Effects of voxelization on dose volume histogram accuracy

Kyle Sunderland, Csaba Pinter, Andras Lasso, Gabor Fichtinger

Laboratory for Percutaneous Surgery, School of Computing, Queen's University, Kingston, Canada

ABSTRACT

PURPOSE: In radiotherapy treatment planning systems, structures of interest such as targets and organs at risk are stored as 2D contours on evenly spaced planes. In order to be used in various algorithms, contours must be converted into binary labelmap volumes using voxelization. The voxelization process results in lost information, which has little effect on the volume of large structures, but has significant impact on small structures, which contain few voxels. Volume differences for segmented structures affects metrics such as dose volume histograms (DVH), which are used for treatment planning. Our goal is to evaluate the impact of voxelization on segmented structures, as well as how factors like voxel size affects metrics, such as DVH.

METHODS: We create a series of implicit functions, which represent simulated structures. These structures are sampled at varying resolutions, and compared to labelmaps with high sub-millimeter resolutions. We generate DVH and evaluate voxelization error for the same structures at different resolutions by calculating the agreement acceptance percentage between the DVH.

RESULTS: We implemented tools for analysis as modules in the SlicerRT toolkit based on the 3D Slicer platform. We found that there were large DVH variation from the baseline for small structures or for structures located in regions with a high dose gradient, potentially leading to the creation of suboptimal treatment plans.

CONCLUSION: This work demonstrates that labelmap and dose volume voxel size is an important factor in DVH accuracy, which must be accounted for in order to ensure the development of accurate treatment plans.

Keywords: Voxelization, Radiation therapy, Dose volume histogram, Treatment planning

1. PURPOSE

When creating radiation therapy (RT) treatment plans, medical images such as computed tomography (CT) scans are used to delineate structures such as targets and organs at risk as a series of 2D planar contours. In order to be used in various treatment planning steps, the contours for each structure must be converted into a binary labelmap, which is comprised of a volume containing a large number of voxels that indicate whether a particular region is considered to be inside the segmented structure.

The process of converting the contours into a binary volume is a required step that is implemented in all treatment planning systems (TPS). The implementations are not identical between the packages however, which can lead to differences in the labelmap volumes. Although these differences are small in large structures, they can become quite large in small structures that are comprised of only a small number of voxels, for example small organs at risk (OAR) such as optic chiasm or even rectum, which is emphasized in high dose gradient regions, where dose gradient is a measure of the rate of change in dose intensity. This can lead to significant differences in the resulting dosimetry and DVH, and may negatively impact the treatment planning procedure.

It is important that the DVH derived from the structures are as accurate as possible [1], since, they are used to evaluate the effectiveness and impact of a treatment plan. In automated TPSs which use inverse planning, such as the one by Li et al., DVHs are used as a metric in the automated process of plan optimization [2]. This means that an inaccurate DVH could potentially result in a suboptimal treatment plan being presented.

Our goal is to evaluate how the errors introduced by the process of voxelization impact the quality of the labelmaps that are used to represent structures of interest, such as targets and OAR in TPSs. In addition, we want to determine how factors like voxel size impact metrics, such as DVH, that are used in the decision making process during treatment planning.

2. METHODS

Implicit Function

We create a series of implicit functions in order to approximate a set of anatomical structures. Each of the structures are implicitly defined by a mathematical equation, which allows us to determine whether any given point is inside or outside of the structure. These functions are treated as the ground truth analytical representation for our structures that have infinitely high resolution. This allows us to construct labelmaps of a higher resolution than those typically used in clinical procedures. Analytical phantoms have been used previously as a phantom for TPS [3], and for comparing the effects of voxelization on dose calculation accuracy [4]. Each of the simulated targets and OAR that we use is created from the simple mathematical representations for spheres, cones and cylinders. We also use simple Boolean operations to add and subtract the functions, which allows us to create more complex representations (Fig. 1). One downside to using the composite implicit functions is the appearance of infinitely sharp edges. These edges occur at the intersection between implicit functions and can never be correctly voxelized at any resolution. This means that the resulting voxelized structures will have a volume that is slightly higher than the original implicit function.

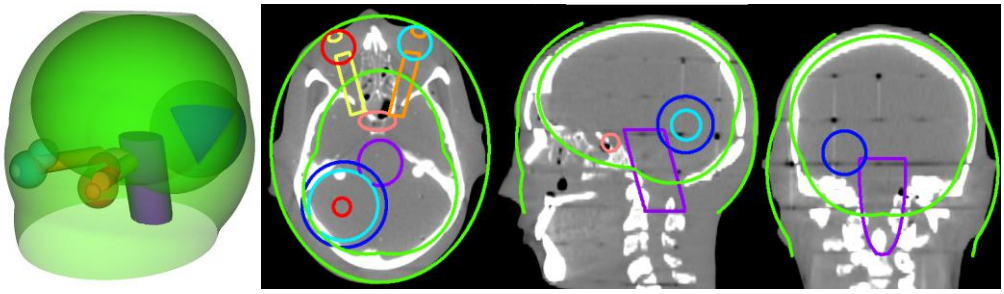

Fig. 1. Models representing the implicit functions that are used as the ground truth structures for the head and neck phantom in our experiments, and CT images showing the phantom that is being approximated.

Binary Labelmap

We convert these implicit functions into binary labelmaps by sampling the implicit functions at the center of each voxel in a volume of a given resolution. Voxels from the volume in which the implicit function evaluated to less than or equal to zero at the center are considered to be part of the structure, otherwise outside the structure. We iterate through each voxel in the volume for all of the implicit functions representing individual structures, resulting in a set of binary labelmaps (Fig. 2).

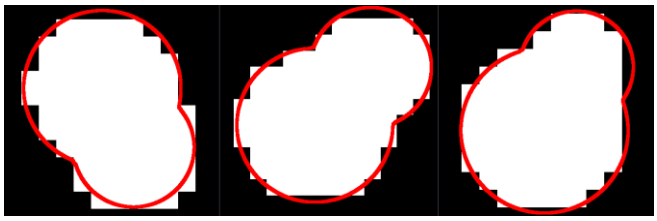

Fig. 2: Axial, Sagittal and Coronal views for the labelmap representing an implicit function at a 2.5mm voxel size.

We construct two binary labelmaps from our implicit functions. One labelmap at a traditional clinical resolution using cubical voxels with an edge length of 2.5mm, and another labelmap with smaller voxel size of 0.05mm, 0.15mm, or 0.3mm, with higher volume structures utilizing the larger voxel size. This difference in voxel size for the higher resolution structures is required in order to allow the images for large structures to be contained within memory. In order to quantitatively evaluate the differences between the labelmaps at two resolutions, we utilize two metrics: Dice coefficient and Hausdorff distance.

Dice similarity coefficient

The Dice similarity coefficient is a metric which is often used to compare the similarity between two sets of items with binary classification. In the case of binary labelmaps, it is a measure of the overlapping voxels between the two labelmaps [5]. In order to calculate the Dice coefficient, we resample the labelmap with the lower resolution into the voxel size of the high resolution labelmap. In order to correctly resample the low resolution as a binary volume, we need to use nearest neighbor interpolation, which offers a binary value for each pixel. We then traverse the labelmap in order to determine the number of true positive, false positive and false negative voxels, which we use to calculate the Dice coefficient.

Hausdorff distance

The Hausdorff distance is a quantifiable measure of difference between two sets of points, such as vertices or pixels in an image [6]. We calculate the Hausdorff distance by comparing the labelmaps as two sets of points in 3D space, with each point representing a border voxel in the labelmap. We find the shortest distance from each point in the first labelmap to a point in the second, as well as the reverse. The Hausdorff distance is considered to be the maximum of these distances.

Dose Volume

In order to evaluate the effect of labelmap voxel size on the DVH, we need to construct a simulated dose volume at an arbitrarily high resolution, which we can use in our calculations. We create a dose volume by simulating a plan containing multiple beams with a source-axis distance of 100cm and a focal region of $10cm^2$ (Fig. 3). We use a clinically observed tissue phantom ratio, as well as clinical cross-plane and in-plane dose profiles for our dose calculation. The voxels in the dose volumes are sampled at the same resolution as their corresponding labelmaps (Fig. 4). These volumes are not generated to be clinically perfect, however, we consider them to be qualitatively sufficient for our evaluation. Using the generated labelmaps, we construct dose volumes at the same resolution, which we can then use to calculate DVH for the structures. The DVHs are compared for each structure using a comparison algorithm from [10], which returns an agreement acceptance percentage, which indicates the number of bins in the DVH that are within a specified threshold. We used the algorithm two different thresholds of 3% volume-difference and dose-to-agreement (DTA) criterion, and a volume-difference and DTA of 1%.

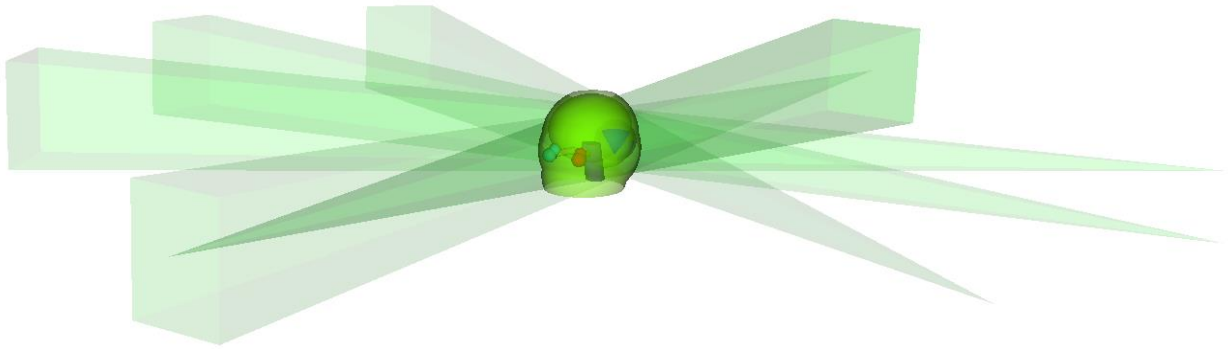

Fig. 3: Beam layout which was used to generate dose volumes at any specified resolution.

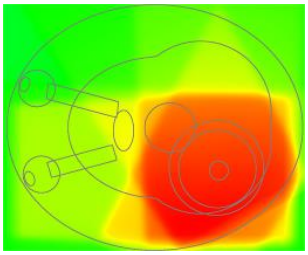

Fig. 4: Axial view of a dose volume generated from a dose plan of five simulated beams at a 0.3mm voxel size.

3. RESULTS AND DISCUSSION

In order to evaluate the effect that voxel size has on the DVH calculation process, we developed several modules in the SlicerRT [7] RT research toolkit which is based on the widely used medical image analysis and visualization platform 3D Slicer [8]. The modules are implemented as a combination of Python scripted modules and C++ loadable modules. These modules serve as a platform for implementing the steps required to perform the voxelization analysis. We created a series of implicit functions within 3D Slicer by using the VTK implicit function libraries [9]

We observed that the Hausdorff distance remained relatively constant across structures of varying size. The Dice Similarity Coefficient was found to be reduced in small structures such as the optic lenses and nerves. These differences could be potentially used to identify structures that are represented with low accuracy.

When the DVH for the high and low resolution labelmaps was plotted on the same graph, we observed differences between the two (Fig. 5).We found that there were only small differences in the DVH for large structures which contained many voxels. Even when they were sampled using a large voxel size, in large structures such as the body, brain, CTV, GTV, left femoral head, and right femoral head, there were no differences found between the DVH. In some structures that were a medium size, there was a large difference found when comparing the DVH using a 1% dose and volume to agreement criteria, however they were found to be much more accurate when using a relaxed comparison of a 3% dose and volume to agreement criteria. In small structures, such as the optic lenses and nerves, there were much larger differences found at both 1% and 3% dose and volume to agreement criteria. The differences in the DVH for the structures was amplified if the structures were located in regions with a high dose gradient (Table 1). This caused several types of artifacts to appear in the DVH for small structures, such as a staircase-like effect, as well as consistent differences between the low resolution and high resolution DVH (Fig. 6).

These effects can potentially cause high errors in the DVH V and D metrics, which are calculated from the dose and volume levels, and are used in treatment planning to assess the quality of a plan. If there are differences in the labelmap and volume for small targets or OARs, they could potentially create large errors that cause a suboptimal plan to be accepted. If the TPS uses inverse planning, then the process of determining if a plan meets the specified criteria could be compromised, since the DVH metrics are used as criteria for evaluating a plan effectiveness. This problem might not be noticed by the physician, since the DVH values would support the results that were provided by the TPS, except the problem would be caused by errors within the DVH itself.

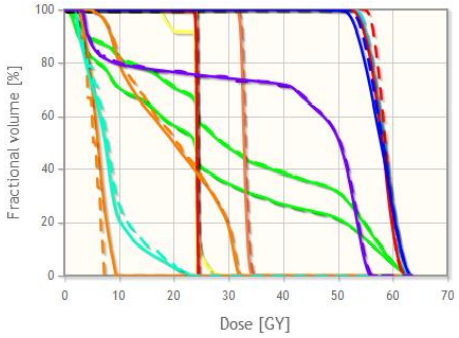

Fig. 5: DVH showing the differences between the high resolution (solid) and low resolution (dashed) labelmaps.

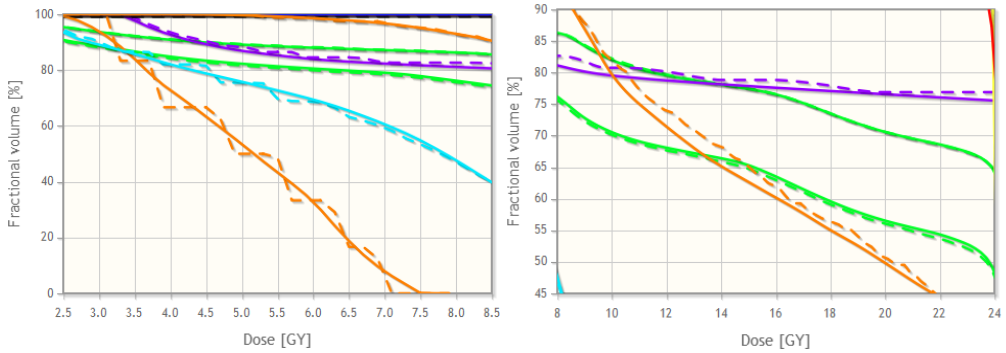

Fig. 6: DVH for the high resolution (solid) and low resolution (dashed) labelmaps showing the staircase effect of voxelization (left) and consistent differences between DVH (right).

Table 1: Hausdorff distances, Dice Similarity Coefficients and agreement acceptance percentage comparing labelmaps and DVH from a small reference voxel size against labelmaps with a clinical voxel size of 2.5mm.

	Reference Voxel Size (mm)	Hausdorff Distance (mm)	Dice Similarity Coefficient	Agreement Acceptance % (1% DTA)	Agreement Acceptance % (3% DTA)
GTV (Head Phantom)	0.150	2.970	0.946	97.77	100.0
CTV (Head Phantom)	0.150	2.783	0.970	100.0	100.0
PTV (Head Phantom)	0.150	2.920	0.956	100.0	100.0
Left Lens	0.050	2.861	0.794	99.19	99.19
Right Lens	0.050	2.901	0.808	51.22	53.66
Left Orbit	0.050	3.021	0.929	75.00	79.84
Right Orbit	0.050	3.021	0.929	50.42	76.47
Left Optical Nerve	0.050	3.085	0.866	90.24	94.51
Right Optical Nerve	0.050	3.229	0.884	28.48	59.39
Optical Chiasm	0.050	2.941	0.897	92.56	93.02
Brain Stem	0.150	3.015	0.958	59.23	100.0
Brain	0.300	2.858	0.987	100.0	100.0
Body	0.300	5.555	0.990	100.0	100.0
GTV (Prostate Phantom)	0.150	2.960	0.952	99.50	100.0
CTV (Prostate Phantom)	0.150	2.739	0.955	99.50	100.0
PTV (Prostate Phantom)	0.150	2.729	0.958	99.50	100.0
Left Femoral Head	0.150	3.021	0.946	100.0	100.0
Right Femoral Head	0.150	3.021	0.953	100.0	100.0
Rectum	0.150	3.402	0.925	80.20	100.0

4. CONCLUSION

We evaluated the errors introduced by the voxel size variation in order to quantify the resulting errors in the DVH that are produced from these structures. We found that the effect of varying voxel size had a small effect on DVH for large structures or in regions with a low dose gradient, however we observed an agreement acceptance percent as low as 28.48% (1% DTA) and 53.66% (3% DTA) for small structures and in regions with a high dose intensity variation. This indicates that size of voxels is an important factor to consider relative to the total volume of a structure.

5. FUTURE WORK

Future work will include investigating the impact of voxelization errors on treatment planning systems, as well as potential methods to mitigate the effects of voxel resolution on DVH. One potential approach is the use of fractional labelmaps as a functional alternative to the current use of binary labelmaps and supersampling.

Fractional Labelmap

Since pixels in binary labelmap images are often represented using a full byte of data, there is an opportunity to greatly increase the amount of information contained in a labelmap without increasing the storage requirements. Using a fractional labelmap, we can use each pixel in the image to represent the fraction of a voxel that is contained within a segmented structure.

In order to create fractional labelmaps, one method is to first construct regular binary labelmaps at a high resolution. This labelmap is then resampled into a lower resolution volume, with each voxel value representing the fraction of the voxel that is filled by the binary labelmap resulting in a fractional labelmap containing integer values between 0 and 255, rather than a binary value (see Fig. 7).

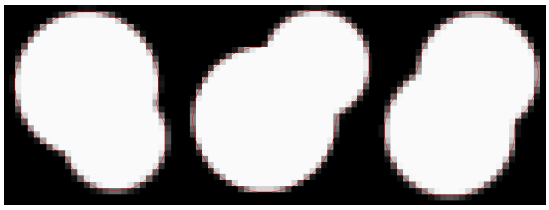

Fig. 7: Axial, Sagittal and Coronal views of a fractional labelmap with a 1mm voxel size, sampled from a binary labelmap with a 0.1666mm voxel size.

ACKNOWLEDGEMENTS

Gabor Fichtinger is supported as a Cancer Care Ontario Research Chair in Cancer Imaging.

REFERENCES

[1] Fraass, B., et al., “American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: quality assurance for clinical radiotherapy treatment planning,” Medical physics, 25(10), 1773-1829 (1998).
[2] Li, N., et al., “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] Zaidi, H., & Tsui, B. M., “Review of computational anthropomorphic anatomical and physiological models,” Proc. IEEE 97(12), 1938-1953 (2009).
[4] Peter, J., et al., “Analytical versus voxelized phantom representation for Monte Carlo simulation in radiological imaging,” Proc. IEEE Transactions on Medical Imaging 19(5), 556-564 (2000).
[5] Zou, K. H., et al., “Statistical validation of image segmentation quality based on a spatial overlap index,” Academic radiology, 11(2), 178-189 (2004).
[6] Huttenlocher, D. P., et al., “Comparing images using the Hausdorff distance,” Pattern Analysis and Machine Intelligence 15(9), 850-863 (1993).
[7] Pinter, C., et al., “SlicerRT: radiation therapy research toolkit for 3D Slicer,” Med. Phys. 39(10), 6332-6337 (2012).
[8] Fedorov, Andriy, et al. "3D Slicer as an image computing platform for the Quantitative Imaging Network." Magnetic resonance imaging, 30(9), 1323-1341 (2012).
[9] Schroeder, W., et al., [Visualization Toolkit: An Object-Oriented Approach to 3D Graphics], Kitware, Clifton Park, New York (2006).
[10] Ebert, M. A., et al., “Comparison of DVH data from multiple radiotherapy treatment planning systems,” Physics in medicine and biology, 55(11), 337-346 (2010).