang XH, Good WF, Fuhrman CR, et al: Stereo CT image compositing methods for lung nodule detection and characterization. *Acad Radiol* 12:1512–1520, 2005
11. Wang XH, Durick JE, Lu A, et al: Characterization of radiologists' search strategies for lung nodule detection: slice-based versus volumetric displays. *J Digital Imaging* 21(Supplement 1):39–49, 2008
12. Stereo 101—Stereoscopic Display Technology. Available at: External link <http://www.planar3d.com/3d-technology/stereoscopic-101/>. Accessed March 31, 2010