Robust Method for Estimating Normals on Point Clouds Using Adaptive Neighborhood Size

E.A. Leal and N.E. Leal

Faculty of Engineering, Universidad Autónoma del Caribe, Barranquilla Colombia

ABSTRACT

Normal estimation on sampled curves or surfaces is a basic step of many algorithms in computer graphics, computer vision, and especially in recognition and reconstruction of three dimensional objects. This paper presents a simple and intuitive method for estimating normals on point based surfaces. The method is based on Robust Principal Component Analysis (RPCA) therefore is capable to deal with noisy data and outliers. In order to estimate an accurate normal on a point, our method takes a neighborhood of variable size around the point. The neighborhood size depends on local properties of the sampled surface. It is shown that the estimation of the tangent plane on a point is more accurate using a neighborhood of variable size than using a fixed one.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Line and Curve Generation).

1. Introduction

Algorithms for surface reconstruction, surface segmentation and edge extraction on point based surfaces, needs accurate estimation of the normals. Estimation of normal is important for detecting and extracting geometric features and preprocessing steps in surface reconstruction. In order to estimate the normal, is necessary filter the noise of the points; although, this causes loss of the sharp features in the surface. To avoid loss of sharp features is important design good algorithms capable of estimate accurate normal and preserve sharp feature even on noisy data.

The normal estimation problem has been addressed by different research areas, such as computer graphics, mathematics and image processing. A common step is estimating a local neighborhood around each sampled point, approximating the local surface at each point. Several algorithms used to surface reconstruction, takes a fixed neighborhood size, which are determined experimentally, [MVDe03], [ABCO*01], [HDeD*92], ignoring, the variation of curvature and the sampling density around each point in the cloud.

A way to get a better normal approximation and therefore the tangent plane is not taking the neighborhood size fixed, but variable, and depending on intrinsic characteristics on the surface from which the sample was taken.

This paper presents a simple method for normal estimation on point based surfaces. The proposed method takes into account the noise in the point cloud, the curvature variation to each point, and the neighborhood size. The proposed method also analyzes the normal estimation on sharp features, like corners and edges, achieving a better estimating of the normal.

2. Related Work

Many researchers have used a fixed size of neighborhood, or k nearest points [PGK02], [HDeD*92], [ABCO*01], [MVDe03], [FCS04], [LSK*10] to estimate the normal vector at a given point.

The pioneer work of Hoppe et al. [HDeD*92] for surface reconstruction, fits a tangent plane by least-squares, at each point of the cloud, using k nearest neighbors, the value of k is selected experimentally. This method fails in point clouds with presence of noise and outliers, because the tangent plane is estimates using a classical PCA.

Mitra and Nguyen [MN04], made eigen analysis of points belonging to each neighborhood, their analysis involves the curvature, sampling density and noise, thus achieving an adaptive neighborhood size estimation, the approach depends on parameters entered by users, which makes the algorithm depends on the intrinsic features strongly which are found in different data sets, the PCA approach is not robust to noise and outliers.

Several Works [ABCO*01], [MVDe03], [FCS04], have used a combination of linear and polynomial approximations, these approaches use the Moving least-squares (MLS) [Lev98], which makes them robust to the noise and outliers, therefore can estimate good normals. MLS-based methods have the inconvenient of over smooth the surfaces they process, Mederos et al [MVDe03] and Fleishman et al [FCS04] solve this problem, although they need to tune local kernel parameters, which is crucial for final results.

Another drawback is that MLS-based methods are computationally expensive, these approaches uses also a fixed neighborhood size. In [LSK*10] resolve the problem introducing a kernel density, this method is robust to outliers and noise, producing good approaches to normal estimation.

Methods based on triangular meshes, the normal estimating on a vertex V_i, is derived from the average of the normals to each faces or triangles adjacent to vertex V_i. Approaches like [Tau02], [Van02], [WKWL02] go in this direction. In [PSK*02], have proposed an improvement to the standard estimation method on triangular meshes, called "Normal Vector Voting", using a geodesic neighborhood of triangles, around the vertex V_i. In general these methods have the disadvantage of being sensitive to noise, also imply the process of building a mesh of triangles over the surface. Another disadvantage is that such methods do not describe a smooth and continuous surface.

Delaunay/Voronoi, are based on geometrical concepts, several approaches use this concepts [DS06], Amenta y Bern [AB99]. The methods to estimate normals from noise free point clouds, use the idea of pole; specifically, these approaches reveals that the farthest Voronoi vertices from poles, provide a good estimating of normal. Dey and Sun [DS04], extends the earlier idea of poles to noisy data. Ouyang and Feng [OF05] build a Voronoi local mesh based on 3D Voronoi diagrams and proposes a heuristic rule of mesh growing, to fit quadratic curve groups through which are obtained the tangent vectors. These methods do not describe how to treat normals at edges and corners.

The paper is organized as follow: in section 3 we presents the mathematical foundations and explain the proposed method in detail; in section 4 we show the results and illustration how the algorithm work; in section 5 our conclusion are presented.

3. Proposed Method

Our method is roughly made up of two stages: the first one finds the suitable size of the neighborhood to estimate the normal on each point. This process begins with a neighborhood of size equal to 4, and iteratively increases the size as long as a pre-established criterion is met. Then fits a regression plane to the neighborhood and estimate the normal. The second one corrects normals estimated in on critical points located in sharp features like borders, edges and corners where the normal cannot be estimated correctly. In next subsections, our method is explained in more detail.

3.1 Problem Statement

Given a set of points $P = \{(x, y, z)^T\} \subset R^3$, which probably were sampled with noise from an unknown curve or surface S, i.e. $z_i = f(x_i, y_i) + \varepsilon_i$, where ε_i, is noise in the sampled points.

The goal is estimate the normal at each point $p_i = (x_i, y_i)$ and the error estimation is minimum respect to the true normal.

We start choosing a k-nearest neighbor to each p_i in the cloud. Once the neighborhood has been obtained, the next step is to fit a plane or curve to surface, using variation of "Robust Principal Component Analysis" RPCA, as shown below.

$$CM_w = \frac{1}{n-1}\sum_{i=1}^{n}(p_i - \overline{p}_w)(p_i - \overline{p}_w)^T W \quad (1)$$

$$\overline{p}_w = \frac{\sum w_i p_i}{\sum w_i} \quad (2)$$

$$w_i = \frac{1}{d_i \cdot \sum_{j=1}^{n} \frac{1}{d_j}} \quad (3)$$

Where CM_w in (1), is the weighted covariance matrix, $W = \{\sqrt{w_1}, \ldots, \sqrt{w_n}\}$, are the weights associated to each point in the neighborhood of p_i. $\overline{p}_w$ in (2) is the weighted mean. The weights w_i in (3) are estimated using a proportional inverse distribution, where d_i, is the Euclidean distance between the point p_i and each neigborhood point.

We obtain a robust version of PCA, called RPCA, the robust PCA handled noise and outliers contain in the point clouds, which not depend on user entry parameter how is used to be in classical RPCA approaches[HRB05].

3.2 Size of neighborhood

The choice of neighborhood size can directly affect the estimation of normal at point $p_i \in S$. When we choose a large k-neighborhood, a large surface area is approximate, producing a very smooth surface estimation, resulting in loss of fine details and lack of distinction between the separations of two adjacent surfaces. On the other hand, if a small neighborhood is taken, this leaves an inaccurate estimation of the normal, mainly close to sharp features like corners, crest and edges. This uncertainty is even greater when the sampled data are noisy, because it is necessary to take a larger neighborhood to reduce the noise, results in loss of fine details.

3.3 Adaptive neighborhood

For a better understanding of method, and without loss of generality, we use a manifold in R^2 instead one in R^3. A generalization of this analysis can be made directly for R^3.

To obtain a good approximation of the normal, must also take into account the surface curvature at each point. In order to find the relationship between the size of the neighborhood and the curvature at point p_i, the following analysis will be made according to the definition of curvature given in [AKL74].

The curvature in a particular point *O*, is stated by (4)

$$k = \lim_{l \to 0} \frac{2h}{l^2} \quad (4)$$

where h is the distance from a second point p_c, on the curve to tangent space to *O* and *l* is the length of the tangent segment from the point *O* to the projection of point p_c onto the tangent plane (Figure 1).

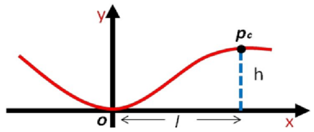


Figure 1: *Curvature approximation at point O.*

Based on above, we choose a rectangular coordinate system such that the origin *O*, is matching with p_i, (point in the curve which the normal will be estimated), and the *OX* axis is matching with the tangent space (Figure 2), note that $y' = 0$ and $k = |y''|$, spanned the function $y = f(x)$, using second order Taylor expansion, (since *f(x)* is unknown), we obtain $y = \frac{1}{2} y'' x^2 + \varepsilon x^2$, let $y' = 0$. We have $\varepsilon \to 0$, when $x \to 0$, then $k = |y''| = \lim_{x \to 0} 2|y| / x^2$, therefore as $|y| = h$, and $x^2 = l^2$ implies that $k = \lim_{l \to 0} \frac{2h}{l^2}$.

We can adapt the above analysis, to our neighborhood size problem as shown in Figure 2.

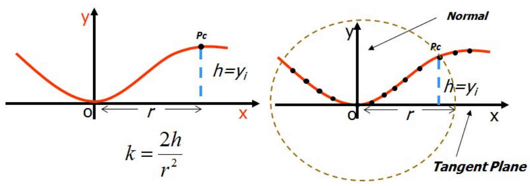


Figure 2: *Relationship between curvature and neigborhood at point p_i*

Figure 3 illustrates an intuitive idea of the relationship between the projections of each point (belonging to the neigborhood), into the plane and its right size.

We observed, for highest curvatures, the neighborhood radius r_i, is much smaller than the projection of a point y_i on the curve, to plane, i.e. $(r_i < y_i)$, the opposite happen when the curvature is smooth, is observed that the projection of point y_i, is much smaller than the radius r_i, i.e. $(r_i > y_i)$.

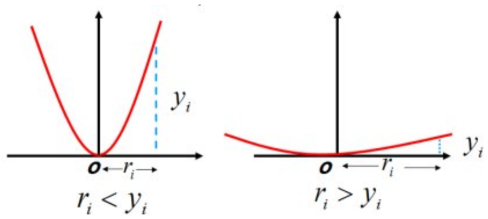


Figure 3: *neigborhood size depends on the curvature.*

From the above relations, we can find the mathtematical expression, which leads us to obtain a variable neighborhood size. Having obtained the eigenvectors from the covariance matrix, Our next step is projecting the vectors $(p_i - \overline{p}_i)$ formed by each p_i of the neighborhood and its mean point $\overline{p}$ to the tangent plane. We denoted to simplify $P_{roj}(p_i - \overline{p}_i)$. Figure 4 show the vectors projection.

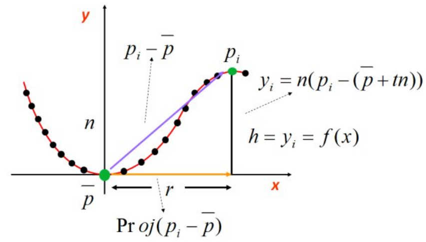


Figure 4: *Vector projection on the tangent plane*

We are looking for an expression that relating the curvature at the origin ***O***, with heights $y_i = n(p_i - (\overline{p} + tn))$ of p_i and the radius of neighborhood, $r_i = \max \| P_{roj}(p_i - \overline{p}) \|$. Where *n* is the normal vector and *t*, is displacement in the direction of normal, the above is made with the finality of approximate; the mean of neigborhood to the surface point p_i.

To find the expression, we must start with the relation from Section 3.2. i.e., $k = \frac{2h}{l^2} = \frac{2y}{r^2}$, now assuming that the curvature in the neighborhood is small, namely $r_i >> y_i$, we can establish that $| y_i | \leq k \frac{r_i^2}{2}$, to ensure, that high curvature points, the neighborhood must shrink until r_i is greater than y_i in the range $r_i \in [-r, r]$. The right hand side of above expression, identify a parabola, which are a quadratic approximation of curve on the origin ***O***. $|y_i| \leq k\frac{r_i^2}{2}$ can be used to establishing a adaptability criterion, that produce a neighborhood size variation depending on the curvature at each point on the cloud.

Now carrying the above expression to our notation, we obtain (5):

$$\| n(P - (\overline{P} + tn)) \| \leq k/2 \| \Pr oj(P - \overline{P}) \|^2 \quad (5)$$

The curvature k (maximum local curvature), is approximated using (6) [PGK02].

$$k = \frac{\lambda_0}{\lambda_0 + \lambda_1}, \text{with } \lambda_0 \leq \lambda_1 \quad (6)$$

and λ_0, λ_1, are eigenvalues of covariance matrix CM_w; This method give us a local curvature estimation (not very accurate).

The expression (5), operates in the following way, it takes a neighborhood large enough, if inequality is not satisfied, we proceed to remove one point from the neighborhood, until the inequality be satisfied. The result of this iteration is the right size of neighborhood at each point p_i.

3.4 Normal Estimation on Sharp Features

Fitting algorithms base on PCA have a drawback when estimate normals on sharp features like edge or corners. The drawback is presented because there is a discontinuity in the derivative of the surface on sharp features, and these algorithms estimate the normal in the same way they do on smooth surfaces, i.e. computing the mean of points located in adjacent sides to the discontinuity, instead of using only points located just in one side

3.4.1 Critical points detection

Critical points can be identified trough principal component analysis, the covariance matrix CM_w have three eigenvalues $\lambda_0 \leq \lambda_1 \leq \lambda_2$ and their correspondent eigenvectors v_0, v_1, v_2, for which it holds that v_0 is an approximation to the normal on p_i, and λ_0 is close to zero, but if p_i is located on a sharp feature λ_0 is large compared to λ_2 therefore the ratio between λ_0 and λ_2 can be used to estimate the probability that p_i lie on a sharp feature i.e. if $\lambda_0 / \lambda_2 > \varepsilon$ then p_i is a critical point. Where ε is a given constant.

Once found the critical points, the next step is to identify the type of feature (corner, edge, or border) where the point is located. An detailed explanation can be founded in [GWX01].

3.4.2 Suitable Neighborhood on Sharp Features

As it was mentioned before in section 4, in order to estimate a good normal on a sharp feature, the neighborhood of points to calculate the normal must be conformed only by points belonging to one of the adjacent sides to the critical point. To do this, a k-means clustering analysis is made for knowing to which side of the sharp feature belongs to the critical point. If the sharp feature is an edge or border, will be selected two clusters, if it is a corner, will be selected 3 clusters. Once established the cluster of the critical point, a regression plane is fitted to the cluster (neighborhood) and the normal to the plane will be the normal on the critical point. In this way the normals on the critical points are fixed.

4. Analysis and Results

In this section, it is replicated the work made in (Mitra, 2003), to show the efficiency of our method. We started sampling a set of points, from the curve $y = k\frac{x^2}{6}$, in the range $x \in [-1,1]$ using different k values (curvature).

The direction of true normal is the "y" axis, i.e. [0 1], we can calculate the error between the true normal and the estimated normal like the angle between them. Additionally we add noise to sampled points, in "y" direction. Figure 5 shows, the error as neighborhood size function $r_i = \max \| P_{roj}(p_i - \overline{p}) \|$, where n = 0.05 (noise), for three different k values, k = 0.4, 0.8 and 1.2, the graph shows as the radius is growing (more neighbors), as the error increases. Figure 5 shows values for r > 0.2

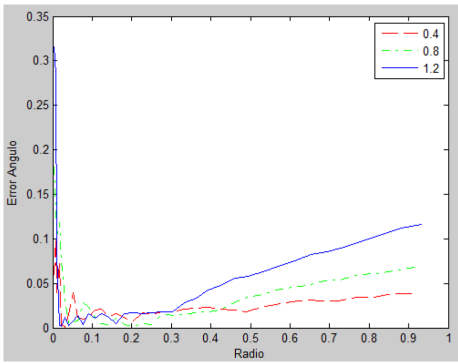


Figure 5: *The normal estimating error grows as r increase. r > 0.2*

Figure 6, Indicates the estimated error as function of neighborhood size r. For small r values, when k = 1.2, and n = 0.017, 0.033, and 0.05. It is noted, the error tends to decrease as r increases, r <0.08. This confirms that neighborhood size depends on features found around a point.

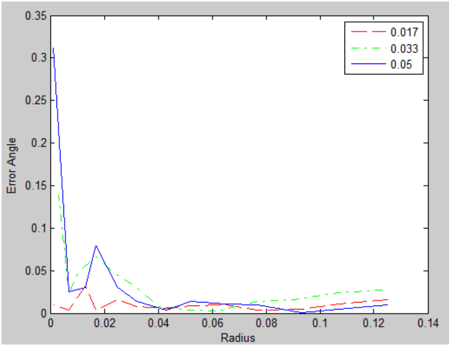
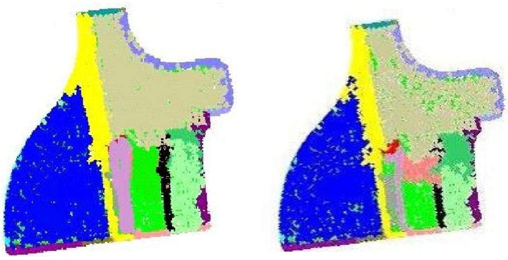


Figure 6: *The normal estimating error decrease, as n decrease and increase r. r > 0.2*

Figure 6 shows there is always a range in which the number of points is the adequate, to keep the normal estimating error below a threshold, outside of range, both above and below, the error tends to increase. This happens in areas with a high curvature. The right size of neighborhood is given by the normal estimating method, presented in this paper.

Our method efficiency is shown, when is applied to surface segmentation algorithm, based on normals, can be seen in Figure 7a, how the use of variable neighbor-hood and RPCA, produces in areas of edges and corners, better boundaries surface definition, compared to normal estimating, with fixed neighborhood size and simple PCA Figure 7b.



(a) (b)

Figure 7: *Normal estimation in fandisk (a) proposed method (b) not using proposed Method*

4.1 Conclusions

This paper has presented a simple and robust method for estimating normals in noisy point clouds with sharp features. The method uses a robust version of principal component analysis RPCA, which does not require extra parameters entered by the user. The proposed method uses a neighborhood sized variable, around each point in the cloud. The eighborhood size depends on the local surface characteristics that representing the point cloud.

It is shown numerical and visual results, about the validity of our method. For example, how the estimation error, decrease when neighborhood size is variable as opposed to fixed one.

The proposed method also takes into account the preservation of fine details such as edges and corners as well as the correct estimation of normal. Future work should be directed to incorporate a statistical model of noise that gives robustness and validity to our method.

References

[AKL74] Aleksandrov A.D., Kolmogorov A.N., Laurentiev, M.A. et al.: *La matemática: su contenido, métodos y significado, volumen 2*. Alianza Editorial, Madrid 1974.

[ABCO*01] Alexa M., Behr J., Cohen-Or D., Fleishman S., Levin D., Silva C.T.: Computing and rendering point set urfacees. *IEEE Transaction on Visualization and Computer Graphics, 9(1):3-15,* 2003.

[AB99] Amenta N., Bern J.: Surface Reconstruction by Voronoi Filtering. *Discrete Computational Geometry*. 22 (1999), 481-504.

[DS06] Dey T.K., Sun J.: Normal and Feature Approximation from Noisy Point Clouds. Lecture Notes in Computer Science, 2006, vol. 4337/2006, 21-32.

[DS04] Dey T.K., Goswami S.: Provable Surface Reconstruction from Noisy Samples. *Proc. 20th Annual Symposium on Computational. Geometry*. (2004), 330-339.

[FCS05] Fleishman S., Cohen-Or D., Silva C.T.: Robust Moving Least-squares Fit-ting with Sharp Features. *Proceeding of SIGGRAPH, TOG,* vol. 24 Issue 3. July, 2005.

[GWM01] Gumhold S., Wang X., Macleod R.: Feature Extraction From Point Clouds. *Proceedings, 10th International Meshing Roundtable, Sandia National Laboratories*, pp.293-305, October 7-10 2001.

[HDeD*92] Hoppe H., Derose T., Duchamp T., Mcdonald J., Stuetzle W.: Surface Reconstruction from Unorganized Points. *Computer Graphics*, 26(2):71-78, 1992.

[HRB05] Hubert, M., Rousseuw, P.J., Branden, K.V., 2005. ROBPCA: A New Approach to Robust Principal Component. *Analysis. Technometrics*, February, vol. 47, No. 1.

[Lev98] Levin D.: The approximation Power of Moving Least-squares. *In Math. Comp.* vol. 67, No. 224, pages 1517-1531, 1998.

[LSK*10] Li B., Schnabel R., Klein R., Cheng Z., Dang G., Jin S.: Robust normal estimation for point clouds with sharp features. *Computer & Graphics* 34(2010) 94-106.

[MVDe03] Mederos B., Velho L., De figueiredo L. H.: Robust Smoothing of Noisy Point Clouds. *In Proc. SIAM Conference on Geometric Design and Computing* (Seattle, USA), Nashboro Press. 2003.

[MN04] Mitra N.J., Nguyen AN., Guibas L.: Estimating Surface Normals in Noisy Point Cloud Data. *International. Journal* *of Computational Geometry and Applications*, (2004), vol. 14 pp. 261-279.

[OF05] OUYANG D., FENG H.Y.: On the normal vector estimation for point cloud data from smooth surfaces. *Computer-Aided Design* 2005; 37(10):1071–9.

[PSK*02] PAGE D.L., SUN Y., KOSCHAN A.F., PAIK J., ABIDI M.A.: Normal Vector Voting: Crease Detection and Curvature Estimation on Large, Noisy Meshes. *Journal of Graphical Models Elsevier,* pp. 199-229, vol. 64, 2002.

[PGK02] PAULY M., GROSS M, KOBBELT L.: Efficient Simplification of Point-Sampled Surfaces. *In Proceedings of IEEE Visualization,* pp 163-170. 2002.

[Tau02] TAUBIN G.: Estimating the Tensor of Curvature of a Surface From a Polyhedral Approximation, *in Proceedings of the Fifth International Conference on Computer Vision*, 1995, pp. 902–907.

[Van02] VANCO M.: *A Direct Approach for the Segmentation of Unorganized Points and Recognition of Simple Algebraic Surfaces*. PhD Thesis. University of Technology Chemnitz 2002.

[WKWL02] WOO H., KANG E., WANG, SEMYUNG KWAN., LEE H.: A New Segmentation Method for Point Cloud Data. *Interna-tion-al Journal of Machine Tools & Manufac-ture. Pergamon, Elseivier Science*, pp. 167-178, 2002.