

INTERPOLATION CENTERS' SELECTION USING HIERARCHICAL CURVATURE-BASED CLUSTERING

Juan C. Rodríguez^{*}
Diego Del Portillo Z^{**}
Germán Sánchez Torres^{***}

Recibido: 05/05/2010

Aceptado: 08/10/2010

ABSTRACT

It is widely known that some fields related to graphic applications require realistic and full detailed three-dimensional models. Technologies for this kind of applications exist. However, in some cases, laser scanner get complex models composed of million of points, making its computationally difficult. In these cases, it is desirable to obtain a reduced set of these samples to reconstruct the function's surface. An appropriate reduction approach with a non-significant loss of accuracy in the reconstructed function with a good balance of computational load is usually a non-trivial problem. In this article, a hierarchical clustering based method by the selection of center using the geometric distribution and curvature estimation of the samples in the 3D space is described.

Key words: clustering, point simplification, range data.

^{*} System Engineer, Master student of System and Informatics at National University of Colombia. Correo electrónico: jcrodriguez@unal.edu.co.

^{**} System Engineer, Master student of System and Informatics at National University of Colombia. *Investigación y Desarrollo en Nuevas Tecnologías de la Información y la Comunicación*, Magdalena University, Santa Marta, Colombia. Correo electrónico: delportillozaur@unimagdalena.edu.co

^{***} PhD candidate at National University of Colombia, Professor Faculty of Engineering at Magdalena University. Researchgroup: *Investigación y Desarrollo en Nuevas Tecnologías de la Información y la Comunicación*. Correo electrónico: gsanchez@unimagdalena.edu.co.

SELECCIÓN DE CENTROS DE INTERPOLACION MEDIANTE AGRUPAMIENTO JERÁRQUICO BASADO EN CURVATURA

RESUMEN

Es ampliamente conocido que algunos campos relacionados con aplicaciones de gráficos realistas requieren modelos tridimensionales altamente detallados. Las tecnologías para esto están bien desarrolladas, sin embargo, en algunos casos los escáneres láser obtienen modelos complejos formados por millones de puntos, por lo que son computacionalmente intratables. En estos casos es conveniente obtener un conjunto reducido de estas muestras con las que reconstruir la superficie de la función. Obtener un enfoque de reducción adecuado que posea un equilibrio entre la pérdida de precisión de la función reconstruida, y el costo computacional es un problema no trivial. En este artículo presentamos un método jerárquico de aglomeración a través de la selección de centros mediante la geométrica, la distribución y la estimación de curvatura de las muestras en el espacio 3D.

Palabras clave: agrupamiento, simplificación de puntos, datos de rango.

INTRODUCTION

Three-dimensional models reconstruction involves a research area that has gained an important interest in recent years due to its great potential and applicability in different fields such as the medicine, industrial automation, robotics, security and others.

Three-dimensional models reconstruction process is performed using a set of stages; many of these are sub-fields of research by themselves. The typical stages of three-dimensional reconstruction are acquisition, registration, integration, fitting and, in some particular applications, the texturing stage. The surface fitting phase aims to obtain a digital description that is accurate, concise and approximate to the real surface. The phase of range image adjustment can be viewed as four optimization problems as follows [1]:

Criterion: To choose the best function to optimize (maximize or minimize).

Estimation: Choosing the best method to optimize the selected function.

Design: Making an optimal deployment of the best method to obtain the best parameters estimation

Modeling: To determine the mathematical model that best describes the digital system, including a model error process.

Recently, one of the most widely used methods for surfaces adjustment is the interpolation with radial basis functions, because these bring good results as they have an equations system associated that is invertible even if the original data are scattered. The theory of interpolation is well defined and developed. There are extensive studies concerning with the efficiency, performance and accuracy [2, 3]; however using these theories into application fields, in which the number of samples collected from the unknown function are significantly large in relation to the processing capabilities of a classical computer, has computational limitations. In these cases it is desirable to obtain a reduced set samples from which reconstruct the function.

Reduce this samples-set without a significant accuracy loss of the reconstructed function, with a proper balance in the computational load is often a nontrivial problem. Thus, one of the biggest problems of such techniques is the selection of interpolation centers by making, a good selection of these, generates good results by reducing the computational cost.

This work focuses on the selection of interpolation centers because this is an open problem in the area of 3D image reconstruction.

Some previous works have shown a way to get the correct center set from a continuous function. In [4] is described an ideal positioning of interpolation centers from a mathematical function but avoiding considerations like data characteristics, computational costs, implementation difficulties and the intrinsic geometry of the object in the data. Those are nontrivial considerations, because in general it takes into account some characteristics about the data-set that cannot be guaranteed in the type of data used, named range data. For example: the uniformity of the center-set. According to the definition 1.1 in [4], the set of centers is perfect if the following considerations come true. For explanation convenience we take some definition in the original paper:

Let a center set characterization like

$$\mathcal{X}_{d=\{\mathbb{X} \in \mathbb{R}^3\}}$$

And h_x the fill distance and q_x the separation distance.

• *Definition 1.1.* Let $d \geq 1$ and $\mathcal{X} \in \mathcal{X}_d$ For a given $h_0 > 0$, we stay that \mathcal{X} is perfectly separated with respect to h_0 , if $h_x \leq h_0$ and

$$\sup q_Y = q_X$$

$$Y \in \mathcal{X}_d \tag{1}$$

$$h_Y \leq h_0$$

hold true.

For a fixed $q_0 > 0$, X is said to be perfectly dense with respect to q_0 , if $q_x \geq q_0$ holds.

However, it is not clear how achieving a perfectly separated and perfectly dense center-set from a scattered, non-uniform, noisy and with anomalies data-set.

Characteristics like these are presents in the data that describe objects of free geometry.

In general, a method of center selection that does not take into account the intrinsic geometry in the data can result in a poor accuracy related to geometrical details. The purpose of center selection procedures is the point's concentration at regions with geometry highly detailed and the decrease of these at geometrically simple regions, in order to obtain an adequate-precision interpolation process.

1. RADIAL BASIS FUNCTIONS INTERPOLATION

The problem of interpolation of scattered point clouds can be posed formally as [5]:

Given a set of distinct points $X = \{x_i\}_{i=1}^N \subset \mathbb{R}^3$ and a set of scalars find an interpolant such that:

$$s(x_i) = f_i \quad i = 1, \dots, N \quad (2)$$

Note that the simplified notation $X = (x, y, z)$ is used for the image point $X \in \mathbb{R}^2$.

The main idea of the radial basis functions is to select a function (usually called radial basis function) ϕ , and a norm $\|\cdot\|$ in \mathbb{R}^d , such that the interpolant s is calculated as follows:

$$s(x_i) = p(x) + \sum_{i=1}^N \lambda_i \phi(\|x - x_i\|) \quad (3)$$

where $p(x)$ is a linear low degree polynomial, the coefficients λ_i are real numbers, $\|\cdot\|$ is usually the Euclidean norm in \mathbb{R}^3 and x_i is the center of interpolation.

2. HIERARCHICAL CURVATURE-BASED CLUSTERING

An interesting alternative to dealing with the mathematical limitations in center selection field is including some center's heuristic-based searches. Clustering methods are widely used in computer vision for classifying unattended data according to a given criterion. A few of the most widely used clustering methods are k-means and hierarchical clustering.

The proposed method uses a hierarchical clustering implementation with the Euclidean distance and curvatures estimation as similarity measure. It works in two phases as follows: the algorithm starts by assigning each point into a group, from each group calculates the mean or centroid (*for the first iteration it is the same point*), then, it looks for the two groups whose similarity measure is minimal, Once they are identified the two closest groups proceed to join them and update the resulting group average, this procedure is repeated until it reach the level of α clustering (see algorithm 1).

The search for the closest groups and the closest points to the centroid of the groups is done by using the nearest neighbor algorithm implemented with *kd-tree* data structures which enable to find closest points in dense clouds points with low response times.

One of the most important parameters of the algorithm is the α parameter which allows to establish the desirable degree of grouping, i.e. when the parameter α has a value of 30 means that only 30% of the data are grouped, we should clarify that although only an alpha percentage of the data is clustered, the algorithm maps the full data set uniformly; making it in the case of the above example achieving a reduction of 70% of the image points. This parameter is user-defined and adjusts the degree of reduction. It is related to achieve greater or lesser centre's selection quantity.

The curvature as criteria selection of centers focuses on finding the pair of points whose curvatures have a very high level of similitude and that

the aforementioned points are very close, of more formal way it is:

Once a group G_i is given, a point p_1 and his nearest neighbor p_2 , each of which has their corresponding curvatures c_1 and c_2 , say that $(p_1, p_2) \in G_i$, if:

$$1 - \varepsilon \leq \frac{c_1}{c_2} \leq 1 + \varepsilon$$

where ε is a user-introduced value of tolerance.

2.1 Curvature Estimation

In the area of computer vision and 3D image reconstruction have been used geometrical invariant characteristics of the surfaces and images for pattern recognition and images characterization. The invariant feature of differential geometry most commonly used for these purposes is the curvature.

The curvature is one of the simplest and one of the most important properties of the curve, which can be defined for a curve according to [2] as:

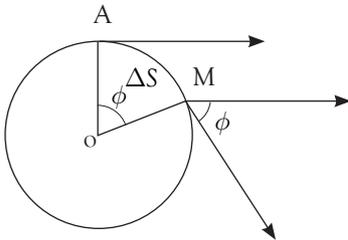


Figure 1. Angle ϕ between the tangents.

Source: [2]

Denote γ as a smooth curve in \mathbb{R}^3 . Let the point A in γ and M a closed point to A. The angle ϕ between the tangents (see figure 1) of these points expresses the directions' curve variation related to the arc A to M. The average variation is $\frac{\phi}{\Delta s}$, where the arc-length is Δs . Then the curvature is when $\Delta s \rightarrow 0$ is defined like:

$$k = \lim_{\Delta s \rightarrow 0} \left(\frac{\phi}{\Delta s} \right) \quad (4)$$

In order to estimate the surface curvature in a point, we used the approximation described in [4, 6] where given a covariance matrix defined over a closed neighborhood $N(p) = \{p_1, \dots, p_x\}$:

$$M = \sum_{i=1}^N (p_i - \bar{p})(p_i - \bar{p})^T \quad (5)$$

where, \bar{p} is defined thus:

$$\bar{p} = \frac{1}{n} \sum_{i=1}^n p_i$$

The eigenvalues λ_i measure the variation of p_i along the direction of the corresponding eigenvectors. Assuming $\lambda_0 \leq \lambda_1 \leq \lambda_2$ it follows that the plane

$$T(x) : (x - \bar{p}) \cdot V_0 = 0$$

Through \bar{p} minimizes the sum of squared distances the neighbors of p. Thus V_0 approximates the surface normal in p, and

$$C = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} \quad (6)$$

Quantitatively describe the variation along the surface normal. Its approximation have been used [3, 7].

3. RESULTS

The experiments were carried out using mathematical functions and real images for this way to obtain a good validation of the method.

The algorithms were implemented in C++ using the library ANN to estimate the nearest neighbor's search. All the results and the times of executing for each one of the tests were obtained using a processor Intel® Core Duo 1.66 GHz and 2Gb Ram size.

```

Centroids = Hierarchical_Clustering ( $\alpha$ ,  $n$ , datos)
Calculating Curvatures ( $n$ , datos)
Assign each point to a group
Update the group means
Limit = totalGroups *  $\alpha$  / 100
While %Groups > 1 & %Groups > Limit
  x, y = ClusteringCriteria ( $\epsilon$ , datos, curvatures)
  Join (x, y)
  Update the resulting group average
  Update the %Groups
End While
End

Curvatures = Calculating_Curvatures ( $r$ , data)
For each data
   $p_i$  = data(i)
  Neighborhood = getNeighborhood ( $p_i$ )
   $\bar{p}$  = CalculateAverage (Neighbourhood)
   $M = \sum_{i=1}^N (p_i - \bar{p})(p_i - \bar{p})^T$ 
  Curvatures ( $i$ ) =  $\frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}$ 
End For
End

[x, y] = ClusteringCriteria ( $\epsilon$ , data, curvatures)
shortDist[ pos] = nearestNeighbor (0, data)
For each data
  [tmpDist, tmpPos]= nearestNeighbor (i, data)
  Likeness = likenessCriteria(curvatures(pos),
  curvatures(tmpPos),  $\epsilon$ )
  If tmpDist < menDist & Likeness = True
    menDist = tmpDist
    pos = i
    x = i
    y = tmpPos
  End If
End For
End

S = likenessCriteria (curvature1, curvature2,  $\epsilon$ )

$$s = \begin{cases} 1 & \text{if } 1 + \epsilon > \frac{\text{Max}(\text{curvature1}, \text{curvature2})}{\text{Min}(\text{curvature1}, \text{curvature2})} > 1 + \epsilon \\ 0 & \text{Otherwise} \end{cases}$$

End

```

Figure 2. Algorithm 1. Procedure for center selection with distance-curvature criteria

Source: The authors

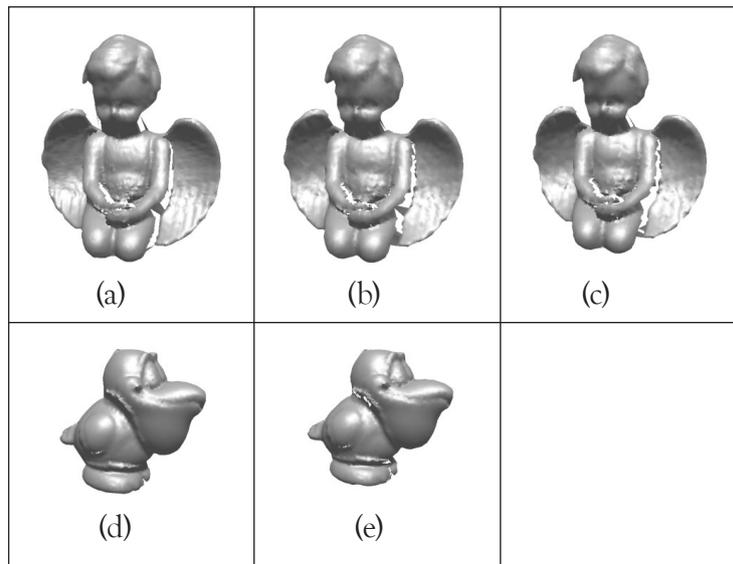


Figure 3. Visual results of point reduction. The 40, 50, 60 percent for Angel model (a-c), and bird model (d-f).

Source: The authors

4. ERROR BEHAVIOR

The procedure to calculate the interpolation error it is based on the calculation of the average quadratic error between the original group of data and the group of interpolated data, as it is shown in the equation 7:

$$s = \frac{1}{N} \sqrt{\sum_{i=1}^N x_i - s(x_i, y_i)} \quad (7)$$

Note that the error in this case is estimate only in the z component, thus, $s(x_i, y_i)$ is the interpolation result of a not center points.

In the figure 3a-c, the Angel model was used, for a fixed ratio-value the obtained error was 0.32, 0.41 and 0.47 for a α -value of 40%, 50 and 60%. Similar result was obtained with Bird models, where the errors were 0.24, 0.36 and 0.4. A general average behavior of error relates with alpha reduction value of a few mathematical function is shown in figure 4. The continue line show the average values error and a fitting by polynomial regression is describe by dash-dot line. As expected to high reduction percent the error increasing. However, the curvature based method show a smooth error value increment and avoids a lineal accuracy drop.

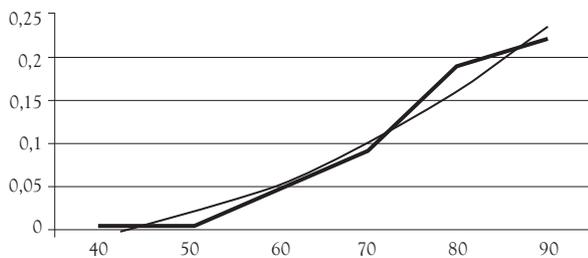


Figure 4. Average error behavior

Source: the authors

6. CONCLUSIONS

In this paper is described a hierarchical curvature-based method to point reduction for center

interpolation selection. This method uses an implementation of hierarchical clustering using similarity measurement both the Euclidean distance and curvature criterion to improve the selection result.

The advantage of the proposed method toward a classic hierarchical procedure is that permit to analyze the intrinsic geometric inside de data. It permits that in neighborhoods with a high variation in the curvature, including most points, and those regions less curved fewer points are select. This allows model's representation with a multi-resolution. Some loss of high detailed regions was observed. However, the general geometry of model was maintained.

It is important to note that appropriate choice of parameters such as radio and tolerance that are used influences the final representations. A relationship of these parameters with the final accuracy is still an open issue. Additionally, no noise treatment was introduced in the procedure. An option to face the noise is to improve the curvature estimation using WPCA instead PCA.

REFERENCES

- [1] G. Sánchez, y J. Branch, "Ajuste de superficies de objetos 3D de forma libre a partir de datos de rango utilizando técnicas de triangulación", presentado a Encuentro de Investigación sobre Tecnologías de Información Aplicadas a la Solución de Problemas - EITI, Medellín, 2004, vol.
- [2] V. Toponogov, *Differential Geometry of Curves and Surfaces: A Concise Guide*, Berlin: Birkhauser, 2006,
- [3] I. Jolliffe, *Principle Component Analysis*, New York: Springer-Verlag, 1986,
- [4] J. C. Carr *et al.*, "Reconstruction and representation of 3D objects with radial basis functions", presentado a Proceedings of the 28th Annual Conference on Computer Graphics and interactive Techniques SIGGRAPH '01, New York, 2001, vol.
- [5] M. Pauly *et al.*, "Efficient simplification of point-sampled surfaces", presentado a Proceedings of IEEE visualization 2002, 2002, vol.

- [6] M. D. Buhmann, *Radial Basis Functions: Theory and Implementations*, Cambridge: Cambridge University Press, 2003,
- [7] G. Sanchez *et al.*, “Reconstrucción de objetos de topología arbitraria mediante selección de centros para la interpolación con fbr”, *DYNA*, vol. 73, no. 150, pp. 189-201, 2006.