Classification of Aneurysm Morphology

Yuting Tang
Electrical and Computer Engineering, UCLA
Los Angeles, USA
tyt2110@g.ucla.edu

Xiao Peng
Electrical and Computer Engineering, UCLA
Los Angeles, USA
xpeng94@g.ucla.edu

***Abstract*—This project mainly applies the 3D shape context as the descriptor of the aneurysm. The results of the average of three types of total cost are analyzed to find the differences between the growing aneurysm and the stable aneurysm.**

Keywords—aneurysm, shape context, cost matrix, distance

I. Methods

The data we use totally contains 10 3D-shapes which is shown in the form of the .stl files. Among them, 10 figures have been divided into 2 categories, namely growing and stable. By analyzing the data, we are trying to find out whether the results of the shape context of the two categories would be different or not.

The specific steps in our project are as follows.

A. Reading the .stl File

Firstly, we need to read the .stl files and transform them into the form that could be used in Python. Since there are some irrelevant information to our project, we could only extract the location information which should be analyzed later from the 3D-figures.

B. Sampling

Sampling is quite necessary in this project, since there are too many data points in each 3D figure, which would also slow down the operation of the program. We have applied the random sampling to the data to gain the sample which is appropriate for the process. And random sampling is also a good way to gain a unbiased sample of the data.

C. Computing the Shape Context

Computing the shape context is the key step in the project. In this part, instead of only do the calculate the bins of the center point, we create the bins of every data point in the sample dataset. By creating a spherical coordinate for each data point in the 3D graph, we could calculate the coordinates for each other point. Therefore the bins of every point in the 3D figure consists of the bins of the whole 3D shape which is the main feature of the graph of the aneurysm we would use later.

D. Computing the Cost Matrix

As the results of shape context are distributions represented as histograms, in order to compare two different graphs, we need to use the "shape context cost" of matching the two points.

$$Cs = \frac{1}{2}\sum_{k=1}^{K}\frac{\left[g(k)-h(k)\right]^2}{g(k)+h(k)}$$

Now for each point in the first shape and another point in the second shape, the cost as described is calculate, consisting the cost matrix.

E. Finding the Best Matching

In this part, we applied Munkres algorithm to find the matching that minimizes total cost, which could also be called the best matching. The basic principle is shown as below.

$$H(\pi) = \sum_i C(p_i, q_{\pi(i)})$$

F. Computing the Shape Distance

A shape distance between two 3D shapes P and Q need to be computed, which is actually the minimum total cost we have calculated before.

G. Setting the Threshold

A threshold need to be set up to find out whether the two 3D shapes match or not. If the shape distance is smaller than the threshold, it could be concluded that the two shapes belong to the same category, while the two shapes belong to different category if the shape distance is bigger than the threshold.

According to the data we have, we have compared every two figures of all figures and averaged the shape distances respectively of two growing aneurysms, two stable aneurysms and one for each growing and stable aneurysms.

II. Results

Since the points are sampled by random, two images are not totally identical even though they are sampled from the same graph. The example of the images of BI8022 with 500 random sample points when we compare it with itself is shown below.

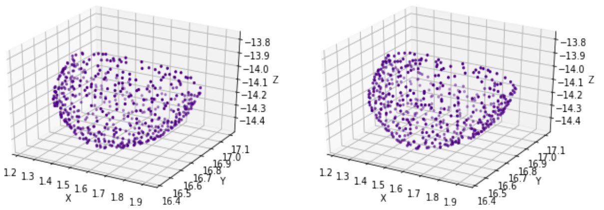

Fig. 1. BI8022 with 500 random sample points when compared with itself

By applying the method we designed to the data we have, we could calculate the total cost for every two of ten files. The results are listed in the tables below.

	BI8022	MM8894	RS4668	VR1339	WC610
BI8022		46707	54855	52060	59034
MM8894			95314	56731	76777
RS4668				75419	41376
VR1339					73389
WC610					

Fig. 2. The total costs for every two of stable aneurysms

	FE1546	LR1592	LR4789	TK928	XJ3252
BI8022	63017	65733	75497	36062	88859
MM8894	42852	25065	28242	63571	49489
RS4668	108471	86477	97571	81052	60228
VR1339	49844	64332	51060	67897	50466
WC610	80420	74236	75075	61712	69266

Fig. 3. The total costs for one for each growing and stable aneurysms

	FE1546	LR1592	LR4789	TK928	XJ3252
FE1546		49272	52095	47560	67064
LR1592			30826	54494	42690
LR4789				52081	57335
TK928					48082
XJ3252					

Fig. 4. The total costs for every two of growing aneurysms

III. Discussion

Our project works on applying the classification of the aneurysm morphology on the target .stl file based on the existing labelled trainset. But out of the confidential consideration for patients' privacy, we as well as Professor Fabian can only obtain 5 .stl files for stable aneurysm case and growing aneurysm case respectively as the trainset of our classification system. Definitely, only 5 files for each class is far from enough for training algorithm, so we change our goals to extract the features of every stable and growing aneurysm cases, then use the processing methods mentioned above to calculate the average features for stable and growing cases. So we are doing the analysis and the discussion based on the results we got.

Firstly, we have compared every two figures of all .stl files and calculated the shape distances respectively of two growing aneurysms, two stable aneurysms, and one for each growing and stable aneurysms. If the threshold is set as 60000, which is relatively appropriate according to the results we have got, we could find that the shaded results in the three tables listed above are correctly classified. And the accuracies could be calculated. As for stable aneurysm cases, 65.7% (23/35) of them are correctly classified. As for the growing aneurysm cases, 74.3% (26/35) of them are correctly classified. And the total accuracy is 71.1% (32/45). What is more, the classification results are quite poor for certain cases such as MM8894 and VR1339, which also shows the unsatisfying application range and error-tolerant for different shapes. Therefore, the method with 3D shape context used in the project could be a relatively reliable method for classification of aneurysm morphology.

Besides, as it has been shown in the table, the accuracies of the results are not perfectly ideal. The main reasons of imperfect performance are the deficiency of the data and insufficient sample points due to the limited computing resource we have. The former will be mentioned in the end, while latter one is rather practical in our project. We now set the amount of the sample points as 500, but we've done the experiment to verify that with higher amount of sample points, like 2000, the more distinguishing total costs we can get for different category, which obviously facilitates threshold setting and leads to better classification performance.

In the end of the report, we would like to discuss in the future what we can do to improve the performance of the aneurysm classification system. The prime concern is we need more labelled samples as our input training dataset. Only with enough trainset, can the predictions get more accurate result. And furthermore, we should try some more specific methods according to the characteristic features of the shapes to be tested. To be more precise, the shapes we have contain a semi sphere representing the aneurysm, and the round-like cross-section representing the joint surface of the vessel and the aneurysm. In the project, we take the semi sphere as the only feature and neglect the features lying in the cross-section. In fact, the cross-section in the shapes can provide us with abundant information of the aneurysm, such as the location of the cross-section on the semi sphere as well as the relative size of the cross-section and the whole aneurysm. Predictably, the information like these can greatly facilitate the prediction process and promote the classification performance in some sense. The methods like this are what we would like to make attempts at.

Actually, shape context can't be the only descriptor of 3D shape matching, and there are other ways to implement the feature extraction. Shape context is one kind of local descriptor that analyze the shape in overview perspective and unable to capture the specific details. So for new descriptors, we can explore other global descriptors like cord and angle histograms or shape distributions which are also usually used in 3D shape matching nowadays. The idea of shape distributions is to use shape distribution sampled from a shape function to represent the signature of an object as a measuring global geometric properties of an object. The main advantage of it is effectively simplifying the shape matching to a possibility comparison problem without need to do the feature correspondence or model fitting.

REFERENCES

[1] Serge Belongie, Jitendra Malik. Matching with Shape Contexts. 0-7695-IEEE 0695-X/00 2000.

[2] J.W.H. Tangelder and R. C. Veltkamp, "A survey of content based 3D shape retrieval methods," Proceedings Shape Modeling Applications, 2004., 2004, pp. 145-156.doi: 10.1109/SMI.2004.1314502.

[3] S. Belongie, J. Malik and J. Puzicha, "Shape matching and object recognition using shape contexts," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 509-522, Apr 2002.doi: 10.1109/34.993558.

[4] Tombari F., Salti S., Di Stefano L. (2010) Unique Signatures of Histograms for Local Surface Description. In: Daniilidis K., Maragos P., Paragios N. (eds) Computer Vision – ECCV 2010. ECCV 2010. Lecture Notes in Computer Science, vol 6313. Springer, Berlin, Heidelberg.

[5] L.Z., M.J.F., A.F., "Survey on 3D Shape Descriptors," in FCT.