Fractional labelmaps for computing accurate dose volume histograms

Kyle Sunderland, Csaba Pinter, Andras Lasso, Gabor Fichtinger
Laboratory for Percutaneous Surgery, School of Computing, Queen's University, Kingston, Canada

ABSTRACT

PURPOSE: In radiation therapy treatment planning systems, structures are represented as parallel 2D contours. For treatment planning algorithms, structures must be converted into labelmap (*i.e.* 3D image denoting structure inside/outside) representations. This is often done by triangulated a surface from contours, which is converted into a binary labelmap. This surface to binary labelmap conversion can cause large errors in small structures. Binary labelmaps are often represented using one byte per voxel, meaning a large amount of memory is unused. Our goal is to develop a fractional labelmap representation containing non-binary values, allowing more information to be stored in the same amount of memory.

METHODS: We implemented an algorithm in 3D Slicer, which converts surfaces to fractional labelmaps by creating 216 binary labelmaps, changing the labelmap origin on each iteration. The binary labelmap values are summed to create the fractional labelmap. In addition, an algorithm is implemented in the SlicerRT toolkit that calculates dose volume histograms (DVH) using fractional labelmaps.

RESULTS: We found that with manually segmented RANDO® head and neck structures, fractional labelmaps represented structure volume up to 19.07% (average 6.81%) more accurately than binary labelmaps, while occupying the same amount of memory. When compared to baseline DVH from treatment planning software, DVH from fractional labelmaps had agreement acceptance percent (1% ΔD, 1% ΔV) up to 57.46% higher (average 4.33%) than DVH from binary labelmaps.

CONCLUSION: Fractional labelmaps promise to be an effective method for structure representation, allowing considerably more information to be stored in the same amount of memory.

Keywords: Fractional labelmap, Dose volume histogram, Radiation therapy, Treatment planning

1. PURPOSE

In radiation therapy treatment planning systems (TPS), structures such as targets or organs at risk are stored as 2D planar contours, which are segmented on parallel 2D image slices from computed tomography scans. Before they can be used in various treatment planning and evaluation algorithms, the structures must be converted into binary volumes in which each voxel indicates whether it is contained within the structure. These binary 3D volumes are referred to as binary labelmaps. In the process of conversion, planar contours are often first converted into a triangulated (closed) surface mesh, before being converted into a binary labelmap (see Figure 1). This allows the structures to be easily visualized as surfaces and creates a continuous surface for the structure between the sparse planar contours. This conversion results in errors that are caused by the existence of false positive and false negative regions within each of the border voxels. In large structures, these errors only have a small effect due to the relative size of the volume difference to the total structure volume, however in small structures, these differences can have a large effect on the total volume of the structure.

The errors that are caused by voxelization can potentially have a negative effect on the accuracy of metrics derived from the structure. Dose volume histogram (DVH) is a metric that is calculated using a dose distribution image and a labelmap, and are used to measure the intensity and distribution of radiation that each structure receives. It is important that the DVH for each structure is as accurate as possible [1], since DVH metric values such as the percentage of the structure volume that receives 20 Gy of radiation (V20), and the minimum dose that is received by 5cc of the structure (D5cc) could be skewed by differences in the DVH. In inverse treatment planning methods, such as the one by Li *et al.* [2] metrics such as V20 and D5cc are used to evaluate whether a treatment plan meets the desired standards, and to guide the inverse planning process towards higher quality plans. If there are significant errors in the DVHs provided

to the system, it may result in the selection of a sub-optimal plan. This mistake may not be found before treatment, since the observed DVH metrics would support the current plan selection.

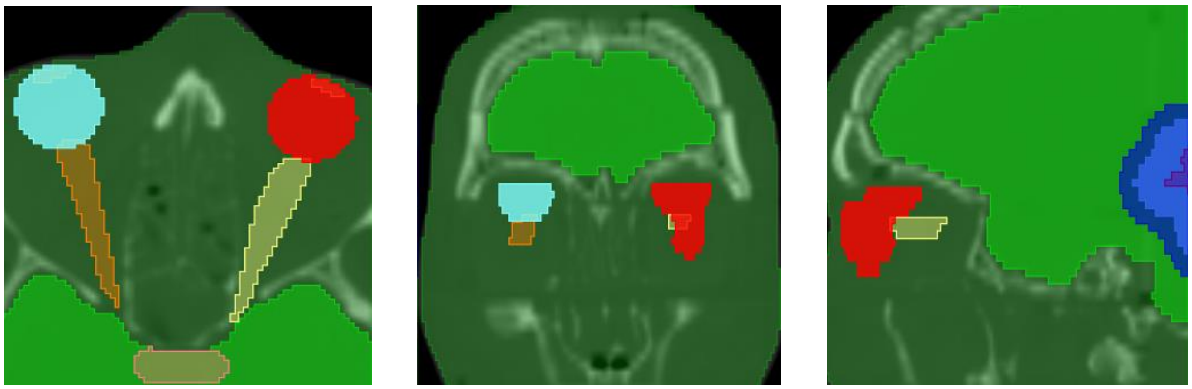

Figure 1. Binary labelmap created from head and neck phantom dataset. The labelmap has the same resolution as the CT volume

In many medical image analysis applications such as the widely used open-source medical imaging platform 3D Slicer [3], binary images are stored using a single byte of memory to represent each voxel. This presents an opportunity to utilize the remaining memory to increase the accuracy of stored segmentation information. This suggests an alternative, more accurate type of labelmap, called a fractional labelmap, in which the value of each voxel represents the fraction of the voxel that is filled by the structure, rather than a simple binary inside/outside classification.

Fractional labelmap representations have been previously used most frequently for automated image segmentation [4] [5]. In this method of representation, the voxel values represent the probability that the structure is contained inside [4], or represents the probability of whether the tissue is normal or abnormal [5]. These methods cannot be used to create fractional labelmaps from closed surfaces however, since they only operate on the original image data. This means that a closed surface to fractional labelmap conversion algorithm is required for uses such as in radiotherapy TPS algorithms.

2. METHODS

The algorithm for converting closed surfaces to fractional labelmaps was implemented within the Segmentations module in 3D Slicer. The structure contours were first converted into surface mesh [6], which were then converted into fractional labelmap representations. When we represented the fractional value of each voxel in a single byte of memory, there were a maximum of 256 possible values, between 0 and 255. To ensure that the resolution was increased uniformly in all dimensions, the resolution could only be increased by a maximum factor of six (6^3=216), since a factor of seven (7^3=343) would exceed the allowed range of values. This led to the first implemented approach for creating a fractional labelmap from a closed surface mesh, in which the surface was converted into a binary labelmap representation that had a resolution that was 6 times higher in each dimension than the desired final resolution. The binary labelmap was then converted into a fractional volume by summing the number of non-zero binary voxels that were contained in each of the large voxels in the fractional labelmap. This resulted in an image that contained a range of 217 possible values between 0 and 216 in each voxel. For large structures, this brute force oversampling method did not work, since for large structures, the high resolution binary labelmap was too large to reliably allocate in memory.

To avoid this problem, the same principle was used in a different approach to instead construct 216 binary labelmaps at the original resolution, rather than a single high resolution labelmap. Since the value of a voxel in a binary labelmaps reflects whether the centre of the voxel is contained within the structure, it is possible to determine if any point is within the structure by shifting the origin of the labelmap so that the center of the voxel is moved to the desired location. This principle was applied so that in each of the 216 generated labelmaps, the origin was moved to one of 216 evenly spaced points that form a cubic grid pattern within the voxels of the original labelmap (see Figure 2). For voxels along the border of the surface, only some of the points will be inside the structure. The number of points that are inside will

indicate the fraction of the total voxel volume that is occupied by the structure. The 216 binary labelmaps can be calculated independently, meaning that only one offset labelmap needs to be allocated in memory at a time. Since the labelmaps are calculated independently, this also presents an opportunity to multi-thread the algorithm, allowing multiple labelmaps to be calculated in parallel. Once a binary labelmap has been calculated, the values of the voxels are added to the fractional labelmap. In the final labelmap, each voxel represents the sum of the values in the same voxel for all the binary labelmaps.

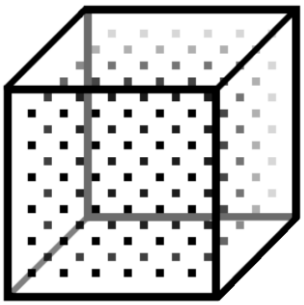

Figure 2. 216 evenly spaced offset points within the original voxel. The origin is shifted so the voxel centres represent different points

In order to evaluate the accuracy of the fractional labelmap, the volume difference between the original closed surface mesh and the binary/fractional labelmaps was measured. To test this, a closed surface mesh was converted to both fractional and binary labelmaps. The total volume of the closed surface was then calculated, and served as the ground truth value. The volume of the binary labelmap was found by calculating the number of non-zero voxels and multiplying by the voxel volume. Similarly, the volume of the fractional labelmap was found by summing the fractional value of the voxels, where each voxel was represented as a floating point number between zero and one, and multiplying by the voxel volume. The representations were evaluated by calculating the difference between the volume of the closed surface and the binary/fractional labelmap as a percentage. Metrics such as Dice Similarity Coefficient and Hausdorff distance were not used, since they are traditionally only utilized to evaluate differences between binary labelmaps, and are unable to be used for fractional labelmaps without modification.

The effectiveness of fractional labelmaps was also evaluated by creating a modified algorithm DVH within the open-source radiation therapy research toolkit SlicerRT [7]. These modified DVH calculations can utilize the extra information from the fractional labelmap to more accurately calculate the volume of the structure that receives a specific dose of radiation. The DVH generated using the binary and fractional labelmaps are then compared to baseline values generated from two radiotherapy TPS: Eclipse™ (Eclipse™ radiation therapy treatment planning system, Varian Medical Systems, Inc.) and CERR (Computational Environment for Radiotherapy Research) [8]. The comparisons use the metrics developed by Ebert *et al.* [9], which are implemented in the DVH Comparison module of SlicerRT.

An algorithm was also implemented for converting fractional labelmaps back into a closed surface mesh. Using the marching cubes algorithm, the surface is created at the fractional value which represents voxels which are 50% occupied by the structure. The surface that is generated is a smoother and more accurate representation of the surface than a closed surface generated in the same way from a binary labelmap, which will contain many sharp edges.

To determine the accuracy of the closed surfaces generated from the conversion, the results were evaluated by calculating the absolute closest point to point distance from the reconstructed surface to the original surface mesh. The same calculations are then performed between the surface created from the binary labelmap and the original closed surface for the same structures to provide a comparison between the two representations. These comparisons were performed using the using the Model To Model Distance extension in 3D Slicer.

3. RESULTS AND DISCUSSION

The accuracy of the closed surface to fractional labelmap conversion algorithm was tested by comparing binary and fractional representations for numerous different structures from artificial phantom datasets (see Figure 3). The conversion from closed surface to fractional labelmap took ~9.21s, while the closed surface to binary labelmap conversion took ~0.58s on the same machine (Intel® Core™ i7-4790K CPU @ 4.00GHz, 16GB ram).

Figure 3. Fractional Labelmap created from head and neck phantom dataset. The labelmap has the same resolution as the CT volume

When the volume of structures was calculated from both the binary and fractional labelmaps, it was found that the volume from fractional labelmaps was substantially more accurate than the volume of binary labelmaps calculated at the same resolution (see Table 1) compared to the surface mesh volume. This is because binary labelmaps are only capable of representing whether or not the centre of the voxel is within the structure, while each voxel in a fractional labelmap represent 216 different points. In large structures, the improvements were small relative to the structure size, however in small structures, the improvement in accuracy between binary and fractional labelmaps was found to be as high as 19.07% of the closed surface volume for structures such as the left optic nerve in the head and neck phantom.

Table 1. Difference between the volume of closed surface and binary/fractional labelmaps from the head and neck phantom dataset

Structure	Closed Surface Volume (cc)	Binary Volume (cc)	Fractional Volume (cc)	Binary Volume Difference (%)	Fractional Volume Difference (%)	Improvement Between Binary and Fractional Volume (%)
Body	8031.29	8053.87	8031.24	0.28	0.00	0.28
Brain	1113.30	1114.65	1113.12	0.12	0.02	0.10
Brain Stem	30.99	31.42	30.98	1.41	0.01	1.40
PTV	126.18	127.03	126.17	0.68	0.01	0.67
CTV	69.52	70.11	69.52	0.84	0.01	0.84
GTV	8.00	8.46	7.99	5.86	0.07	5.79
Lens – Lt	0.12	0.14	0.12	17.59	0.20	17.39
Lens – Rt	0.11	0.13	0.11	10.58	0.06	10.52
Optic Chiasm	1.01	1.15	1.00	14.21	0.08	14.13
Optic Nerve-Lt	1.55	1.76	1.54	13.69	0.04	13.65
Optic Nerve - Rt	1.89	2.26	1.89	19.32	0.25	19.07
Orbit - Lt	8.30	8.50	8.29	2.42	0.03	2.39
Orbit - Rt	8.56	8.76	8.56	2.37	0.02	2.36

To evaluate the effectiveness of fractional labelmaps compared to binary labelmaps for radiotherapy TPS algorithms, the DVH calculation algorithm in SlicerRT was modified to allow fractional labelmaps to be utilized. Using a manually contoured RANDO® head and neck phantom structures and an associated dose volume, DVHs were calculated using both binary and fractional labelmap representations (see Figure 4).

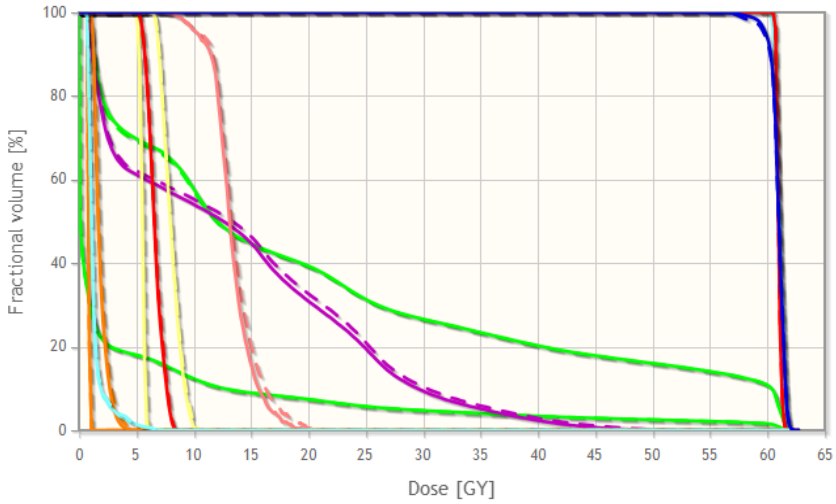

Figure 4. DVH from fractional (solid) and binary (dashed) labelmaps

Using the DVH comparison module in SlicerRT, DVH calculated from the fractional and binary labelmaps were compared against DVHs created from the same structures by Eclipse™ and. The DVH from the fractional labelmap had a higher or equal agreement acceptance percent (1% dose and 1% volume criterion) than the DVH from binary labelmaps in most structures (see Table 2). The brain stem structure was found to improve the most, increasing from 41.59% and 55.10% to 99.04% and 96.50% for Eclipse and CERR respectively. In the cases where the agreement acceptance percent was lower in fractional labelmap than binary labelmap, the difference was less than a fraction of a percent, which can be attributed to differences in contour interpolation and calculation error.

Table 2. Agreement acceptance % (1% ΔD, 1% ΔV) between DVH calculated in SlicerRT and DVH calculated from Eclipse™ and CERR the difference between fractional and binary agreements are calculated as: difference = fractional - binary

Structure	Binary and Eclipse™ (%)	Fractional and Eclipse™ (%)	Eclipse™ Improvement (Fractional - Binary) (%)	Binary and CERR (%)	Fractional and CERR (%)	CERR Improvement (Fractional - Binary) (%)
Body	100.00	100.00	0.00	100.00	100.00	0.00
Brain	100.00	100.00	0.00	100.00	100.00	0.00
Brain Stem	41.59	99.05	57.46	55.10	96.50	41.40
PTV	98.73	99.68	0.95	97.45	97.13	-0.32
CTV	98.41	99.05	0.63	97.45	97.45	0.00
GTV	98.41	98.41	0.00	98.09	98.09	0.00
Lens - Rt	99.68	99.68	0.00	99.36	99.36	0.00
Lens - Lt	98.73	98.41	-0.32	98.73	98.73	0.00
Optic Chiasm	79.37	80.63	1.27	81.21	80.57	-0.64
Optic Nerve-Lt	86.67	92.70	6.03	86.94	92.04	5.10
Optic Nerve - Rt	93.33	93.33	0.00	93.63	93.63	0.00
Orbit - Lt	95.87	95.24	-0.63	94.59	94.90	0.32
Orbit - Rt	96.19	96.83	0.63	93.31	93.95	0.64

The runtime of the DVH calculation algorithm for both binary and fractional labelmaps was compared by constructing fractional and binary labelmap representations from the head and neck structures, and measuring the execution time of the algorithm on both representations. The fractional labelmap DVH calculation was found to take a comparable amount of time compare to the binary labelmap method. For the body, which was the largest structure in the dataset, the DVH calculation using fractional labelmap took ~0.177s, compared to the binary labelmap's execution time of ~0.134s.

For the fractional labelmap visualization, we decided to use the variable opacity voxels that are shown in Figure 3 since the structures that are being represented have discrete, real-world boundaries. Instead, the visualization for fractional labelmaps in 3D Slicer was implemented by displaying a high resolution 2D binary image slice generated from the fractional labelmap for the 2D slice view. The fractional image is resampled at the display resolution using linear interpolation, before being thresholded at the 50% occupancy value. Pixels that contain 50% occupancy or greater were displayed as inside, while pixels that were lower than 50% occupancy are displayed as being outside the structure. The resulting image is then displayed in the slice view (see Figure 5). The resulting edges are smoother than the visualization of binary labelmaps at the same resolution, and more closely represents the original closed surface that the fractional labelmap is derived from.

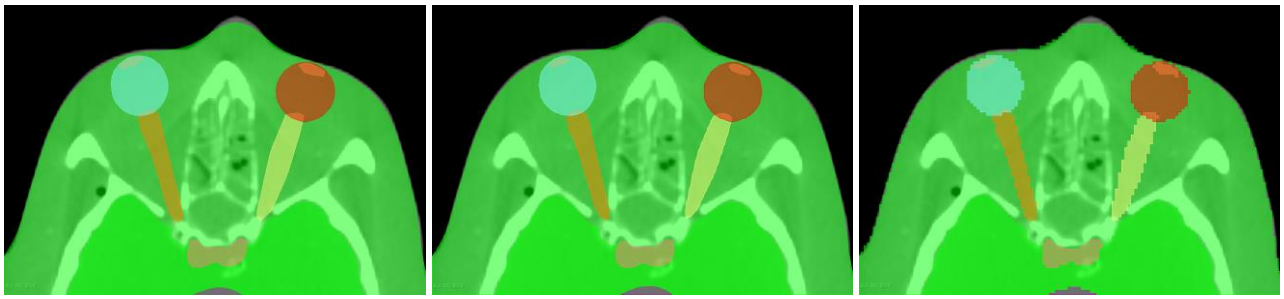

Figure 5. 2D Slice view representations in Slicer for original closed surface (left) as well as the fractional labelmap (center), and binary labelmap (right) created from the closed surface

When the fractional labelmap to closed surface conversion was run on the sample structures, it was found to be a visually closer representation of the original closed surface than the closed surface created from a binary labelmap at the same resolution (see Figure 6).

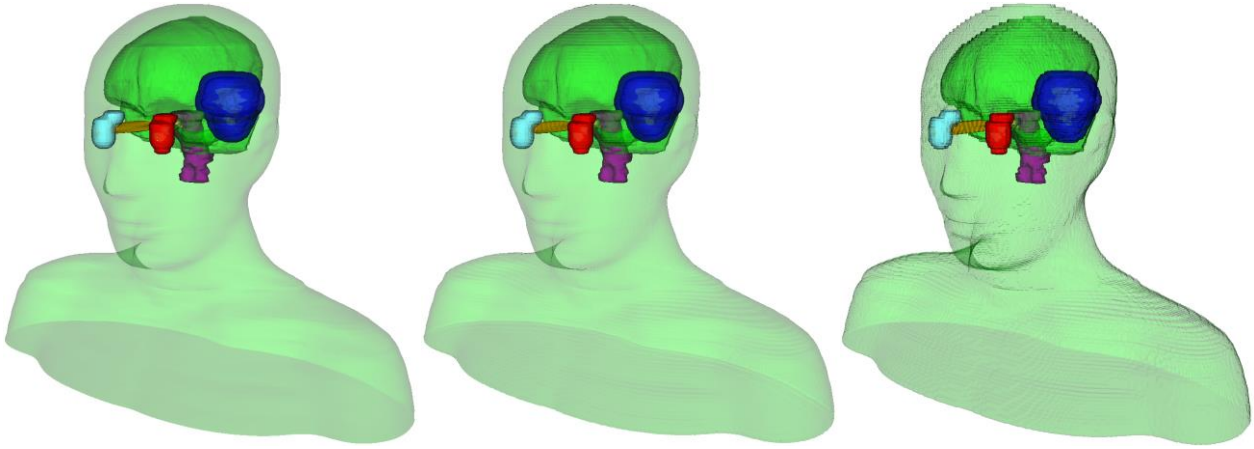

Figure 6. 3D closed surface representations for original closed surface (left), and closed surface created from fractional labelmap (center) and binary labelmap (right), without smoothing

The distance between the reconstructed surface and the closed surface was calculated using the Model To Model Distance extension in 3D Slicer. We found that the distance from the reconstructed surface created from the fractional labelmap representation had both a lower average, as well as a lower average distance for all structures when compared to the distance from the binary labelmap reconstruction (see Table 3).

Table 3. Absolute closest point to point distance between the reconstructed surfaces generated from fractional and binary labelmaps, and the original closed surface mesh

	Binary Generated Surface To Original Surface Distance		**Fractional Generated Surface To Original Surface Distance**		**Improvement (Binary - Fractional)**	
Structure	**Maximum (mm)**	**Average (mm)**	**Maximum (mm)**	**Average (mm)**	**Maximum (mm)**	**Average (mm)**
Body	1.25	0.29	0.60	0.07	0.65	0.22
Brain	1.35	0.32	0.82	0.08	0.52	0.24
Brain Stem	1.21	0.29	0.54	0.12	0.67	0.17
PTV	1.23	0.30	0.32	0.09	0.91	0.21
CTV	1.23	0.31	0.47	0.10	0.76	0.21
GTV	1.23	0.40	0.67	0.18	0.56	0.22
Lens - Lt	0.93	0.25	0.41	0.13	0.52	0.12
Lens - Rt	0.83	0.21	0.58	0.13	0.25	0.09
Optic Chiasm	1.16	0.31	0.54	0.14	0.63	0.17
Optic Nerve-Lt	1.21	0.28	0.51	0.13	0.70	0.16
Optic Nerve - Rt	1.25	0.42	0.70	0.17	0.55	0.25
Orbit - Lt	1.20	0.28	0.39	0.09	0.81	0.18
Orbit - Rt	1.17	0.26	0.28	0.08	0.89	0.18

4. CONCLUSION

We successfully implemented a method for representing segmented structures using fractional labelmaps within 3D Slicer. This method of representation is more accurate than a traditional binary labelmap, containing 216 times more information, while still being stored in the same amount of memory as a binary labelmap of the same resolution.

Segmented structures that were created from the closed surface to fractional labelmap conversion were found to be more accurate at representing volume than structures from binary labelmaps at the same resolution. In small structures, such as optic lenses and nerves, we found an increase in the total volume accuracy of up to 19.32% of the original structure volume, with an average improvement of 6.81%. The conversion from the closed surface mesh to the fractional labelmap representation was completed in ~9.21 for head-neck phantom structures. This process currently takes considerably longer than a regular closed surface to binary labelmap conversion (~0.58s), however with further optimization the execution time for this process will be reduced. The conversion from fractional labelmap back to closed surface mesh using marching cubes was found to produce a surface that was more accurate than the same conversion performed on a binary labelmap.

By modifying the DVH calculation algorithm to use fractional labelmaps, we have effectively shown that fractional labelmaps can be utilized for the same algorithms as binary labelmaps, with better results. The improvement in DVH accuracy for binary labelmaps from not always large, however in some structures, such as the brain stem, which previously had issues with DVH accuracy, we saw improvement of up to 57.46%, with an average improvement of 5.08% when compared to Eclipse**™**, and 3.58% when compared to CERR.

The conversion from fractional labelmap to a closed surface mesh was found to be more accurate than the same conversion performed on a binary labelmap when comparing the distance from the generated surface to the original surface. The maximum point-to-point distance between the structures had an improvement up to 0.91mm, with an average improvement of 0.65mm for all structures, while the average point-to-point distance had an improvement of up to 0.25mm with an average improvement of 0.19mm for all structures

Fractional labelmap representations are a segmentation modality which allows more structural information to be stored in the same amount of memory as binary labelmap representations. They have qualitatively better visualizations, and can be used for many of the same algorithms that use binary labelmaps by applying fuzzy logic principles. These modified algorithms have access to more data, and have an execution time that is still comparable to the original algorithm. Future work will be conducted on the implementation of fractional labelmap editor effects within the Segment Editor module in 3D Slicer, such as the "smooth brush" effect. This will allow for easy manual segmentation of fractional labelmaps that do not require an existing closed surface representation. In addition, more work will be done evaluate the accuracy of these conversion methods using metrics such as modified Dice and Hausdorff [10] [11].

5. ACKNOWLEDGMENTS

This work was supported in part Discovery Grants Program of the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Applied Cancer Research Unit program of Cancer Care Ontario with funds provided by the Ontario Ministry of Health and Long-Term Care. Gabor Fichtinger was supported as a Cancer Care Ontario Research Chair in Cancer Imaging. Kyle Sunderland was supported by the Ontario Graduate Scholarship (OGS).

REFERENCES

[1] Fraass, B., Doppke, K., Hunt, M., Kutcher, G., Starkschall, G., Stern, R., and Van Dyke, J., "American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: quality assurance for clinical radiotherapy treatment planning," Med. Phys. 25(10), 1773-1829 (1998).

[2] Li, N., Zarepisheh, M., Uribe-Sanchez, A., Moore, K., Tian, Z., Zhen, X., Graves, Y. J., Gautier, Q., Mell, L., Zhou, L., Jia, X., and Jiang, S., "Automatic treatment plan re-optimization for adaptive radiotherapy guided with the initial plan DVHs," Physics in medicine and biology 58(24), 8725-8738 (2013).

[3] Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J. C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Alyward, S., Miller, J. V., Pieper, S., and Kikinis, R., "3D Slicer as an image computing platform for the Quantitative Imaging Network," Magnetic resonance imaging, 30(9), 1323-1341 (2012).

[4] Noe, A., and Gee, J. C., "Partial volume segmentation of cerebral MRI scans with mixture model clustering," Biennial International Conference on Information Processing in Medical Imaging, 423-430 (2001).

[5] Warfield, S. K., Westin, C. F., Guttmann, C. R., Albert, M., Jolesz, F. A., and Kikinis, R., "Fractional segmentation of white matter," International Conference on Medical Image Computing and Computer-Assisted Intervention, 62-71 (1999).

[6] Sunderland, K., Woo, B., Pinter, C., and Fichtinger, G., "Reconstruction of surfaces from planar contours through contour interpolation," SPIE Medical Imaging 2015, pp. 94151R (2015).

[7] Pinter, C., Lasso, A., Wang, A., Jaffray, D., and Fichtinger, G., "SlicerRT: radiation therapy research toolkit for 3D Slicer," Med. Phys. 39(10), 6332-6337 (2012).

[8] Deasy, J. O., Blanco, A. I., and Clark, V. H., "CERR: a computational environment for radiotherapy research," Med Phys. 30(5), 979-985 (2003).

[9] Ebert, M. A., Haworth, A., Kearvell, R., Hooton, B., Hug, B., Spry, N. A., Bydder, S. A., and Joseph, D. J., "Comparison of DVH data from multiple radiotherapy treatment planning systems," Physics in medicine and biology, 55(11), 337-346 (2010).

[10] Zou, K. H., Wells III, W. M., Kaus, M. R., Kikinis, R., Jolesz, F. A., and Warfield, S. K., "Statistical validation of automated probabilistic segmentation against composite latent expert ground truth in MR imaging of brain tumors," International Conference on Medical Image Computing and Computer-Assisted Intervention, 315-322, (2002).

[11] Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M., Kaus, M.R., Haker, S.J., Wells, W.M., Jolesz, F.A. and Kikinis, R., "Statistical validation of image segmentation quality based on a spatial overlap index 1: Scientific reports," Academic radiology 11(2), 178-189 (2004).