Segmentation Edit Distance

Daniel Pucher and Walter G. Kropatsch

Pattern Recognition and Image Processing Group (PRIP), TU Wien

Austria

Abstract—In this paper, we present a novel distance metric called Segmentation Edit Distance (SED) and its use as a segmentation evaluation metric. In segmentation evaluation, the difference or distance of a test segmentation and the associated ground truth segmentation are measured in order to compare different algorithms. Our proposed edit distance extends the idea of other edit distances such as the string edit distance or the graph edit distance to the domain of image segmentations. The distance is based on the cost of edit operations that are needed to transform one segmentation into another. Only one edit operation, the deletion of an error region, is considered. Different to other edit distances, the costs assigned to this operation are based on properties of the error regions and the image processing method used to delete a region. As a segmentation evaluation metric, it combines the assessment of accuracy and efficiency into a single metric. Evaluations on synthetic and real world data show promising results compared to other state of the art segmentation evaluation metrics.

I. INTRODUCTION

Image segmentation is the process of partitioning a digital image into sets of pixels or segments which share certain meaningful characteristics. It is an important process in image analysis which has resulted in extensive research and various proposed methods. Since this aggravates the selection of an appropriate algorithm for a given task, the importance of methods for evaluating the quality of image segmentation algorithms has increased. In segmentation evaluation, the task is to compare a given test segmentation to a ground truth segmentation by measuring the distance or similarity between them. Heckel et al. [1] categorize segmentation algorithms by the degree of automation in the following way:

1) *Fully automatic methods*, requiring no user interaction.
2) *Semiautomatic methods*, where the algorithm is initialized or parameterized by the user.
3) *Interactive methods*, where the user steers and corrects computer-generated segmentation results.
4) *Manuel tools*, where the segmentation is painted by hand, for example.

Fully automatic methods are the ultimate goal in image segmentation, since they minimize the user effort and provide reproducable results [1] and can be objectively evaluated through benchmarks[1]. If an automatic segmentation algorithm is not available or has failed, interactive methods can be used instead. Especially if accurate semantic objects are required. In order to assess the quality of interactive segmentation algorithms they are typically evaluated in terms of user experiments which are effective but time-consuming and labor

[1] http://homepages.inf.ed.ac.uk/rbf/CVonline/Imagedbase.htm#segmentation

intensive [2]. Consequently, methods to automate the evaluation have been researched and successfully applied in the past.

Heckel et al. [1] evaluated segmentation editing tools in the context of tumor segmentation in computer tomography. McGuinness and O'Connor [3] proposed a method to automate the evaluation of interactive segmentation algorithms with the focus on natural images. In order to achieve this, they drive the segmentation by automatically generating the user interactions from the error between the current segmentation and the ground truth. Moschidis et al. [4] presented an evaluation framework to assess the quality of interactive segmentation algorithms in the field of medical image analysis.

When using interactive segmentation algorithms, the user is required to provide input to the algorithm in the form of edit operations such as seed pixels [1], [3], [4]. In order to recieve a comprehensive evaluation, *accuracy*, *efficiency* and *repeatability* are criteria that are commonly used in the context of interactive segmenatation algorithms [1]. *Accuracy* is the degree of resemblance between a segmentation and the ground truth. It is measured using evaluation metrics such as the Jaccard index [3]. *Efficiency* is the amount of effort required in order to perform the segmentation and is related to the edit operations. Heckel et al. [1] state that the goal of an interactive segmentation algorithm is to finish a given segmentation with as few steps as possible, where one step can be seen as a user interaction or edit operation. It is measured in terms of run time or number of edit operations. Consequently, statements about the efficiency can only be made after the whole evaluation has been performed. *Repeatablity* expresses how well the same result for a specific segmentation task can be repeated over different sessions.

We want to focus on the first two criteria and propose a novel distance metric that combines the assessment of accuracy and efficiency into a single metric, the Segmentation Edit Distance (SED).

The remainder of this paper is organized as follows. Section II presents a short literature review of current state of the art segmentation evaluation metrics. Section III presents our proposed distance metric. Section IV discusses the segmentation datasets used to evaluate the new combined descriptor and studies the results of the evalution. Conclusions are given in Section V.

II. STATE OF THE ART

Taha and Hanbury [5] give an overview of 20 different evaluation metrics for volume segmentation. All these metrics can be expressed in one of two terms: The four basic overlap

cardinalities, namely True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) that reflect the overlap between two segmentations. Or the spatial position of voxels. Since the overlap cardinalities only take the number of pixels in different overlap regions into account they do not consider any spatial information like shapes of FP or FN regions. For the test segmentations in Figure 3 metrics based on these cardinalities give the same score to every shape, as the number of false positives is the same for every shape. Spatial distance based metrics take into account the spatial postion of false negatives and false positives. Taha and Hanbury present three distance metrics, namely the Hausdorff distance, the Average distance and the Mahalanobis distance [5]. The Hausdorff distance is sensitive to outliers. The Average distance, which is the Hausdorff distance averaged over all points, is stable and less sensitive to outliers but still effected by them. The Mahalanobis distance measures the distance between two segmentations by comparing estimates of them, thereby considering only the general shape and alignment and ignoring the boundary details.

Peng et al. [6] and Dung and Binh [7] used metrics to evaluate the quality of image segmentation that were not reviewed by Taha and Hanbury [5], but they have the same properties.

Deng et al. [8] proposed a new segmentation evaluation method for tumor segmentation in CT images, where they simulate the subjective evaluation of radiologists. Compared to subjective assessments, objective metrics are easy to compute but may not provide a good evaluation results. This is due to the fact that each metric only captures certain aspects of the difference between a segmentation and the ground truth. Therefore, their method uses a set of objective metrics which are combined in a linear fashion to form a new composite metric. With this new metric they achieve better results for evaluating segmentations than with the individual objective metrics. In order to construct the composite metric, the following three objective metrics were chosen based on the highest correlation with the subjective rating of radilogists: Volume overlap, absolute value of normalized volume difference and RMS surface distance (see [8] for definitions). The first two of these metrics are based on the basic overlap cardinalities and the last metric is based on the contour resulting in the same properties as the evaluation metrics reviewd by Taha and Hanbury [5].

Funke et al. [9] proposed an error measure for segmentation evaluation where small errors around the boundary are considered as tolerable and larger errors have to be corrected manually. The manual correction is reflected as a minimal weighted sum of split and merge operations that are needed to transform one segmentation into another. Since only splits and merges between regions are counted and no spatial information is considered, this error measure assigns the same value for all test segmentations in Figure 3.

III. Proposed Edit Distance

The SED measures the distance between a test segmentation and the corresponding ground truth segmentation based on the cost of edit operations. Edit operations correspond to actions that have to be performed in order to fix error regions in the test segmentation to transform it into the corresponding ground truth segmentation.

The concept of edit distances is not new and has originally been proposed for string sequences [10]–[12]. The string edit distance defines a dissimilarity measure based on the number of edit operations needed to transform a source string into a target string. Common edit operations are insertion, deletion, substitution and transposition of characters. Sanfeliu et al. [13] extended the edit distance from strings to graphs. Common edit operations are insertion, deletion and substitution of nodes and edges. An edit path defines a sequence of edit operations that transform one graph into another. Cost functions measure the strength of an edit operation. The graph edit distance is then the minimum cost edit path needed to transform a source graph into a target graph.

For our proposed SED, the only edit operation is the deletion of an error region. Costs are assinged to this edit operation based on the method used to perform the deletion. In this paper we focus on interactive segmentation methods where the user has to select seed pixels in order to label different regions of the segmentation [3]. Therefore the deletion of an error region corresponds to the interactive action of selecting seed pixels inside an error region.

However, segmentation of error regions close to the boundary of the ground truth is harder than the segmentation of regions that are further away from the boundary. Therefore we split the error regions into two sets, one with big regions and the other with thin regions that tightly fit the boundary. Additionally, pixel accurate ground truths are generally impossible to create for natural images, since the true border position along edges of an object cannot be obtained with absolute certainty [3]. Therefore, the costs for the thin error regions are calculated independently of the costs for the big error regions.

For the first set of big error regions we compute the number of different error regions, based on the assumption that at least one user interaction, e.g. placing seed pixels, is needed for every distinct region. Since the number of user interactions depends on properties of the error regions, we define the following cost functions:

1) Connections between error regions and the ground truth: Since segmentation close to the boundary of the ground truth is more demanding, higher costs are assigned for longer connections (see Figure 1 for an example)
2) Area of error regions:
   The area of an error region is related to the amount of user interactions that are needed. Larger error regions are linked to bigger or longer brush strokes [3] and therefore higher costs.
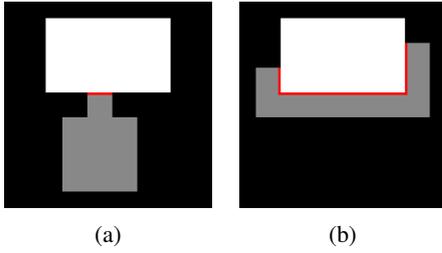
Fig. 1: Example for connections (red) between error regions (grey) and the ground truth (white).

For the second set of thin error regions that tightly fit the boundary, costs are assigned based on the distance between the error region boundary and the boundary of the ground truth.

We now define the computation of the SED for binary segmentations, where an image is only partitioned in foreground and background segments. After that we extend the approach to non-binary segmentations.

A. Computation for Binary Segmentations

In order to measure the similarity between a test foreground segment S_T and the corresponding ground truth foreground segment S_G of a binary segmentation, the following sets are computed:

1) The sets of inner boundary pixels B_T and B_G of the segmentations S_T and S_G based on the discrete algorithm by Suzuki et al. [14].

2) The set of connected components

$$C = ccl(S_T \setminus S_G) \cup ccl(S_G \setminus S_T) \qquad (1)$$

from FP and FN error regions, where ccl is the connected component labeling [15]. Each connected component is a set of pixels connected by 8-neighbourhood.

3) The set of thin connected components close to B_G

$$C_T = \{X \in C | h(X, B_G) \leq \theta\} \qquad (2)$$

where

$$h(A, B) = \max_{a \in A} \min_{b \in B} ||a - b|| \qquad (3)$$

and $||\cdot||$ is the Euclidean norm. The thinness parameter θ defines the maximum distance of points in the connected components from the boundary B_G. We define $\theta = \sqrt{2}$ to restrict the connected components in this set to thin 8-connected lines tightly fitting the boundary.

4) The set of big connected components

$$C_B = C \setminus C_T \qquad (4)$$

5) The partial boundary

$$B_P = \bigcup_{X \in C_B} B_T \setminus X \qquad (5)$$

This boundary is used in the distance based objective metric. By removing parts of the original boundary B_T that intersect the large connected components C_B, the maximum Hausdorffdistance between this boundary and the boundary B_G is defined by $\theta = \sqrt{2}$, removing any outliers.

The binary SED_2 is defined as

$$SED_2(S_T, S_G, \theta) = k \cdot (c_1 + c_2) + c_3 \qquad (6)$$

where θ is the thinness parameter, k is the number of big error regions

$$k = |C_B| \qquad (7)$$

and c_i are the following costs, where $|\cdot|$ gives the number of elements:

1) Length of the boundary between big error regions and the ground truth (c_1)

$$c_1 = \frac{\left| \bigcup_{X \in C_B} B_G \cap X \right|}{|B_G|} \qquad (8)$$

It is normalized to the length of the ground truth boundary B_G.

2) Area of big error regions (c_2)

$$c_2 = \frac{\sum_{X \in C_B} |X|}{|S_G|} \qquad (9)$$

It is normalized to the area of the ground truth segmentation S_G. That makes it comparable to the metric c_1 which is also normalized to the ground truth.

3) Distance between test and ground truth boundaries (c_3)

$$c_3 = \frac{max(h_m(B_G, B_P), h_m(B_P, B_G))}{\theta} \qquad (10)$$

where B_P is used instead of the full test boundary in order to exclude big regions. θ is the thinness parameter corresponding to the maximum Hausdorff distance between the two boundaries ensured by (2). h_m corresponds to the modified median Hausdorff distance introduced by Chetverikov et. al. [16].

$$h_m(M, R) = \frac{1}{N_M} \sum_{a \in M} \min_{b \in R} ||a - b|| \qquad (11)$$

The costs c_1 and c_2 are calculated for the number of big error regions and are therefore multiplied by k. The cost c_3 is independent of the big error regions and is therefore added at the end.

B. Extention to non-binary Segmentations

In order to extend the previously presented approach to segmentations where an image is partitioned into more than two segments, the SED_2 is computed for every pair of test and ground truth segment seperately. In the end all individual scores are summed up to give the score for the
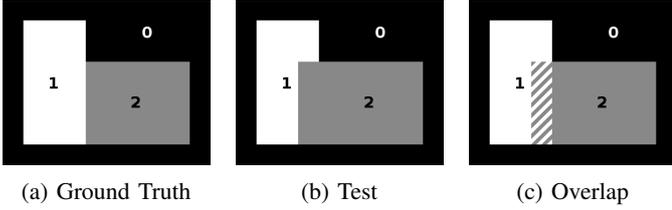
(a) Ground Truth   (b) Test   (c) Overlap

Fig. 2: Example segmentation with three segments. The error region between segment 1 and segment 2 (striped) is a FN region for segment 1 and a FP region for segment 2.

whole segmentation. When computing the SED_2 for individual segment pairs, one problem arises: Error regions may be shared between different segments, and therefore influence the edit distance score of more than one segment pair. Figure 2 shows an example of a segmentation with three segments and overlapping error regions.

In order to fix this, overlapping error regions from all segments in the test segmentation are removed accourding to Algorithm 1.

The general SED is then defined as

$$SED(T, G, \theta) = \sum_{i=1}^{N} SED_2(T_i, G_i, \theta) \qquad (12)$$

where T and G are the sets of segments from the test and ground truth segmentations respectively and N is the number of segments in G.

Algorithm 1 Remove overlapping error regions

Input: set of test segments T,
        set of ground truth segments G
Output: updated set of test segments T
 1: $N \leftarrow |G|$
 2: **for** $i = 1$ to N **do**
 3:     $TP_i \leftarrow G_i \cap T_i$
 4:     $FP_i \leftarrow T_i \setminus G_i$
 5:     $FN_i \leftarrow G_i \setminus T_i$
 6: **end for**
 7: **for** $j = 1$ to N **do**
 8:     **for** $k = j + 1$ to N **do**
 9:        $FP_k \leftarrow FP_k \setminus FP_j$
10:        $FN_k \leftarrow FN_k \setminus FN_j$
11:     **end for**
12:     $T_j \leftarrow TP_j \cup FP_j \cup FN_j$
13: **end for**
14: **return** T

IV. RESULTS

We tested our proposed edit distance on two datasets, one with synthetic segmentations, referred to as synthetic dataset and the other with segmentations from natural images of horses, referred to as horse dataset. For comparison, we chose two metrics from Taha and Hanbury [5]. The first metric is

(a) Seg01   (b) Seg02   (c) Seg03   (d) Seg04

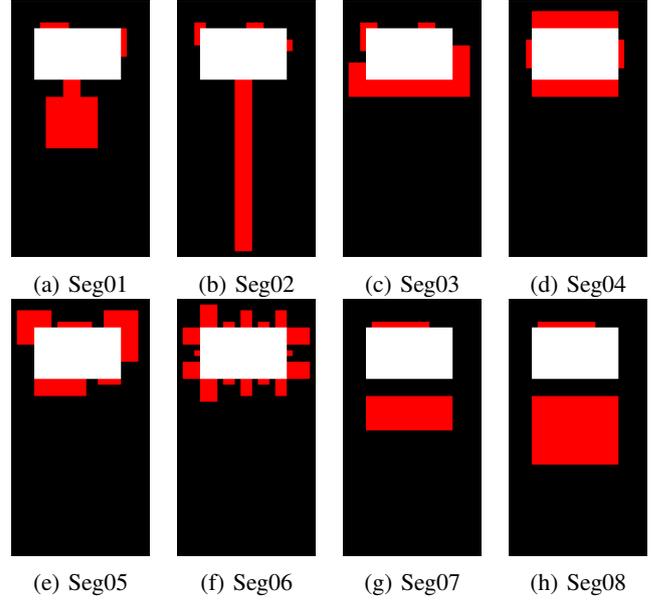(e) Seg05   (f) Seg06   (g) Seg07   (h) Seg08

Fig. 3: Synthetic dataset with ground truth (white) and error regions (red).

the Dice Coefficient (DICE), a spatial overlap based metric defined by

$$DICE = \frac{2TP}{2TP + FP + FN} \qquad (13)$$

It is based on the four basic overlap cardinalities. For our evaluation, this metric also covers other state of the art metrics such as the Rand Index or the Variation of Information, since they can be expressed by the same basic overlap cardinalities [5] and feature the same drawbacks as discussed in Section II. The second metric is a spatial distance based metric, the Average Distance (AVD), defined by

$$AVD(A, B) = max(h_m(A, B), h_m(B, A)) \qquad (14)$$

where h_m is the modified median Hausdorff distance (11).

A. Synthetic Dataset

The synthetic dataset is composed of eight different synthetic segmentations that are compared to the same ground truth (Figure 3). The area of the error regions is the same for every segmentation in this dataset, except $Seg08$, but they are distinguishable by different properties such as the number of components or the size of the connections to the ground truth boundary.

Table I shows the results of the synthetic dataset for the DICE, AVD and our proposed SED_2. As expected, the DICE index is not able to distinguish the different segmentations, except $Seg08$, since the error regions are the same size in every other example of the dataset. The AVD is sensitive to outliers, assigning the largest score to $Seg02$ and the lowest to $Seg06$. Our proposed metric assigns the lowest score to $Seg07$ since only one error region is present and there is no

TABLE I: Distance scores for the synthetic dataset (best scores are bold per column).

Segm.	DICE	AVD	SED_2	k	c_1	c_2	c_3
Seg01	**0.730**	3.455	0.903	1	0.068	0.667	0.168
Seg02	**0.730**	9.331	0.906	1	0.068	0.667	0.171
Seg03	**0.730**	2	1.473	1	0.500	0.667	0.307
Seg04	**0.730**	2	2.908	2	0.591	0.667	0.393
Seg05	**0.730**	1.937	3.726	3	0.477	0.667	0.295
Seg06	**0.730**	**1.917**	7.321	6	0.500	0.667	0.321
Seg07	**0.730**	3.134	**0.827**	1	0.000	0.667	0.161
Seg08	0.587	5.160	1.494	1	0.000	1.333	0.161

Fig. 4: One example image from the horse dataset.

(a) Background subtraction

(b) Two-stage Otsu

(c) Grabcut

Fig. 5: Ground truth (white) and error regions (red) for a part of the segmentation of *c_allora*.

connection to the boundary of the ground truth. The table also shows k, c_1, c_2 and c_3 for the proposed edit distance. $Seg07$ and $Seg08$ show big error regions that are not connected to the ground truth segmentatio, which is regared as the best case for error regions. However, the SED_2 score for $Seg08$ is still larger than SED_2 scores for other segmentations in the dataset. This is due to the fact that the area ($c2$) is much larger, negating the effect that the error region is not connected to the ground truth.

B. Horse Dataset

The horse dataset consists of eight natural images of different horses taken from video sequences with three test segmentations per image that are compared to the same ground truth (Figure 4 shows an example image of the dataset). The images are from a project were image processing methods are used to increase the accuracy, objectivity and throughput of phenotypic estimation of Lipizzan horses [17]. Traits of the horses are interpreted based on the contour of the segmented horse shape. Therefore, the quality of the segmentation is important. Three algorithms were used to create an initial segmentation of the horse from the background. The first algorithm is a simple background subtraction between the current frame of the video and a background frame without horse, taken with a stationary camera. The second algorithm is a two-stage Otsu [18] that selects an optimal threshold. The last method is GrabCut [19], an interactive iterative segmentation algorithm based on graph cuts where seed points are used to mark foreground and background regions. The first automatic segmentation of GrabCut is improved by results from the two-stage Otsu.

Table II shows the results of the horse dataset for DICE, AVD and SED comparing the three segmentation algorithms background subtraction (BG-Sub.), two-stage Otsu (Otsu) and Grabcut. The segmentation by the Grabcut algorithm should have the lowest score for every image of the dataset, since it gives the best approximation of the ground truth border. Figure 5 shows examples of the horse *c_allora* from the horse dataset. Only the SED_2 assigns the lowest score to all segmentation by the Grabcut algorithm, although the DICE and AVD metrics assign a lower value to the two-stage Otsu algorithm only for one image of the dataset. Looking at the individual metrics for this dataset, the scores for DICE and AVD lie much closer
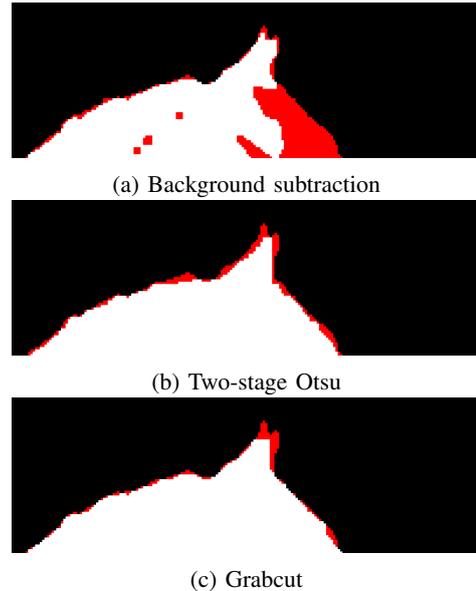
together than the scores of the SED_2 which gives clearer results between the three segmentation algorithms.

Table III shows SED results for the horse *c_allora* with all objectiv metrics of the proposed edit distance. The number of big error regions and therefore the number of minimal user interactions.

V. CONCLUSION

This paper proposes a novel distance metric called Segmentation Edit Distance (SED). The basic concept of edit distances such as the string edit distance or graph edit distance is extended to the domain of image segmentations. It measures the distance between a test segmentation and the associated ground truth segmentation based on the cost of edit operations that are needed to fix error regions between the two segmentations. The only edit operation is the deletion of error regions and costs are assigned to it based on properties of the error regions and the image processing method used for the deletion. This work focuses on interactive segmentation methods, where users have to input seed pixels and the deletion

TABLE II: Results of the DICE, AVD and SED_2 scores on the horse dataset (best scores are bold per score for every line).

Horse	DICE			AVD			SED_2		
	BG-Sub.	Otsu	Grabcut	BG-Sub.	Otsu	Grabcut	BG-Sub.	Otsu	Grabcut
c_allora	0.894	**0.960**	0.953	12.611	**2.890**	4.532	40.346	47.067	**11.305**
c_bradamanta	0.920	0.943	**0.987**	10.420	3.443	**0.891**	23.569	32.362	**5.307**
c_samira	0.936	0.946	**0.976**	7.923	4.595	**2.247**	28.076	48.716	**13.564**
c_wanda	0.930	0.950	**0.978**	15.086	3.829	**3.006**	26.884	35.337	**12.524**
m_fantasca_b	0.909	0.956	**0.978**	8.755	2.511	**1.872**	24.195	43.232	**10.092**
m_malina	0.937	0.940	**0.979**	9.807	4.028	**2.247**	21.687	41.753	**7.287**
m_rustica	0.926	0.942	**0.977**	13.590	3.815	**2.824**	31.297	29.613	**11.136**
n_aga	0.816	0.935	**0.978**	17.962	4.648	**2.060**	25.710	51.153	**9.1**

TABLE III: Results of the SED score for the horse *c_allora* (best scores are bold for every column).

Method	SED_2	k	c_1	c_2	c_3
BG-Sub.	40.346	57	0.516	0.187	0.305
Otsu	47.067	58	0.732	**0.073**	0.384
Grabcut	**11.305**	23	**0.391**	0.092	**0.199**

of an error region corresponds to the interactive action of placing such seed pixels inside an error region.

The costs associated with these interactive actions are related to properties of the error regions such as length of the boundary between the error region and the ground truth or the area of the region. Furthermore, costs for imprecisions close to the boundary of the ground truth are assigned based on the distance between these thin regions and the boundary. By including additional metrics, such as shape descriptors like the rectangularity or circularity, the analysis of error regions could be extended.

Moreover, currently only the binary shapes of error regions are considered for the cost calculation. As a future research direction, the results could be improved by incorporating information from the feature space, for example texture information of the error regions from local binary patterns [20]. This would allow more precise statements about the costs associated to the edit operation and could relax the assumption that at least one edit operation is needed for every distinct error region.

In order to use the proposed edit distance for other methods than interactive segmentation with seed pixels, the costs have to be adapted based on properties of the new method.

REFERENCES

[1] F. Heckel, J. H. Moltz, H. Meine, B. Geisler, A. Kießling, M. DAnastasi, D. P. dos Santos, A. J. Theruvath, and H. K. Hahn, "On the evaluation of segmentation editing tools," *Journal of Medical Imaging*, vol. 1, no. 3, pp. 034 005–034 005, 2014.

[2] K. McGuinness and N. E. Oconnor, "A comparative evaluation of interactive segmentation algorithms," *Pattern Recognition*, vol. 43, no. 2, pp. 434–444, 2010.

[3] K. McGuinness and N. E. OConnor, "Toward automated evaluation of interactive segmentation," *Computer Vision and Image Understanding*, vol. 115, no. 6, pp. 868–884, 2011.

[4] E. Moschidis and J. Graham, "Simulation of user interaction for performance evaluation of interactive image segmentation methods," in *Medical Image Understanding and Analysis*, 2009, pp. 209–213.

[5] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC medical imaging*, vol. 15, no. 1, p. 29, 2015.

[6] B. Peng, L. Zhang, and D. Zhang, "A survey of graph theoretical approaches to image segmentation," *Pattern Recognition*, vol. 46, no. 3, pp. 1020–1038, 2013.

[7] N. D. T. Dung and H. T. T. Binh, "Using contour information for image segmentation," in *Soft Computing and Pattern Recognition (SoCPaR), 2013 International Conference of*. IEEE, 2013, pp. 258–263.

[8] X. Deng, L. Zhu, Y. Sun, C. Xu, L. Song, J. Chen, R. D. Merges, M.-P. Jolly, M. Suehling, and X. Xu, "On simulating subjective evaluation using combined objective metrics for validation of 3d tumor segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2007, pp. 977–984.

[9] J. Funke, J. Klein, F. Moreno-Noguer, A. Cardona, and M. Cook, "Ted: A tolerant edit distance for segmentation evaluation," *Methods*, vol. 115, pp. 119–127, 2017.

[10] R. W. Hamming, "Error detecting and error correcting codes," *Bell Labs Technical Journal*, vol. 29, no. 2, pp. 147–160, 1950.

[11] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," in *Soviet physics doklady*, vol. 10, no. 8, 1966, pp. 707–710.

[12] F. J. Damerau, "A technique for computer detection and correction of spelling errors," *Communications of the ACM*, vol. 7, no. 3, pp. 171–176, 1964.

[13] A. Sanfeliu and K.-S. Fu, "A distance measure between attributed relational graphs for pattern recognition," *IEEE transactions on systems, man, and cybernetics*, no. 3, pp. 353–362, 1983.

[14] S. Suzuki *et al.*, "Topological structural analysis of digitized binary images by border following," *Computer vision, graphics, and image processing*, vol. 30, no. 1, pp. 32–46, 1985.

[15] K. Suzuki, I. Horiba, and N. Sugie, "Linear-time connected-component labeling based on sequential local operations," *Computer Vision and Image Understanding*, vol. 89, no. 1, pp. 1–23, 2003.

[16] D. Chetverikov and Y. Khenokh, "Matching for shape defect detection," in *Computer Analysis of Images and Patterns*. Springer, 1999, pp. 82–82.

[17] T. Druml, A. Gabdulkhakova, N. Artner, G. Brem, and W. Kropatsch, "The use of image data in the assessment of equine conformation–limitations and solutions," in *22nd International Conference on Pattern Recognition, Workshop on Visual Observation and Analysis of Vertebrae and Insect Behaviour*, 2014, pp. 21–25.

[18] D.-Y. Huang, T.-W. Lin, and W.-C. Hu, "Automatic multilevel thresholding based on two-stage otsus method with cluster determination by valley estimation," *International journal of innovative computing, information and control*, vol. 7, no. 10, pp. 5631–5644, 2011.

[19] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," in *ACM transactions on graphics (TOG)*, vol. 23, no. 3. ACM, 2004, pp. 309–314.

[20] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010.