# Roughness Index and Roughness Distance for Benchmarking Medical Segmentation

Vidhiwar Singh Rathour, Kashu Yamakazi and T. Hoang Ngan Le

*Department of Computer Science and Computer Engineering, University of Arkansas, Fayetteville, Arkansas 72701, U.S.A.*

Keywords:      Surface Analysis, Roughness Distance, Irregular Spikes/Holes, Medical Imaging, Medical Segmentation, Volumetric Segmentation.

Abstract:      Medical image segmentation is one of the most challenging tasks in medical image analysis and has been widely developed for many clinical applications. Most of the existing metrics have been first designed for natural images and then extended to medical images. While object surface plays an important role in medical segmentation and quantitative analysis i.e. analyze brain tumor surface, measure gray matter volume, most of the existing metrics are limited when it comes to analyzing the object surface, especially to tell about surface smoothness or roughness of a given volumetric object or to analyze the topological errors. In this paper, we first **analysis both pros and cons of all existing medical image segmentation metrics**, specially on volumetric data. We then propose an appropriate **roughness index and roughness distance** for medical image segmentation analysis and evaluation. Our proposed method addresses two kinds of segmentation errors, i.e. (i) **topological errors on boundary/surface** and (ii) **irregularities on the boundary/surface**. The contribution of this work is four-fold: (i) detect irregular spikes/holes on a surface, (ii) propose roughness index to measure surface roughness of a given object, (iii) propose a roughness distance to measure the distance of two boundaries/surfaces by utilizing the proposed roughness index and (iv) suggest an algorithm which helps to remove the irregular spikes/holes to smooth the surface. Our proposed roughness index and roughness distance are built upon the solid surface roughness parameter which has been successfully developed in the civil engineering.

## 1 INTRODUCTION

In this paper we first discuss the pros and cons of various metrics that have been commonly used for benchmarking the medical image segmentation task. We emphasize on the limitations of existing metrics, such as Hausdorff distance when evaluating the volumetric segmentation. Our study shows that the existing volumetric metrics are unable to measure the topological errors specially when irregular spikes/holes are on the surface. We then propose (i) an algorithm that helps to detect irregular spikes/holes that exist on a given object surface; (ii) a roughness index that describes how rough an object is given an object's surface; (iii) a roughness distance that aims at comparing the surfaces between two given objects; (iv) an algorithm that aims at removing the small outliers and the irregular spikes/holes to smooth the surface. As compared to other volumetric segmentation metrics i.e. Hausdorff distance, our proposed roughness distance is able to measure the topological error whereas roughness index evaluates the surface roughness. Fur-

thermore, we conduct the experiment to show that our proposed irregular spikes/holes detection and surface smoothing can be applied as a post-processing step in any image segmentation algorithm to improve the accuracy.

## 2 DESCRIPTION OF PURPOSE

Medical image segmentation is an important research topic in medical analysis and has attracted attention in past couple of years. With the abundance of medical data available it has become easier to perform segmentation task. However, evaluation and validation of medical segmentation, specially volumetric data is still a major concern because majority evaluation metrics have been developed as piece-wise setting for 2D natural images and then extended to medical images including volumetric data. As categorized in (Shi et al., 2013), there are four types of segmentation errors i.e. *quantitative or the number of objects, area of*

*segmentation, contour or the object boundary, and the presence of holes, or irregularities in the boundary of segmentation.* The first type of error, which regards the number of objects, can be mitigated by increasing the training data. Most of the common evaluation metrics (i.e. Dice score , Sensitivity, Specificity, etc) have focused to solve the second type of error, i.e. area of segmentation which is a well-known problem in any segmentation task in both computer vision and medical analysis. For the third type of error, i.e. object contour/boundary error, there are a limited number of metrics that have been developed. Hausdorff distance (HDD) and Average Symmetric Surface Difference (ASSD) (Gerig et al., 2001) are the ones that have been used for calculating errors on object surface. The last error, which is related to topological errors such as holes and spikes, still remains as a challenging problem in medical analysis. Several attempts such as (Joshi et al., 2007) (Li et al., 2006) (Wu and Chen, 2002) has focused on the last error category by considering the smoothness and roughness criteria. **In this work, we address the last two kinds of errors, i.e. (i) topological error on boundary/surface and (ii) irregularities on boundary/surface** as demonstrated in Fig:1.

Different from 2D objects, volumetric objects need the *consistency* and *continuous* between slides. A comparison between consistency-inconsistency and regularity-irregularity in volumetric data is given in Fig:1 where each slide is presented in a cuboid (one volumetric is considered as a set of slices) and $\zeta$ is the distance between the surface and center of gravity. The *inconsistency* or *irregularity* is defined as an abrupt or a sudden spike/hole. In Fig:1, the regular spike/hole is given in the top (Fig:1.a) where spike or hole is gradually formed from slice to slice whereas the irregular spike/hole is given in the bottom (Fig:1.b) where spike or hole suddenly appeared.

Different from the previous works (Joshi et al., 2007) (Li et al., 2006) (Wu and Chen, 2002) which use geometric graph i.e., minimum s-t cut, we make use of **solid surface roughness parameter in civil engineering to propose roughness metric** (Chang et al., 2006) (Tonietto et al., 2019) (Gadelmawla et al., 2002). Our contribution can be summarized as follows:

- Revise and analyze the existing segmentation metrics that have been used in medical analysis (Sec:3).

- Propose an algorithm which helps to detect all irregular spikes/holes on the object surface (Sec:4.1).

- Introduce a **roughness index** that measures the surface roughness given an object in (Sec:4.2).
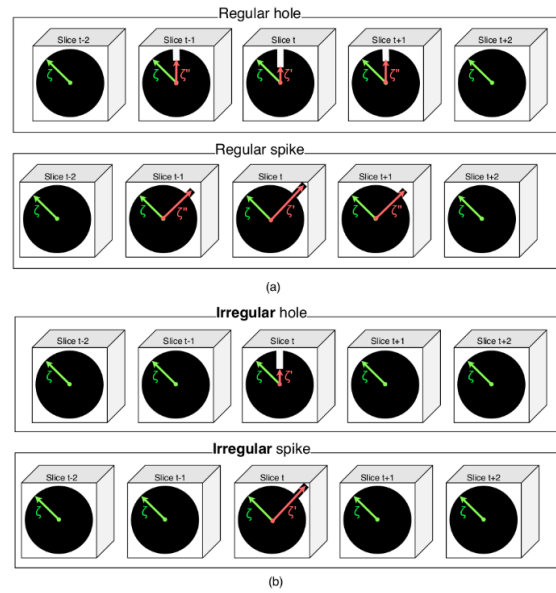


Figure 1: An illustration of a regular spike/hole (a) v.s an irregular spike/hole (b).

Our proposed roughness index is based on the solid surface roughness parameter that has been successfully developed in the civil engineering (Chang et al., 2006) (Tonietto et al., 2019) (Gadelmawla et al., 2002).

- Propose a **roughness distance** metrics which computes the surface distance between two surfaces (Sec:4.3).

- Propose an algorithm which helps to remove the irregular spikes/holes and to smooth the contour (Sec:4.4).

## 3 RELATED WORK

In this section, we will revise all existing segmentation metrics that have been commonly used in medical analysis. We first categorize the existing segmentation metrics into two groups, namely, region-based metrics and boundary-based metrics. We then analyse the pros and cons of each metric in the following subsections.

### 3.1 Region-based Metrics

By definition, region-based metrics are used to evaluate the area occupied by the segmentation. The region-based metrics, which are based on pixel-wise, have been first developed for spatial images (2D) segmentation in computer vision in general and then extended to volumetric (3D) segmentation in medical

imaging. These types of metrics tend to work well when there is clear demarcation with respect to data and when the contour is smooth. However they tend to fail when the the data has holes or boundary is irregular. These metrics tend to evaluate the second type of segmentation error, i.e. area of segmentation. The following is some common region based metrics that are popularly used volumetric segmentation.

**Segmentation Problem Setting:** In the image segmentation problem, evaluation process is performed between the ground-truth G created by the human and segmentation predicted P by some algorithmic model.

**Dice Similarity Coefficient (DSC):** Initially introduced as Dice (Dice, 1945) also known as the F1 score is one of the most commonly used metrics in validating medical image segmentation (Linguraru et al., 2012) (Linguraru et al., 2009) in both spatial images and volumetric data. Lets consider $P$ as the predicted volumetric segmentation vector and $G$ as the ground-truth, then DSC can be calculated as shown in Eq:1.

$$\text{DSC} = \frac{2|P \cap G|}{|P| + |G|} \tag{1}$$

**Symmetric Volume Difference (SVD):** introduced by (Campadelli et al., 2009) and **Jaccard Similarity Coefficient (JSC)** introduced by (Liu et al., 2012) are similar to DSC and can be mathematically computed from DSC as shown in Eq:2 and Eq:3.

$$\text{SVD} = 1 - DSC \tag{2}$$

$$\text{JSC} = \frac{|P \cap G|}{|P \cup G|} = \frac{DSC}{2 - DSC} \tag{3}$$

DSC although works well with data that is clearly demarcated, yet it tends to produce unwanted results if the segmentation boundary is ambiguous. Also DSC cannot tell anything about the boundary information, roughness and smoothness of a volumetric surface or the topological error on the boundary surface. JSC and SVD have the same inherent problems as DSC.

**Precision (Pre), Recall (Rec) and Sensitivity (Sens):** Precision is defined as the volume of correctly segmented volume to the total volume that has been segmented. Recall (also referred to as Sensitivity) is the the ratio of correctly segmented volume over the ground-truth.

$$\text{Pre} = \frac{|P \cap G|}{|P|} \tag{4}$$

$$\text{Rec/Sens} = \frac{|P \cap G|}{|G|} \tag{5}$$

Precision takes into account only the volume that has been segmented correctly but does not consider the under-segmented volume. Recall on the other hand does not consider the over-segmented volume. However these two metrics are extensively being used in computer vision for segregation tasks (Wolz et al., 2012) (Campadelli et al., 2010).

**Specificity (Spec):** Specificity also referred to as Selectivity is the ratio of portion of total volume that is not common to the ground-truth ($G$) and predicted segmentation ($P$) by the portion not included in ground-truth ($G$). True Negative (TN) is the portion of volume that is not common to the ground-truth ($G$) and predicted segmentation ($P$) and False Positive (FP) is the potion of volume belonging to predicted segmentation ($P$) that is not common to ground-truth ($G$):

$$\text{Specificity(Spec)} = \frac{|(P \cup G)^C|}{|G^C|} \tag{6}$$

Here $C$ denotes the compliment component which is illustrated in Fig:2. The segmented volumetric $\mathcal{S}$ contains two parts corresponding to foreground $\mathcal{F}$ and background $\mathcal{G}$, where $\mathcal{G} = \mathcal{F}^C$
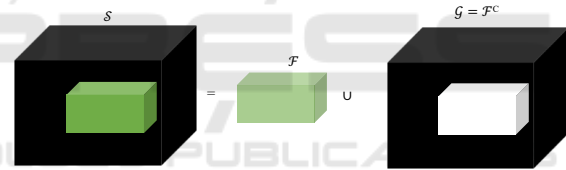


Figure 2: An Illustration of compliment using in Eq:6. Green cuboid $F$ represents the set for which compliment is being calculated, and Black cuboid $S$ represents the universal set of which $F$ is a part.

**Relative Volume Difference (RVD):** RVD is defined as the ratio of absolute difference in volume between the predicted volumetric segmentation vector ($P$) and the ground-truth ($G$) to the ground-truth ($G$). It is commonly used as a reference to other metrics(Heimann et al., 2009) (Linguraru et al., 2012).

$$\text{RVD} = |\frac{|P| - |G|}{|G|}| \tag{7}$$

RVD computed the relative difference in volume between predicted volumetric segmentation vector ($P$) and the ground-truth ($G$) and hence it does not take into consideration the overlap between them.

## 3.2 Boundary-based Metrics

Different from region-based metrics, which are designed to work on entire area, boundary-based metrics focus on boundary or surface only. In this section, we

revise two common boundary-based metrics, namely, Average Symmetric Surface Difference (ASSD) and Hausdorff Distance (HDD) as follows:

**Hausdorff Distance (HDD):** Hausdorff distance (HDD) is defined as the maximum possible distance from a point/voxel on one boundary/surface to the corresponding closest point/voxel on another boundary/surface (Gerig et al., 2001) (Chen et al., 2012b) (Liu et al., 2012) (Chen et al., 2012a). The HDD between the ground-truth boundary/surface $\partial G$ and the predicted segmentation boundary/surface $\partial P$ is defined as follows:

$$\text{HDD} = \max_{x \in \partial G}((|x, \partial P|_{L2})) \qquad (8)$$

where $|x, \partial P|_{L2}$ is the shortest $L_2$ distance between a point/voxel $x$ on the ground-truth boundary/surface $\partial G$ and the predicted segmentation boundary/surface $\partial P$, namely, $|x, \partial P|_{L2} = \min_{y \in \partial P}||(x - y)||^2$. Thus, Eq.9 is rewritten as:

$$\text{HDD} = \max_{x \in \partial G}((|x, \partial P|_{L2})) = \max_{x \in \partial G}((\min_{y \in \partial P}||x - y||^2)) \qquad (9)$$

Because both $P$ and $G$ are symmetric, the bidirectional Hausdorff distance between ground-truth boundary/surface $\partial G$ and the predicted segmentation boundary/surface $\partial P$ is computed as:

$$\text{HDD} = \max(\max_{x \in \partial G}((|x, \partial P|_{L2})), (\max_{x \in \partial P}((|y, \partial G|_{L2})) \qquad (10)$$

Hausdorff distance, which is computed as the maximum distance between two surface, has been commonly used in practice. HDD only tells about the maximum possible distance. However, it is unable to describe the surface roughness as well as detect topological errors which are critical problems in medical imaging. Fig:3 illustrates some limitations of HDD. In this figure, suppose the ground-truth boundary $G$ is presented in blue curve whereas the predicted segmentation $P$ is shown in red curve. Two cases are considered in this example, namely, smooth predicted segmentation (Fig:3(a)) and rough predicted segmentation (Fig:3(b)) with some topological errors on the predicted segmentation boundary. Let denote $D_1$ and $D_2$ as the distance between $G$ and $P$, i.e. $D_1 = \max_{x \in \partial G}((|x, \partial P|_{L2})$ and the distance between $P$ and $G$, i.e. $D_2 = \max_{x \in \partial P}((|y, \partial G|_{L2})$. As shown in Fig:3, the distance $D_1$ and $D_2$ are the same in two cases, thus the HDD is unchanged, i.e. $HDD = max(D_1, D_2)$ even the predicted boundary in Fig:3(b) is different from the one in Fig:3(a). Compared to the predicted boundary in Fig:3(a), the one in Fig:3(b) is rougher and with more topological changes.

Fig:4, 5 further explains the limitations of HDD. In this example, the ground-truth is given in Fig:4, 5
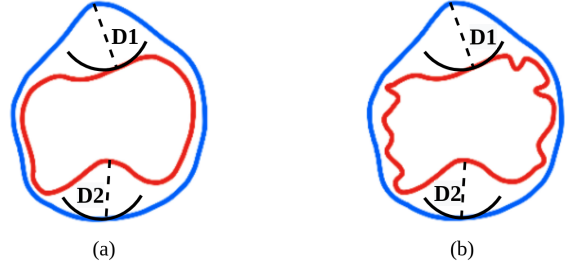


Figure 3: Illustration of HDD in two cases: smooth predicted boundary (a) and rough predicted boundary with topological changes (b). Blue curve is ground-truth boundary $G$ and red curve is predicted segmentation boundary $P$. $D_1$ is distance from $G$ to $P$ and $D_2$ is distance from $P$ to $G$.

(a) and Fig:4, 5 (b) and (c) are two different predicted segmenting results. As shown in Fig:4, 5, there are little irregular spikes on (b) and many irregular spikes on (c), however, the HDD between the ground-truth and the predicted segmenting results are the same.
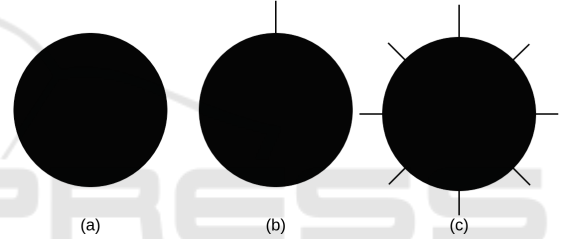


Figure 4: From left to right 2D (a): ground-truth; (b) predicted segmentation with little irregular spikes; (c) predicted segmentation with many irregular spikes.
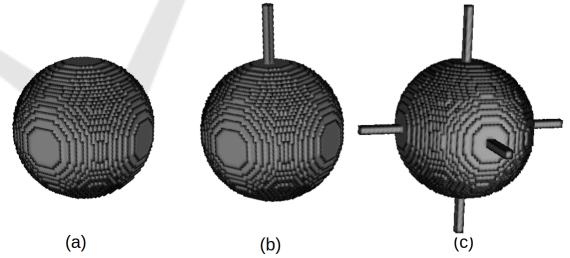


Figure 5: From left to right 3D (a): ground-truth; (b) predicted segmentation with little irregular spikes; (c) predicted segmentation with many irregular spikes.

**Average Symmetric Surface Difference (ASSD):** ASSD (Chen et al., 2012b) (Chen et al., 2012a) (Yokota et al., 2013) is the average of all the distances from points/voxels on the boundary/surface of the ground-truth mask to the boundary/surface of the predicted segmentation mask, and vice versa. Denote $P$ and $G$ as the predicted segmentation mask and the ground-truth mask. The boundary/surface of $P$ and $G$ are then defined as $\partial P$ and $\partial G$. Mathematically, ASSD is computed as follows:

Table 1: Summary of existing metrics on volumetric segmentation:- Red : Predicted Segmentation(P), Blue: ground-truth(G), Purple: True Positive(TP), 'C' in the subscript suggests the compliment of the vector the image description is shown in Fig 6.

| Type | Metrics | Equation | Visualization |
|---|---|---|---|
| Region-Based | DSC | $\frac{2\|P \cap G\|}{\|P\|+\|G\|}$ |  |
| | PREC | $\frac{\|P \cap G\|}{\|P\|}$ |  |
| | JSC | $\frac{\|P \cap G\|}{\|P \cup G\|} = \frac{DSC}{2-DSC}$ |  |
| | REC, SES | $\frac{\|P \cap G\|}{\|G\|}$ |  |
| | SET | $\frac{\|P \cup G\|^C}{\|G\|^C}$ |  |
| | RVD | $\left\| \frac{\|P\|-\|G\|}{\|G\|} \right\|$ |  |
| Contour Based | HDD | $\max\left(\max_{x\in\partial G}((\|x,\partial P\|_{L2})), (\max_{x\in\partial P}((\|y,\partial G\|_{L2}))\right)$ |  |
| | ASSD | $\frac{\sum_{x\in\partial G}(\|x,\partial P\|_{L2})+\sum_{x\in\partial P}(\|x,\partial G\|_{L2})}{\|\partial G\|+\|\partial P\|}$ | |

$$\text{ASSD} = \frac{\sum_{x\in\partial G}(|x,\partial P|_{L2}) + \sum_{x\in\partial P}(|x,\partial G|_{L2})}{|\partial G| + |\partial P|} \quad (11)$$

ASSD is a good metrics for cross distance computation between boundaries of two surfaces however ASSD has the same limitations as HDD that it cannot compute the roughness or smoothness on one particular surface.

The existing metrics can be summarized in Table I where the visualization is further explain in Fig.6.

# 4 PROPOSED METRICS

In this section, our proposed metrics for surface roughness analysis in medical segmentation will be
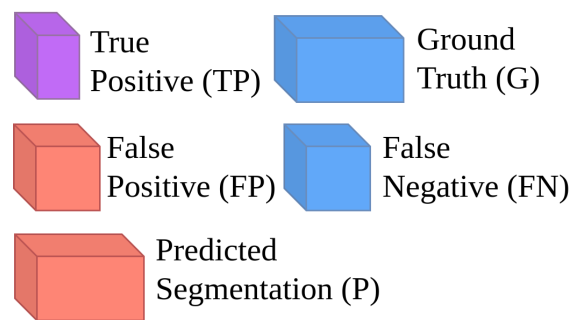


Figure 6: Explanation of annotations & visualization that are used in table 1.

detailed. Our proposed roughness index and roughness distance is based on the real world average roughness parameter(Tonietto et al., 2019) as de-

scribed in Sec:4.1. In civil engineering domain, roughness parameter of a particular surface is calculated using a laser to map the irregularities on the surface (Tonietto et al., 2019). All the symbols and notations used to describe the proposed metrics have been summarized in Table:2.

## 4.1 Irregular Spike/Hole Detection

Roughness is a very important parameter that is frequently used in civil engineering domain (Chang et al., 2006) (Tonietto et al., 2019) (Gadelmawla et al., 2002). Civil engineers use the roughness parameter to measure the inconsistencies on a particular surface such as a slab of concrete or metal. A surface profile gauge or a Digital Holographic Microscope is used to map the fluctuations on the surface (Tonietto et al., 2019). The roughness parameter(Tonietto et al., 2019) in civil engineering domain is defined in Eq:12 and illustrated in Fig:9a , where $\zeta_i$ is the perpendicular distance of a point from the laser plane also referred to as the height coordinate (Tonietto et al., 2019) is calculated using a laser moving on a fixed plane parallel to the object surface and N is the total number of points where height coordinate is calculated.

$$RoughnessParameter = \frac{1}{N} \sum_i^N |\zeta_i| \qquad (12)$$

We extended the term height coordinate to use it in 2D and 3D domain by calculating the distance of the surface point from the center of gravity $C_0$ instead of a plain, as illustrated in Fig: 9c for a closed contour laser plain can be approximated as the center of gravity. We have defined $\zeta$ (Zeta) as the distance of a surface point for a contour $P_{Surface}$ from center of gravity $C_0$ as shown in Eq:13. Here $P_{Surface}$ is the matrix that has value 1 or 0 based on whether the location in the segmentation mask P belongs to the surface or not respectively.

$$\zeta_{ijk} = \begin{cases} |(i,j,k),C_0|_{L2} & P_{Surface}(i,j,k) = 1 \\ 0 & \text{Otherwise} \end{cases} \qquad (13)$$

For roughness in 2D and 3D we use a **Distance Matrix** $\zeta_m$ that contains the distance of each corresponding surface point from the center of gravity $C_0$ as shown in Eq:14. This matrix can be used to detect and correct surface roughness. The main purpose of calculating $\zeta$ is to track the variations in surface. As illustrated in Fig:1 an irregular hole/spike is marked by an abrupt change in $\zeta$ while for a regular hole/spike change in $\zeta$ takes place gradually.

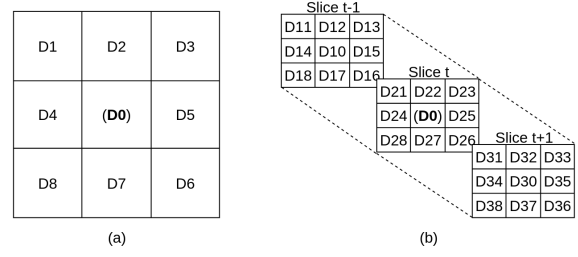$$\zeta_m(i,j,k) = \zeta_{ijk} \qquad (14)$$



Figure 7: Illustration of how neighbors of a reference point D0 are considered for 2D (a) and 3D (b) Distance matrix $\zeta_m$ . In the given figure D0 is a position in distance matrix $\zeta_m$, that belongs to the contour and for which roughness $\Delta\zeta$ needs to be calculated.

To detect roughness we define **Roughness Matrix** $\Delta\zeta_m$ containing roughness value $\Delta\zeta$ (Delta zeta) for each surface location as shown in Eq:17. Roughness $\Delta\zeta$ of a location on surface can be defined as the sum of differences between $\zeta$ and its contour neighbors $\zeta_{Neighbors}$, belonging to the set of neighbors $S_{\zeta_{Neighbors}}$ illustrated in Fig:7 and described in Eq:15 and Eq:16.

$$\Delta\zeta_{ijk} = \sum(\zeta_{ijk} - \zeta_{Neighbors}) \qquad (15)$$

$$\zeta_{Neighbors} \in S_{\zeta_{Neighbors}} \qquad (16)$$

$$\Delta\zeta_m(i,j,k) = \Delta\zeta_{ijk} \qquad (17)$$

lets consider a 2D example for various cases of roughness as shown in Fig:8.

- Case 1: shows the condition of a plain where the neighbors are at the same distance from $C_0$ as the point for which $\Delta\zeta$ needs to be calculated, so $\Delta\zeta$ will be $((D-D)+(D-D)) = 0$

- Case 2: shows a slope where $\Delta\zeta$ will be $((D-D)+(D-D)+(D-(D+1))+(D-(D-1))) = 0$.

- Case 3: is an example of hole where $\Delta\zeta$ will be $((D-D)+(D-D)+(D-(D+1))+(D-(D+1))) = 2$

- Case 4: which is a spike where $\Delta\zeta$ will be $((D-D)+(D-D)+(D-(D+1))) = 1$

Hence it can be easily concluded that $|\Delta\zeta|$ for a location close to zero will denote a smooth surface and greater then zero will refer to rough surface.

## 4.2 Roughness Metrics

Roughness parameter is a term usually used to determine the roughness of a solid surfaces(Chang et al., 2006) (Tonietto et al., 2019) (Gadelmawla et al., 2002). We extended this term to use in 2D and 3D surface vector domain. As shown in Eq:18 The **Roughness Index** (RI) in 3D can be calculated by dividing

Table 2: Symbols along with their descriptions.

| Symbol | Description | Symbol | Description |
|---|---|---|---|
| P | Predicted segmentation mask | G | ground-truth segmentation mask |
| $C_0$ | Center of gravity of a contour | $\partial S^w$ | Segment of an array S with a fixed window size w |
| $\zeta$ $\zeta_i$ $\zeta_{ij}$ $\zeta_{ijk}$ | Distance of a contour position (i,j,k) from $C_0$ | $\zeta_m$ | Distance matrix, $\zeta_m(i,j,k) = \zeta_{ijk}$ |
| $\zeta_{Neighbor}$ | Distance $\zeta$ of neighbor | $S_{\zeta_{Neighbor}}$ | set of neighboring $\zeta_{Neighbor}$ of $\zeta$ |
| $\Delta\zeta$ $\Delta\zeta_i$ $\Delta\zeta_{ij}$ $\Delta\zeta_{ijk}$ | Roughness at a matrix position (i,j,k) | $\hat{\zeta}$ | Difference between $\zeta$ for P and G |
| $\Delta\zeta_m$ | matrix of $\Delta\zeta$ | $\Delta\zeta_{Bm}$ | Rough Boolean matrix, where $\Delta\zeta_{Bm} \in (0,1)$ |
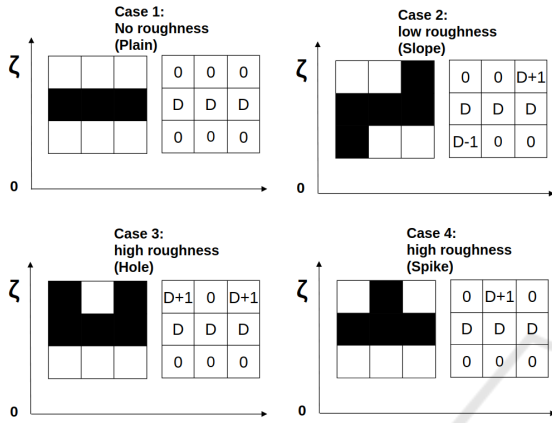


Figure 8: Different cases of roughness when dealing with 2D segmentation mask, In the figure $\zeta$ is the corresponding distance of contour point from center of gravity $C_0$.

the segmentation surface S into small surface element $\partial S^w$ of a fixed window size w, and then calculating the average deviation of $\zeta$ from the mean $\zeta_{Mean}$ for all surface voxels in the surface element $\partial S^w$ as illustrated in Fig:9b. In Eq:18 $\partial S_i^w$ denotes a point on the surface element $\partial S^w$, M is the total no of surface elements $\partial S^w$ that the contour surface is divided into and N is the total no of points $i$ inside each surface element that belong to the contour surface, Here $|\partial S_i^w, C_0|_{L2}$ is equal to $\zeta$ that we calculated in previous section.

$$RI = \frac{1}{M} \sum_{\partial S^w \in S}^{M} \frac{1}{N} \sum_i^N |(|\partial S_i^w, C_0|_{L2}) - Mean(|\partial S_i^w, C_0|_{L2})|$$
(18)

Roughness is a relative quantity. An object that is rough as compared to one surface may be smooth as compared to other. Hence it can be difficult to tell about the roughness of a surface unless we have a baseline to compare the roughness index. Hence we introduce **Roughness Ratio** (RR) that tells about the relative difference between roughness of two objects. The roughness ratio has been defined in Eq: 19 where $\mathcal{R}I_P$ and $\mathcal{R}I_G$ are the roughness index of predicted segmentation and ground-truth respectively.

$$RR = \frac{|RI_P - RI_G|}{RI_G}$$
(19)
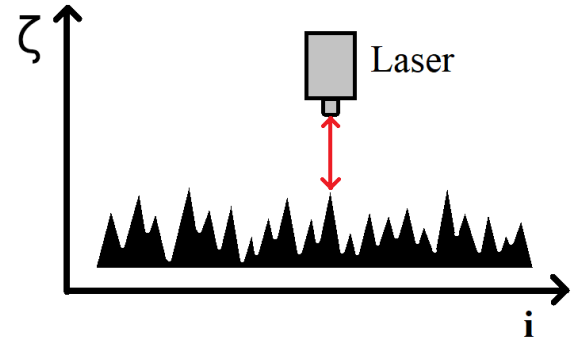


(a) In civil engineering domain Roughness is calculated by moving a laser parallel to the surface to find height coordinate $\zeta$ and then using it to find the roughness parameter of the surface using Eq:12.
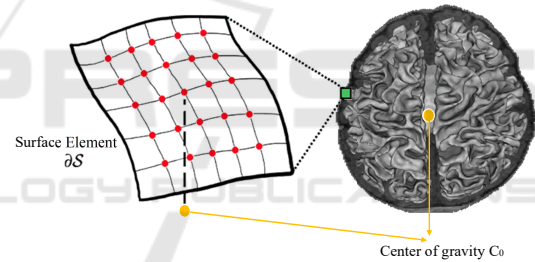


(b) In medical domain the segmentation mask is a closed object as opposed to a flat surface in civil engineering, hence we calculate the distance of object surface $\zeta$ from center of gravity instead of a plane. This $\zeta$ is used to calculate roughness index and surface/roughness distance.
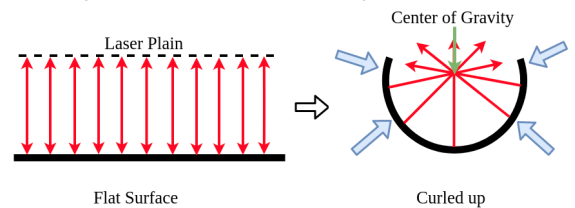


(c) Laser plane can be approximated to center of gravity if the surface is rolled into a closed contour.

Figure 9: Illustration of how to compute roughness (a): Roughness parameter is calculated by moving laser on fixed plane paralleled to the surface, (b): Roughness index and surface/roughness distance is calculate from an origin which is defined as the center of medical image and (c): How the method of calculating roughness can be extended to medical imaging domain.

## 4.3 Roughness Distance

In this section, we propose **Roughness Distance** which is considered as surface distance between two surfaces. Let denote $\hat{\zeta}$ as the difference between the $\zeta$ for predicted segmentation ($\zeta_P$) and ground-truth segmentation ($\zeta_G$) as shown in Eq:20 and **Roughness Distance Matrix** $\hat{\zeta}_m$ as the matrix containing $\hat{\zeta}$ values as shown in equation Eq:21

$$\hat{\zeta}_{ijk} = \zeta_{P_{ijk}} - \zeta_{G_{ijk}} \tag{20}$$

$$\hat{\zeta}_m(i,j,k) = \hat{\zeta}_{ijk} \tag{21}$$

Simply speaking roughness distance matrix $\hat{\zeta}_m$ can be calculated by subtracting distance matrix for ground-truth segmentation $\zeta_{mG}$ from distance matrix for predicted segmentation $\zeta_{mP}$ as shown in equation Eq:22. Roughness distance can be used to calculate the roughness change between two object, ground-truth and predicted segmentation in our case.

$$\hat{\zeta}_m = \zeta_{mP} - \zeta_{mG} \tag{22}$$

We also propose **Average Roughness Distance** (ARD) which as the name suggest is the average surface/roughness distance between two objects as shown in Eq:23. ARD is a metric that tells us about the average difference between the surface of two objects. ARD can be used as a substitute of HDD to compare roughness.

$$\text{ARD} = Mean(|\hat{\zeta}_m|) \tag{23}$$

## 4.4 Surface Smoothing

In this section we will propose a method for smoothing a contour that has roughness on its surface. Smooth contour can be obtained by using either of the two methods which include using roughness matrix $\Delta\zeta_m$ or roughness distance matrix $\hat{\zeta}_m$ that were calculated in previous sections.

For contour smoothing rough boolean matrix $\Delta\zeta_{Bm}$ is used where the value can be one or zero based on whether the position is considered as rough or smooth respectively as shown in Eq:24 where $\kappa$ is the threshold roughness value in range of $(0, \max(\Delta\zeta_m))$.

$$\Delta\zeta_{Bm(i,j,k)} = \begin{cases} 1, & |\Delta\zeta_{ijk}| > \kappa \\ 0, & \text{Otherwise} \end{cases} \tag{24}$$

Similarly $\Delta\zeta_{Bm}$ can also be computed using the roughness distance matrix $\hat{\zeta}_m$ as shown in Eq:25 where $\kappa_c$ is the threshold distance in range of $(0, \max(\hat{\zeta}_m))$.

$$\Delta\zeta_{Bm(i,j,k)} = \begin{cases} 1, & |\hat{\zeta}_{ijk}| > \kappa_c \\ 0, & \text{Otherwise} \end{cases} \tag{25}$$

This rough boolean matrix $\Delta\zeta_{Bm}$ can be used for contour smoothing as shown in Eq: 26 where $P_{Rough}$ is the segmentation mask before contour smoothing and $P_{Smooth}$ is the one after smoothing. It is important to note here that both methods return a smooth contour, However while $\Delta\zeta_m$ requires only the rough segmentation contour, $\hat{\zeta}_m$ also required the corresponding ground-truth segmentation contour.

$$P_{Smooth} = |P_{Rough} - \Delta\zeta_{Bm}| \tag{26}$$

---

**Algorithm 1:** Calculate Roughness index(RI) of a 3D contour.

**Require:** $S \in \{0,1\} \vee S[x,y,z]$
  $(X_0, Y_0, Z_0) \leftarrow 0$
  $N \leftarrow 0$
  $RI \leftarrow 0$
  **for** $(X,Y,Z) \in S$ **do**
    $(X_0, Y_0, Z_0) \leftarrow (X_0, Y_0, Z_0) + (X,Y,Z)$
    $N \leftarrow N + 1$
  **end for**
  $(X_0, Y_0, Z_0) \leftarrow (X_0, Y_0, Z_0)/N$
  $M \leftarrow 0$
  **for** $\partial S^w \in S$ **do**
    $C \leftarrow 0$
    **for** $(X,Y,Z) \in \partial S^w$ **do**
      $D_0 \leftarrow D_0 + |(X,Y,Z),(X_0,Y_0,Z_0)|_{L2}$
      $C \leftarrow C + 1$
    **end for**
    $D_0 \leftarrow D_0/C$
    $N \leftarrow 0$
    $R_s \leftarrow 0$
    **for** $(X,Y,Z) \in \partial S^w$ **do**
      $R_s \leftarrow R_s + |(|(X,Y,Z),(X_0,Y_0,Z_0)|_{L2}) - D_0|$
      $N \leftarrow N + 1$
    **end for**
    $R_s \leftarrow R_s/N$
    $M \leftarrow M + 1$
    $RI \leftarrow RI + R_s$
  **end for**
  $RI \leftarrow RI/M$

---

# 5 DISCUSSION & EXPERIMENTATION

## 5.1 Results and Comparison

The Table 1 summarizes all the metrics that are currently being used for evaluation of 3D medical im-

ages. As stated earlier the region based metrics cannot calculate the roughness and smoothness of a 3D contour. Also Hausdorff distance is capable of finding the maximum distance between the ground-truth and predicted but it fails to capture the small roughness on the surface.

In our experiments we used 2D and 3D segmentation images of size ($100 \times 100$) and ($100 \times 100 \times 100$) as shown in Fig:4 and Fig:5 respectively. In both cases (a) is a smooth segmentation treated as ground-truth, (b) is a segmentation with small spike of 20 pixel length and (c) has many spikes, where the top spike has the same length of 20 as (b) and all other spikes are of smaller length for both 2D and 3D example.

Mathematically the roughness index (*RI*) for a circle and sphere should be 0, but because for images the coordinate system is integral not continuous, even a smooth circle has a small *RI* greater then 0. Hence, we treat it as Residual Roughness Index $RI_{Residual}$. In our experiments we have treated the RI for ground-truth segmentation as residual. The absolute roughness index $RI_{Absolute}$ can be calculated by subtracting $RI_{Residual}$ from *RI* as shown in Eq:27 .

$$\text{RI}_{Absolute} = RI - RI_{Residual} \qquad (27)$$

In our experiments we varied the window size and calculated roughness index for various window sizes. For 2D image we first performed 2D convolution on the image using a kernel and dilation operator to convert the 2D image into a contour. We then used a 2D window and moved it on the image and calculated the variation of distance from each boundary location to the center of gravity of the contour using mean distance for all boundary locations present in the window. We used the strides equal to the window size so that each location is used to calculate roughness index exactly once. The plot of *RI* and *RR* vs window size for images in Fig:4 is shown in Fig:10 and Fig:11 respectively. Similarly for 3D images we first performed 3D convolution on the image using a kernel and dilation operator to convert the 3D image into a contour. We then used a 3D window to calculate the *RI* of each image in Fig:5. The *RI* and *RR* vs window size graph has been shown in Fig:12 and Fig:13 respectively.

It can be inferred from Fig:10 and Fig:11 that window size plays a very important role in RI calculation. In our experimentation we found that the optimal window size must be between 3% to 10% of the image smallest dimension. It is also important to note that Roughness index is a standalone metrics but it can be used to compare the roughness of two image through roughness ratio that we have used in our experiments.
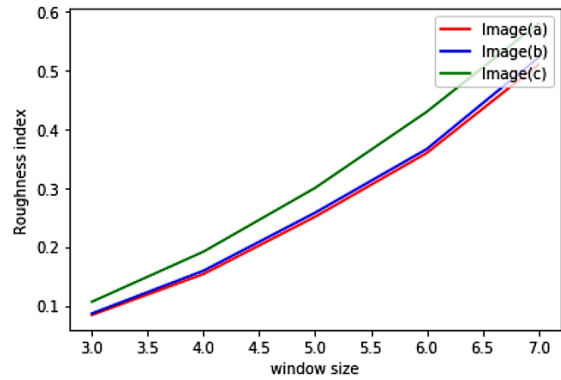


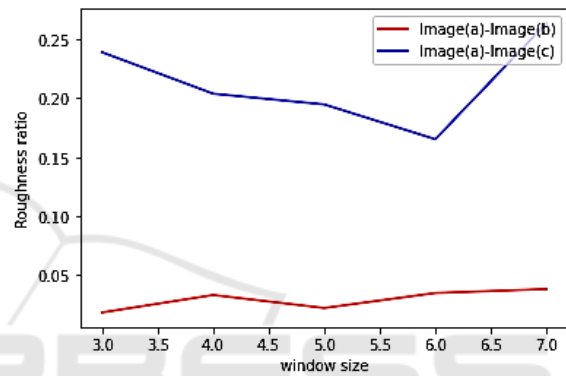Figure 10: Graph of Roughness index Vs Window size for 2D images in Fig:4(a), (b) and (c).



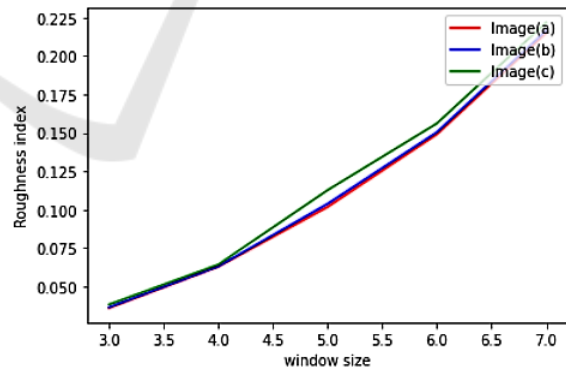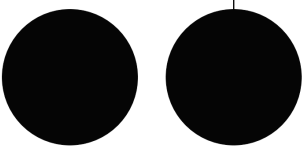Figure 11: Plot of Roughness ratio Vs window size for Fig:4(a)(ground-truth), Fig:4(b)(little roughness predicted segmentation) and Fig:4(a)(ground-truth segmentation), Fig:4(c)(high roughness predicted segmentation).


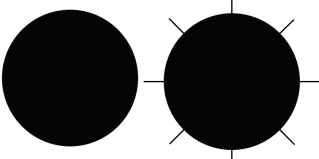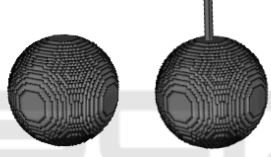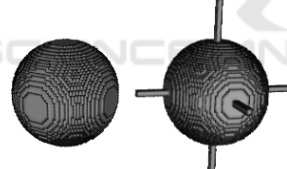
Figure 12: Graph of Roughness index Vs Window size for 3D images in Fig:5(a), (b) and (c).

A tabular comparison between roughness ratio (*RR*), average roughness distance (*ARD*) and Hausdorff distance (*HDD*) has been shown in Table 3. It can be easily inferred from the table that *RR* and *ARD* were capable of finding the difference in roughness for the two image pairs that *HDD* failed to do.

For contour smoothing we used the algorithms discussed in Sec:4.1 and Sec:4.3 to smooth the con-

Table 3: Comparison of Roughness Index RI, Roughness Ratio RR and Average Roughness Distance ARD with respect to Hausdorff Distance, The window size considered for RI calculation is 7% of image size i.e. 7. From top to bottom Fig:4(a)(2D ground-truth segmentation) and Fig:4(b)(2D little roughness predicted segmentation), Fig:4(a)(2D ground-truth segmentation) and Fig:4(c)(2D high roughness predicted segmentation), Fig:5(a)(3D ground-truth segmentation) and Fig:5(b)(3D little roughness predicted segmentation), Fig:5(a)(3D ground-truth segmentation) and Fig:5(c)(3D high roughness predicted segmentation).

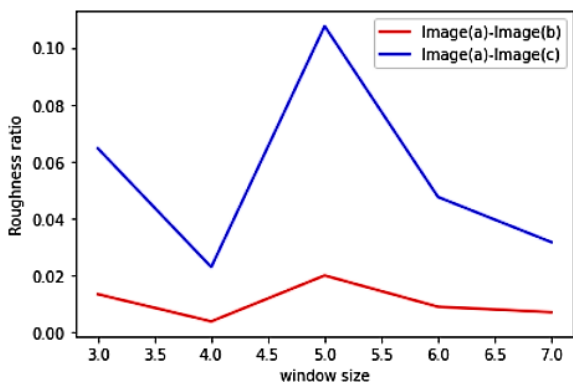| Images | Absolute Roughness Index | Roughness Ratio | Average Roughness Distance | Hausdorff Distance |
|---|---|---|---|---|
|  | 0 and 0.0120 | 0.0235 | 0.3178 | 20 |
|  | 0 and 0.0703 | 0.1377 | 0.7736 | 20 |
|  | 0 and 0.0015 | 0.0070 | 0.0592 | 20 |
|  | 0 and 0.0068 | 0.0317 | 0.0692 | 20 |



Figure 13: Plot of Roughness ratio Vs window size for Fig:5(a)(ground-truth), Fig:5(b)(little roughness predicted segmentation) and Fig:5(a)(ground-truth segmentation), Fig:5(c)(high roughness predicted segmentation).

tours as shown in Fig:14. It is clear from Fig:14 that roughness distance method produces a more satisfying result as compared to roughness matrix method, The reason being that roughness distance uses ground-truth segmentation as reference. However this is also a drawback because roughness distance method is constraint by the need of a reference distance matrix. Furthermore roughness distance method will produce unsatisfactory results if the center of gravity for P and G are not same, i.e. the segmentation masks are not aligned. However this problem can be overcome by using the center of gravity of segmentation mask G for both P and G.

Roughness matrix method is a robust method for detecting and removing surface roughness. For roughness calculation we considered a window size of three which includes a total of eight neighbors for 2D and twenty six neighbors for 3D as shown in Fig:7.

| Ground Truth | Rough Predicted Segmentation | After contour smoothing using roughness matrix | After contour smoothing using roughness distance matrix |
|:---:|:---:|:---:|:---:|
| 0 | 0.1387 | 0.0302 | 0 |
| 0 | 0.1847 | 0.0417 | 0 |

Figure 14: Contour smoothing for two type of segmentation, Circle (top) and star (bottom). Smoothing has been performed using the methods discussed in sections Sec:4.1 and Sec:4.3 and the Absolute roughness index $RI_{Absolute}$ is specified below each image. The window size for RI calculation is 7% of image size i.e. 7 and roughness threshold $\kappa$ , $\kappa_c$ for the given experiments was taken as 0.

this method is capable for detecting irregular spikes of width 1 pixel, However this method can be extended to detect spikes of multiple pixels by increasing the window size and the number of neighbors in neighbors set $S_{\zeta_{Neighbors}}$. The method for smoothing holes is same but in that case we will add a surface point to the surface instead of removing it in case of a spike.

# 6 CONCLUSION

In this paper we first discussed the pros and cons of various metrics that have been commonly used for the medical image segmentation task. We emphasize more on the limitations of existing metrics for volumetric segmentation. We then proposed (i) an algorithm that helps to detect all irregular spikes/holes that exist in the object surface; (ii) a roughness metric that describes how rough of a given object; (iii) a roughness distance that aims at comparing the surfaces between two given objects; (iv) an algorithm that aims at removing irregular spikes/holes to smooth the surface. Compare to other volumetric segmentation metrics i.e. Hausdorff distance, our proposed roughness distance is able to measure the topological error whereas roughness metric present the surface roughness. Furthermore, our proposed irregular spikes/holes detection and surface smoothing can be applied as a post-processing step in any image segmentation algorithm to improve the accuracy.

## REFERENCES

Campadelli, P., Casiraghi, E., and Esposito, A. (2009). Liver segmentation from computed tomography scans: a survey and a new algorithm. *Artificial intelligence in medicine*, 45(2-3):185–196.

Campadelli, P., Casiraghi, E., and Pratissoli, S. (2010). A segmentation framework for abdominal organs from ct scans. *Artificial Intelligence in Medicine*, 50(1):3–11.

Chang, J.-R., Chang, K.-T., and Chen, D.-H. (2006). Application of 3d laser scanning on measuring pavement roughness. *Journal of Testing and Evaluation*, 34(2):83–91.

Chen, X., Udupa, J. K., Bagci, U., Zhuge, Y., and Yao, J. (2012a). Medical image segmentation by combining graph cuts and oriented active appearance models. *IEEE TIP*, 21(4):2035–2046.

Chen, Y., Wang, Z., Hu, J., Zhao, W., and Wu, Q. (2012b). The domain knowledge based graph-cut model for liver ct segmentation. *Biomedical Signal Processing and Control*, 7(6):591–598.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.

Gadelmawla, E., Koura, M., Maksoud, T., Elewa, I., and Soliman, H. (2002). Roughness parameters. *Journal of materials processing Technology*, 123(1):133–145.

Gerig, G., Jomier, M., and Chakos, M. (2001). Valmet: A new validation tool for assessing and improving 3d object segmentation. In *MICCAI*, pages 516–523. Springer.

Heimann, T., Van Ginneken, B., Styner, M. A., Arzhaeva, Y., Aurich, V., et al. (2009). Comparison and evaluation of methods for liver segmentation from ct datasets. *TIP*, 28(8):1251–1265.

Joshi, A. A., Shattuck, D. W., Thompson, P. M., and Leahy, R. M. (2007). Surface-constrained volumetric brain registration using harmonic mappings. *IEEE Trans Med Imaging*, 26(12):1657–1669.

Li, K., Wu, X., Chen, D. Z., and Sonka, M. (2006). Optimal surface segmentation in volumetric images-a graph-theoretic approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):119–134.

Linguraru, M. G., Pura, J. A., Pamulapati, V., and Summers, R. M. (2012). Statistical 4d graphs for multi-organ abdominal segmentation from multiphase ct. *Medical image analysis*, 16(4):904–914.

Linguraru, M. G., Yao, J., Gautam, R., Peterson, J., Li, Z., Linehan, W. M., and Summers, R. M. (2009). Renal tumor quantification and classification in contrast-enhanced abdominal ct. *PR*, 42(6):1149–1161.

Liu, Y., Cheng, H.-D., Huang, J., Zhang, Y., and Tang, X. (2012). An effective approach of lesion segmentation within the breast ultrasound image based on the cellular automata principle. *Journal of digital imaging*, 25(5):580–590.

Shi, R., Ngan, K. N., and Li, S. (2013). The objective evaluation of image object segmentation quality. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, number 3, pages 470–479. Springer.

Tonietto, L., Gonzaga, L., Veronez, M. R., de Souza Kazmierczak, C., Arnold, D. C. M., and da Costa, C. A. (2019). New method for evaluating surface roughness parameters acquired by laser scanning. *Scientific reports*, 9(1):1–16.

Wolz, R., Chu, C., Misawa, K., Mori, K., and Rueckert, D. (2012). Multi-organ abdominal ct segmentation using hierarchically weighted subject-specific atlases. In *MICCAI*, pages 10–17. Springer.

Wu, X. and Chen, D. Z. (2002). Optimal net surface problems with applications. In *Automata, Languages and Programming*, pages 1029–1042.

Yokota, F., Okada, T., Takao, M., Sugano, N., Tada, Y., Tomiyama, N., and Sato, Y. (2013). Automated ct segmentation of diseased hip using hierarchical and conditional statistical shape models. In *MICCAI*, pages 190–197. Springer.