ADAPTIVE GRAPH FORMULATION FOR 3D SHAPE REPRESENTATION

Basheer Alwaely and Charith Abhayaratne

Department of Electronic and Electrical Engineering
The University of Sheffield
Sheffield, S1 4ET, United Kingdom
Email: b.alwaely@sheffield.ac.uk, c.abhayaratne@sheffield.ac.uk

ABSTRACT

3D shape recognition has attracted a great interest in computer vision due to its large number of important and exciting applications. This has led to exploring a variety of approaches to develop more efficient 3D analysis methods. However, current works take into account descriptions of global shape to generate models, ignoring small differences causing the problem of mismatching, especially for high similarity shapes. The present paper, therefore, proposes a new approach to represent 3D shapes based on graph formulation and its spectral analysis which can accurately represent local details and small surface variations. An adaptive graph is generated over the 3D shape to characterise the topology of the shape, followed by extracting a set of discriminating features to characterise the shape structure to train a classifier. The evaluation results show that the proposed method exceeds the state-of-the-art performance by 4% for a challenging dataset.

Index Terms— 3D shape representation, Graph theory, Connectivity, Growing Neural Gas (GNG), Graph spectral analysis.

1. INTRODUCTION

The past decade has witnessed a great interest in analysing and understanding 3D shapes because it contributes to applications, such as, gaming [1–3], security [4, 5] and human activity understanding [6, 7]. The existing work on 3D shape recognition can be categorised into three different areas. Firstly, deep learning techniques [8, 9] show a great improvement in the recognition level. Due to the amount of training and data needed, it is not suitable for all datasets. Secondly, model-based approaches [10, 11] provide an alternative representation such as skeleton representation based on the shape silhouettes. However, high similarity shapes are still challenging for these approaches. Lastly, feature-based approaches [12–15] demonstrate a fast implementation with an efficient recognition level. Usually, more than one type of descriptor is required to describe such complex shapes, and then these features are classified based on machine learning technique. Graphs and complex networks display useful topological features based on the types of connection between their elements. Such features include degree distribution, clustering coefficient, and hierarchical structure. With the emergence of graph signal processing, graph spectral domain feature extraction have been used for 2D shape analysis [16, 17]. Forming a complex network for 3D shape has been considered in [15].

This paper proposes a new method to represent the 3D shapes by capturing local variations into the feature representation. The proposed method begins by extracting a specific number of pixels to represent the shape followed by generating a graph connecting these pixels adaptively to capture the shape structure. The proposed adaptive connectivity is designed to capture the geometric variations of the individual shapes into the graph structure. After that, we create a signature based on the graph's spectral properties to describe each individual shape. Finally, these features are classified using the K-nearest neighbour classifier. The main contributions from this paper are:

- A new method for graph formulation with adaptive connectivity to represent 3D shapes preserving both local and global characteristics of the shape.
- The proposal of a new set of graph spectral features based on the node distribution of the adaptive graph for 3D shape representation.

The rest of the paper is organised as follows: The proposed method is presented in Section 2, followed by the performance evaluations is Section 3 and the conclusions in Section 4.

2. THE PROPOSED METHOD

2.1. Point cloud representation

3D object S contains of n points cloud, with coordinates, (x_i, y_i, z_i), $i = 0, ..., n-1$. Usually, each object contains a large number of pixels on the 3D Cartesian plane. Therefore, we use Growing Neural Gas (GNG) to reduce the complexity

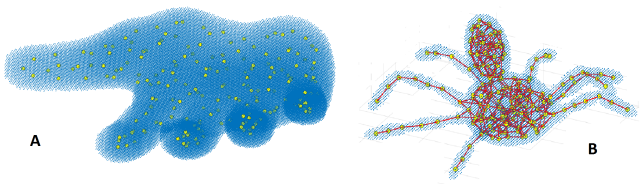

Fig. 1. 3D shape representation using Growing Neural Gas.

and to get the same number of pixels for each object (normalisation). The input data of the GNG are the coordinates of a 3D point cloud. Based on the Euclidean distance between pixels, we train the GNG and it grows gradually inside the shape's region using a pre-determined number of pixels (N). At the end of the training process, the GNG should satisfactory cover the shape regions as can be seen in Fig. 1. The detailed description of GNG can be found in [18].

2.2. Graph preliminaries

We can now represent the nodes on the surface as an undirected graph, $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{A}\}$, where $\mathcal{V}$ is the set of N vertices, $\mathcal{E}$ is the set of edges and $\mathbf{A}$ is the adjacency matrix with edge weights. We consider $\mathcal{E}$ as the Euclidean distance between nodes because it is invariable to rotation and translational operations. We define the weight, $\mathbf{A}_{i,j}$ corresponding to $\mathcal{E}_{i,j}$ connecting between vertices i and j as follows:

$$\mathbf{A}_{i,j} = \begin{cases} \mathcal{E}_{(i,j)}, & \text{if nodes } i \text{ and } j \text{ are connected,} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

We also define the signal $\mathbf{r} : \mathcal{V} \rightarrow \mathbb{R}$, where i^{th} component represents the Euclidean distance from the centre (0,0,0) to the vertex i in $\mathcal{V}$ as follows:

$$\mathbf{r}_i = \sqrt{x_i^2 + y_i^2 + z_i^2}. \quad (2)$$

The graph Laplacian matrix, $\mathbf{L} = \mathbf{D} - \mathbf{A}$, is then calculated, where $\mathbf{D}$ is the diagonal matrix of vertex degrees, whose diagonal components are computed as $\mathbf{D}_{(i,i)} = \sum_{j=0}^{N-1} \mathbf{A}_{(i,j)}$. Since, $\mathbf{L}$ is symmetrically positive, there exists a real unitary matrix, $\mathbf{U}$, that diagonalizes $\mathbf{L}$, such that $\mathbf{U}^t\mathbf{L}\mathbf{U} = \Lambda = diag\{\lambda_\ell\}$ is a non-negative diagonal matrix, leading to an eigenvalue decomposition of $\mathbf{L}$ matrix as follows:

$$\mathbf{L} = \mathbf{U}^t \Lambda \mathbf{U} = \sum_{\ell=0}^{N-1} \lambda_\ell \mathbf{u}_\ell \mathbf{u}_\ell^t, \quad (3)$$

where $\mathbf{u}_\ell$, the column vectors of $\mathbf{U}$, are a set of orthonormal eigenvectors of $\mathbf{L}$ with corresponding eigenvalues, $0 = \lambda_0 < \lambda_1 \leq \lambda_2 ... \leq \lambda_{N-1} = \lambda_{max}$ [19].

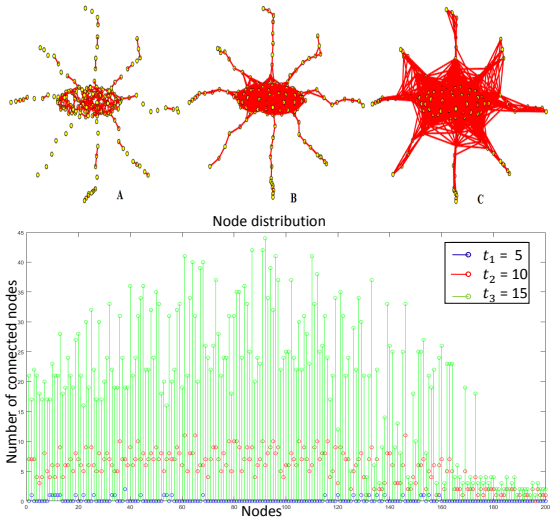

Fig. 2. Dynamic connectivity over the same object (Top row) using, A: $t_1 = 5$, B: $t_2 = 10$, C: $t_3 = 15$ and its node distribution (Bottom row).

2.3. Graph connectivity

The type of connectivity has a direct effect on the graph spectral basis. To design the optimal graph connectivity, we suggest using conditional connectivity; where nodes are linked together if, and only if, they satisfy a certain threshold (t). The generated graph based-threshold $\mathcal{G}(t_i)$ represents a temporary transition state of the object with certain properties associated with the threshold value. Fig. 2 illustrates the result of applying different threshold values over the same object. We consider t as the smallest distance that keeps all nodes connected as one group, which means that there are no separated nodes. The main difference between each $\mathcal{G}(t_i)$ is the connectivity degree of each vertex, which gives a general understanding of the nature of the surface.

To set the optimal threshold (T) a dataset, in our previous work on 2D shape recognition [17], we demonstrated that t is the minimum acceptable solution. Similarly, in [15], an experimentally determined fixed threshold for all shapes in each dataset was considered. In the present paper, we propose using individual t for each shape depending on the surface details as follows:

$$T_s = t + \delta, \quad (4)$$

where δ is the optimal value that can be added to give the optimal possible distinction to all the objects in the dataset. This means that our proposed condition results from local details of each individual samples (t), and the discriminatory value (δ), which is determined based on all objects in the dataset as shown in Section 2.4.

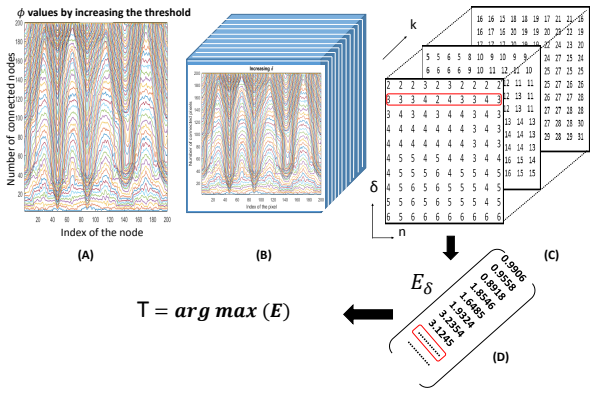

Fig. 3. Dynamic evaluations, A: number of connected pixels for one sample, B: combining all samples, C: matrix representation of the connectivity, D: the variance (σ) of each node distribution (Φ).

Algorithm 1 Computing T

for $i = 0 : (m-1)$ **do**
  $\hat{\boldsymbol{P}}_{\boldsymbol{i}} \leftarrow$ Shape representation as (x_i,y_i,z_i).
  $t \leftarrow$ Minimum distance to link all pixels as a one group.
  for $\delta = 0 : (N-1)$ **do**
    $\boldsymbol{\Phi_\delta} \leftarrow$ Node distribution at $(t+\delta)$.
    $\hat{\boldsymbol{\Phi}}_{\boldsymbol{\delta}} \leftarrow$ Normalising $(\frac{\boldsymbol{\Phi_\delta}}{max(\boldsymbol{\Phi_\delta})})$.
    ${\sigma^2}_\delta \leftarrow$ Compute the variance of $\hat{\boldsymbol{\Phi}}_{\boldsymbol{\delta}}$.
  end
  $T_i \leftarrow Max(\boldsymbol{\sigma^2})$.
end

2.4. The optimal threshold (T).

In this study, we define the optimal threshold (T) as the value that gives the maximum variation of the nodes distribution. Therefore, for each shape S_i with N nodes in the dataset that contains m shapes, we compute the variance of the node distribution at t. Then, we increase δ from $0 \rightarrow N-1$ to find the node distribution corresponding to the maximum variance, as shown in Algorithm 1 and Fig. 3 to get T as follows:

$$T = \underset{\delta}{\operatorname{argmax}}(\sigma_\delta^2). \tag{5}$$

2.5. 3D shapes feature representation

In order to classify 3D objects, it is important to generate a vector feature that describes the geometric details of the shapes surface. To achieve such a task, we explore node domain and spectral domain. Also, this study takes into consideration the global and local details as shown:

1. Local features (F_L) are presented by number of connected elements at each pixel, Φ_i.

$$F_L = \hat{\Phi}_i, \qquad i = 0, \ldots, N-1. \tag{6}$$

2. Global shape features (F_G) are presented by scaling the graph eigenvalues by the distance $(\ominus)$. This combination results in highly discriminating features of the global shape.

$$F_G = \mathbf{r}_i \lambda_i, \qquad i = 0, \ldots, N-1. \tag{7}$$

3. Also, other statistics of the node distribution, are used for detection. These information are invariant to the rotation such as:

   (a) $F_{(3,1)}$ = the mean.
   (b) $F_{(3,2)}$ = the variance.
   (c) $F_{(3,3)}$ = the entropy as in Eq. (8),

$$F_{(3,3)} = \sum_{0}^{n-1} \hat{\boldsymbol{\Phi}}_{\boldsymbol{i}} \log 2 \left(\frac{1}{\hat{\boldsymbol{\Phi}}_{\boldsymbol{i}}} \right). \tag{8}$$

   (d) $F_{(3,4)}$= the summation of square node distribution (Eq. (9)).

$$F_{(3,4)} = \sqrt{\sum_{0}^{n-1} \hat{\boldsymbol{\Phi}}_{\boldsymbol{i}}^{\,\mathbf{2}}}. \tag{9}$$

The total length of the features is $(2N+4)$. In the final step, these features are concatenated and categorised using k-Nearest Neighbour classifier (KNN). This study uses the KNN classifier because it is a very simple classification algorithm based on the nearest model with fast prediction rate.

3. PERFORMANCE EVALUATION

The proposed method has been tested against a variety of shapes using artificial 3D shape dataset [20]. This dataset provides 19 classes$\times$ 20 samples per class $=$ 380 shapes in total. The shapes in each class were presented in different orientations, scales and articulation, making it one of the most challenging datasets. These classes include: Human, Cup, Glasses, Airplane, Ant, Chair, Octopus, Table, Teddy bear, Hand, Plier, Fish, Bird, Mech, Bust, Armadillo, Bearing, Vase, and Four Leg respectively.

Initially, we reduce the number of pixels for all shapes into N = 200 using GNG technique as shown previously. After that, we normalised the distance of each shape in such a way that the maximum Euclidean distance between nodes =1 to avoid scaling variations. Graphs are then generated using a certain threshold T_i associated with individual shape as illustrated in Eq. (4). No attempt was made to determine the

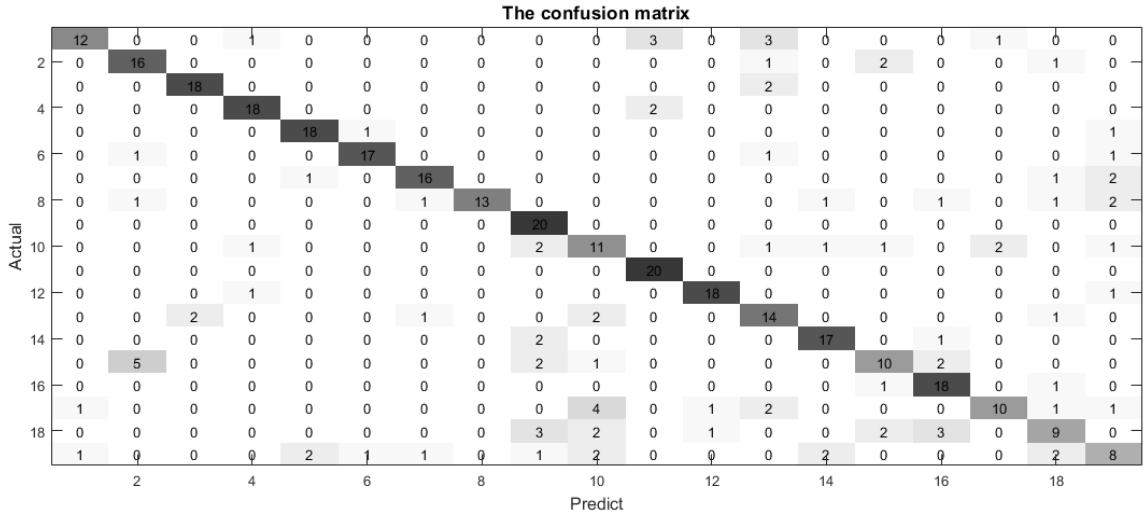

Fig. 4. The confusion matrix of the proposed method using Principle Benchmark dataset of 3D shapes [20].

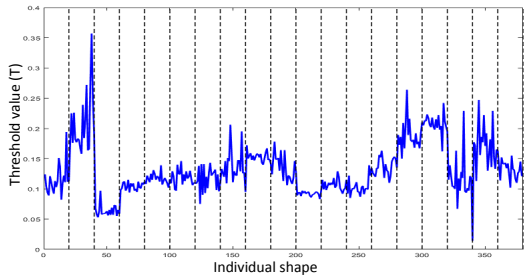

Fig. 5. Our proposed threshold value (T) based on the highest variation. The vertical lines separate different classes in the dataset.

association threshold for each shape in [15]. This work however computes a threshold for each shape, which provides a distinguishing local detail for individual shape and as a result, this leads to better performance. Fig. 5 shows how these threshold values change among different shapes. Also, we can see that T is well correlated in most classes and differs in other, giving an impression of system performance.

A set of features is extracted to represent each shape, S, using their associated T_S. Then, the KNN classifier with $k = 1$ was used to recognise shapes. Using a 10-folds cross validation technique to train and test all the dataset, our method achieves a 74.47% level of accuracy, and the confusion matrix is shown in Fig. 4. From the confusion matrix, we can see that the error samples are highly concentrated in the last five classes, which are very challenging shapes (*E.g.*, Bearing, Vase). This is mainly because the same object appears in different poses, and individual classes have different articulated shapes. However, the proposed global features (*i.e.*, F_G) capture these shapes in terms of global details.

The experiment was implemented using Matlab R2018a on a PC with Intel processor, CPU@3.6GHz and RAM 16GB. After we reduced the complexity of shapes using GNG, the average time for training and testing 380 samples was 6.3 seconds, which is fast for such a large amount of data and the average time for testing a new one sample is 12 milliseconds.

Table 1. Comparison with the different shape descriptors methods.

Method	Classification rate (%)
3D shape histogram [21]	43.42
Shape distribution [22]	67.37
Complex network [15]	70.79
The proposed method	**74.47**

To test the robustness of our proposed solution, we compared our method with well-known approaches based on histograms [21], node distribution [22] and complex networks [15]. Table 1 indicates that our proposed solution offers better performance with an increase in accuracy of 4%.

4. CONCLUSIONS

In this paper, we have proposed a new 3D shape recognition method based on local details. Our approach includes a proposal of new method for graph formulation with adaptive connectivity to represent 3D shapes. The adaptive connectivity preserves both local and global characteristics of the shape. This is followed by a new set of graph spectral and node domain features based on the node distribution of the adaptive graph for 3D shape representation. The performance evaluation based on one of the most challenging 3D dataset showed that the proposed method exceeded the existing node distribution methods by 4%.

5. REFERENCES

[1] S. Tian, L. Zhang, L. Morin, and O. Deforges, "NIQSV: A no reference image quality assessment metric for 3D synthesized views," in *Proc. on International Conference on Acoustics, Speech and Signal Processing (ICASSP),*. IEEE, 2017, pp. 1248–1252.

[2] O. Lézoray, "3D colored mesh graph signals multi-layer morphological enhancement," in *Proc. on International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 1358–1362.

[3] S. Xu, A. Amaravati, J. Romberg, and A. Raychowdhury, "Appearance-based gesture recognition in the compressed domain," in *Proc. on International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 1722–1726.

[4] H. Al-Khafaji and C. Abhayaratne, "Graph spectral domain watermarking for unstructured data from sensor networks," in *Proc. International Conference on Digital Signal Processing (DSP),*. IEEE, 2017.

[5] Z. Lu, X. Jiang, and A. Kot, "A novel LBP-based color descriptor for face recognition," in *Proc. on International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 1857–1861.

[6] H. Liu, Q. He, and M. Liu, "Human action recognition using adaptive hierarchical depth motion maps and gabor filter," in *Proc. on International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 1432–1436.

[7] Q. Huang, S. Sun, and F. Wang, "A compact pairwise trajectory representation for action recognition," in *Proc. on International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 1767–1771.

[8] J. Li, B. Chen, and G. Lee, "SO-Net: Self-organizing network for point cloud analysis," in *Proc. of Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9397–9406.

[9] D. Maturana and S. Scherer, "Voxnet: A 3D convolutional neural network for real-time object recognition," in *Proc. of international conference on Intelligent Robots and Systems (IROS),*. IEEE, 2015, pp. 922–928.

[10] S. Baloch and H. Krim, "Object recognition through topo-geometric shape models using error-tolerant subgraph isomorphisms," *transactions on image processing*, vol. 19, no. 5, pp. 1191–1200, 2010.

[11] H. Huang, S. Wu, D. Cohen, M. Gong, H. Zhang, G. Li, and B. Chen, "L1-medial skeleton of point cloud.," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 65–1, 2013.

[12] D. van Sabben, J. Ruiz-Hidalgo, X. S. Cuadros, and J. R. Casas, "Collaborative voting of 3D features for robust gesture estimation," in *Proc. on International Conference on Acoustics, Speech and Signal Processing (ICASSP),*. IEEE, 2017, pp. 1677–1681.

[13] Z. Li and A. Bors, "3D mesh steganalysis using local shape features," in *Proc. on International Conference on Acoustics, Speech and Signal Processing (ICASSP),*. IEEE, 2016, pp. 2144–2148.

[14] X. Bai, S. Bai, Z. Zhu, and L. Latecki, "3D shape matching via two layer coding," *transactions on pattern analysis and machine intelligence*, vol. 37, no. 12, pp. 2361–2373, 2015.

[15] G. E. da Silva and A. R. Backes, "Characterizing 3D shapes: a complex network-based approach," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2018, pp. 608–612.

[16] B. Alwaely and C. Abhayaratne, "Graph spectral domain feature representation for in-air drawn number recognition," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2017, pp. 370–374.

[17] B. Alwaely and C. Abhayaratne, "Graph spectral domain shape representation," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2018, pp. 603–607.

[18] B. Fritzke, "A growing neural gas network learns topologies," in *Advances in neural information processing systems*, 1995, pp. 625–632.

[19] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, May 2013.

[20] X. Chen, A. Golovinskiy, and T. Funkhouser, "A benchmark for 3D mesh segmentation," in *Acm transactions on graphics (tog)*. ACM, 2009, vol. 28, p. 73.

[21] M. Ankerst, G. Kastenmüller, H. Kriegel, and T. Seidl, "3D shape histograms for similarity search and classification in spatial databases," in *International Symposium on Spatial Databases*. Springer, 1999, pp. 207–226.

[22] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," *Transactions on Graphics (TOG)*, vol. 21, no. 4, pp. 807–832, 2002.