

## A Nonparametric Approach for Noisy Point Data Preprocessing

Yongjian Xi<sup>1</sup>, Ye Duan<sup>1\*</sup>, and Hongkai Zhao<sup>2</sup>

<sup>1</sup>University of Missouri, Columbia, MO 65211, USA

<sup>2</sup>University of California at Irvine, Irvine, CA, 92710, USA

---

**Abstract** – 3D point data acquired from laser scan or stereo vision can be quite noisy. A preprocessing step is often needed before a surface reconstruction algorithm can be applied. In this paper, we propose a nonparametric approach for noisy point data preprocessing. In particular, we proposed an anisotropic kernel based nonparametric density estimation method for outlier removal, and a hill-climbing line search approach for projecting data points onto the real surface boundary. Our approach is simple, robust and efficient. We demonstrate our method on both real and synthetic point datasets.

**Keywords:** Point Preprocessing, Outlier Removal, Surface Reconstruction, Parzen Window, Anisotropic Kernel Density Estimation

---

### 1. Introduction

Despite significant advancement in interactive shape modeling, creating realistic looking 3D models from scratch is still a very challenging task. Recent advancement in 3D shape acquisition systems such as laser range scanner, structural light system and stereo vision system have made direct 3D data acquisition feasible [Seitz 2006].

The obtained point dataset however can be quite noisy. Without preprocessing to deal with noisy data, parametric and/or variational formulations [Hoppe 1992, Bajaj 1995, Curless 1996, Hilton 1998, Whitaker 1997, Zhao 2000, Zhao 2001, Carr 2001, Ohtake 2003, Xie 2003, Levin 2003, Alexa 2003] are usually used. These methods generally composed of both a fitting term for the data and a regularization term for the reconstructed surface. Since all data points, even outliers, are treated equally and can affect the final reconstruction, these approaches will likely fail for highly noisy data.

Another type of algorithms is based on popular computational geometry algorithms such as Delaunay triangulations and Voronoi diagrams to construct triangulated surfaces [Amenta 1998a, Amenta 1998b, Amenta 1999, Edelsbrunner 1998, Boissonnat 1984]. For this kind of method, it is challenging to find the right connections among all data points in three and higher dimensions, especially for noisy data.

Recently, Medioni et al. proposed the tensor voting method [Medioni 2000], which is a nice feature extraction algorithm. By designing an appropriate voting procedure among all data points, a tensor field and an associated

saliency field can be constructed. Coherent geometric information can be extracted from the tensor field and the saliency field. However, tensor voting method is computationally very expensive.

In this paper, we propose a nonparametric approach for noisy point data preprocessing. In particular, we proposed an anisotropic kernel based nonparametric density estimation method for outlier removal, and a hill-climbing line search approach for projecting data points onto the real surface boundary. Our approach is simple, robust and efficient. We demonstrate our method on both real and synthetic point datasets.

### 2. Algorithm

Parzen-window based kernel density estimation

Points outside the object surface are outliers that have to be removed. Since the real object surface is unknown, it is hard to specify a general criterion to detect outliers. In this paper, we propose to employ parzen-window based nonparametric density estimation method for outlier removal.

Parzen window based kernel density estimation is the most popular nonparametric density estimation method. In order to make the paper self-contained, we will briefly review the Parzen window method in the following. For a complete description, please refer to the book by Duda et al. [Duda 2000].

Given  $n$  data points  $x_i, i = 1, \dots, n$  in the  $d$ -dimensional Euclidean space  $R^d$ , the multivariate kernel density estimate obtained with kernel  $K(x)$  and window radius  $h$ , computed in the point  $x$  is defined as:

$$f(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (1)$$

---

\*Corresponding author:  
Tel: +573-882-3951  
Fax: +573-882-8318  
E-mail: duanye@missouri.edu

Without loss of generality, let's assume  $h = 1$  from now on, so we could simplify Equation (1) as:

$$f(x) = \frac{1}{n} \sum_{i=1}^n K(x-x_i) \quad (2)$$

$K(x)$  is generally a spherically symmetric kernel function satisfying:

$$K(x) = C_{k,d} k(\|x\|^2) \quad (3)$$

where  $k(x)$  is the profile function of the kernel  $K(x)$ .  $C_{k,d}$  is the normalizing constant such that

$$K(x) \geq 0 \quad (4)$$

and

$$\int_{R^d} K(x) dx = 1, \quad (5)$$

$\|x\|$  is the  $L_2$  norm (i.e. Euclidean distance metric) of the  $d$ -dimensional vector  $x$ . Employing the profile function notation, Equation (2) can be further rewritten as

$$f(x) = \frac{C_{k,d}}{n} \sum_{i=1}^n K(\|x-x_i\|) \quad (6)$$

There are three types of commonly used spherical kernel functions  $K(x)$ : the Epanechnikov kernel, the uniform kernel, and the Gaussian kernel. The Epanechnikov kernel is defined by the profile function  $k_E(x)$ :

$$k_E(x) = \begin{cases} 1-x & 0 \leq x \leq 1 \\ 0 & x > 1 \end{cases} \quad (7)$$

The uniform kernel is defined by the profile function  $k_U(x)$ :

$$k_U(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & x > 1 \end{cases} \quad (8)$$

The Gaussian kernel is defined by the profile function  $k_N(x)$ :

$$k_N(x) = \exp\left(-\frac{1}{2}x\right) \quad x \geq 0 \quad (9)$$

### 2.2 Outlier removal by anisotropic ellipsoidal kernel

For 3D point cloud obtained by laser scan or stereo vision, the outliers tend to spread in the space randomly, while “real” (we use a quotation here to emphasize the fact that the real surface is unknown) surface points will spread along a thin shell which encloses the real surface object. In other words, the distribution of the outliers is relatively isotropic, while the distribution of the real

surface points is rather anisotropic. Hence in this paper, we propose to employ an anisotropic ellipsoidal kernel based density estimation method for outlier removal.

For anisotropic kernel, the  $L_2$  norm  $\|x-x_i\|$  in Equation (6), which measures the Euclidean distance metric between two points  $x$  and  $x_i$  will be replaced by the Mahalanobis distance metric  $\|x-x_i\|_M$ :

$$\|x-x_i\|_M = ((x-x_i)^t H^{-1} (x-x_i))^{1/2}, \quad (10)$$

here  $H$  is the covariance matrix defined as:

$$H = DD^T \quad (11)$$

and

$$D = (x_1 - x, x_2 - x, \dots, x_n - x). \quad (12)$$

Geometrically,  $(x-x_i)^t H^{-1} (x-x_i) = 1$  is a three-dimensional ellipsoid centered at  $x$ , with its shape and orientation defined by  $H$ . Using Single Value Decomposition (SVD), the covariance matrix  $H$  can be further decomposed as:

$$H = UAU^T \quad (13)$$

with

$$A = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \quad (14)$$

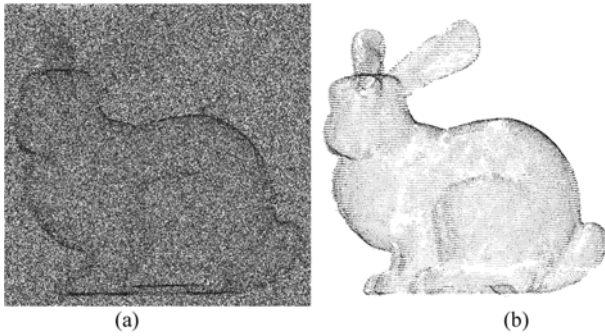
$\lambda_1 \geq \lambda_2 \geq \lambda_3$  are the three eigenvalues of the matrix  $H$ , and  $U$  is an orthonormal matrix whose columns are the eigenvectors of matrix  $H$ . We will detail the SVD-based decomposition (Equation (13)) of the covariance matrix  $H$  in the Appendix.

To compute the anisotropic kernel based density, we will apply an ellipsoidal kernel  $E$  of equal size and shape on all the data points. The orientation of the ellipsoidal kernel  $E$  will be determined locally. More specifically, given a point  $x$ , we will calculate its covariance matrix  $H$  by points located in its local spherical neighborhood of a fixed radius. Without loss of generality, we will assume the radius is 1 (which can be done by normalizing the data by the radius). The  $U$  matrix of Equation (13) calculated by the covariance analysis is kept unchanged to maintain the orientation of the ellipsoid. The size and shape of the ellipsoid will be modified to be the same as the ellipsoidal kernel  $E$  by modifying the diagonal matrix  $A$  as:

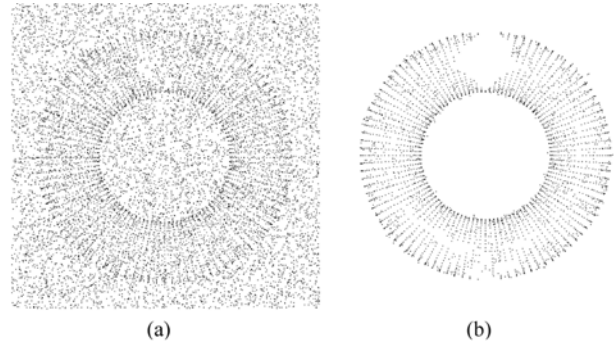
$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & r \end{bmatrix} \quad (15)$$

$r$  is half of the length of the minimum axis of the ellipsoidal kernel  $E$ . After the density value is estimated, we will remove all the points whose estimated density value is smaller than a user-defined threshold. Figure 1 and Figure 2 are two examples of the proposed nonparametric density estimation based outlier removal from synthetic

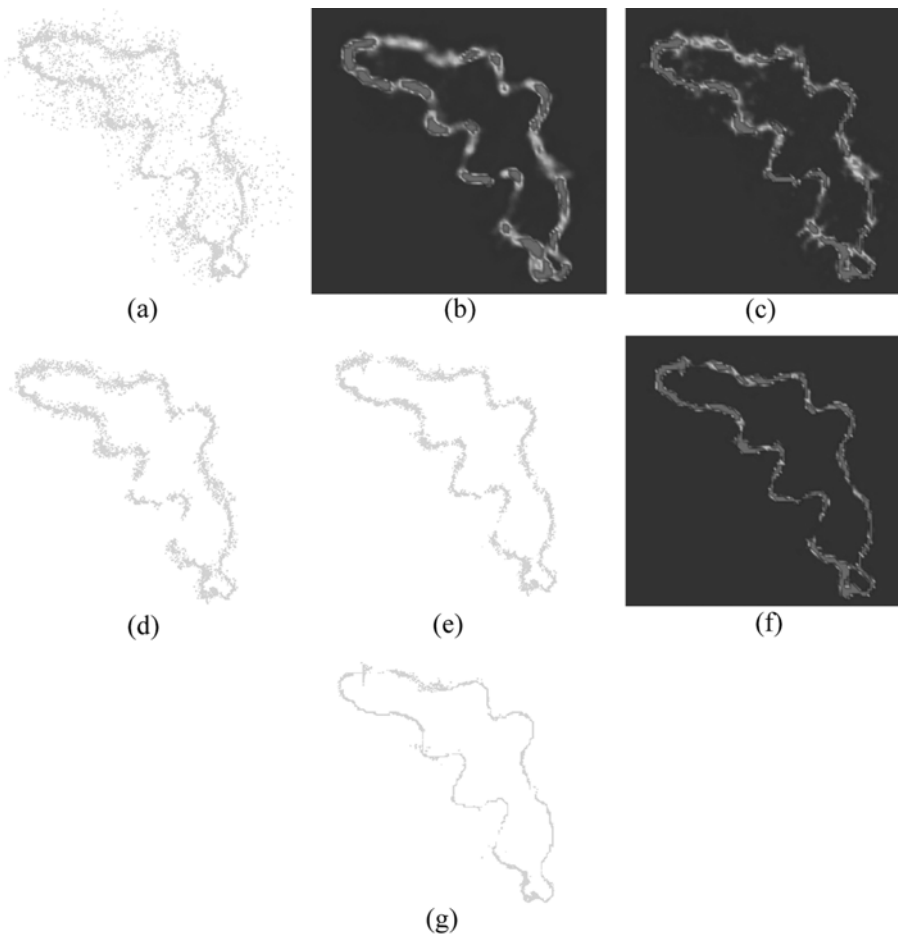
highly noisy data. The input of Figure 4 is a 3D point clouds obtained from multi-view stereo. Figure 4(a) is the original data; Figure 4(b) is the result after outlier removal. Figure 3 is a 2D slice view of the nonparametric data preprocessing process. In particular, Figure 3(a) is



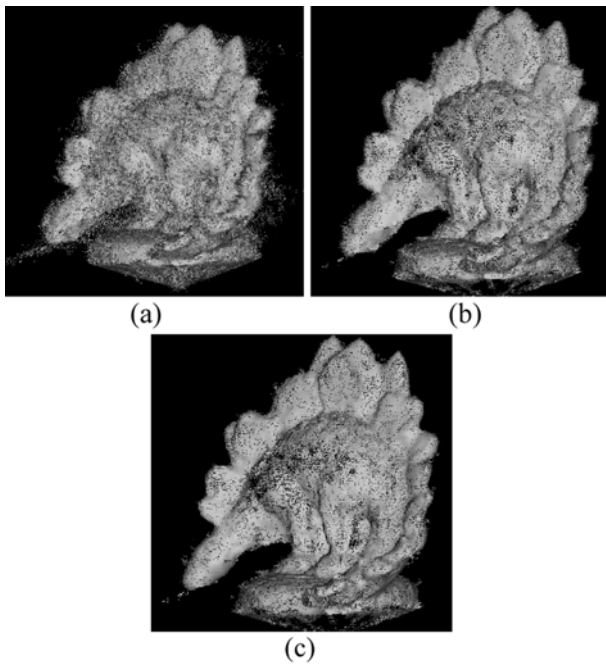
**Fig. 1.** (a) The Stanford bunny (35678 points) with 1000% noise points added in; (b) Result after outlier removal.



**Fig. 2.** (a) A synthetic torus point data (4800 points) which has an open hole in top, with 1000% noise points added in; (b) Result after outlier removal.



**Fig. 3.** A 2D slice view of the nonparametric data preprocessing process. (a) a 2D slice of the original 3D data of Figure 4(a); (b) color-coded density map of (a) estimated by an isotropic Gaussian kernel, with red represents the highest density and blue the lowest density; (c) color-coded density map of (a) estimated by an anisotropic Gaussian kernel of the same scale of (b); (d) after outlier removal using (b); (e) after outlier removal using (c); (f) color-coded density map of (e) estimated by an anisotropic Gaussian kernel of a smaller scale than (c); (g) result of projecting points of (e) by line search based on (f).



**Fig. 4.** (a) 3D point clouds (660,000 points) obtained from multi-view stereo; (b) after outlier removal; (c) after point projection by line search.

a 2D slice of the original 3D data of Figure 4(a); Figure 3(b) is the color-coded density map of Figure 3(a) estimated by an isotropic Gaussian kernel, with red represents the highest density and blue the lowest density; Figure 3(c) is the color-coded density map of Figure 3(a) estimated by an anisotropic Gaussian kernel of the same scale of Figure 3(b); Figure 3(d) is the result after outlier removal using Figure 3(b), and Figure 3(e) is the result after outlier removal using Figure 3(c). As we can see, better result is achieved by using the anisotropic kernel-based outlier removal (Figure 3(e)), comparing with the isotropic kernel-based outlier removal (Figure 3(d)).

### 2.3 Point projection by line search

After the outlier removal step, we will need to further process the remaining point data so that they will be projected into the real surface. We will define the real surface as the set of points whose density value (estimated by the above Parzen-window based kernel density estimation) is maximum along its surface normal direction. This is similar to the concept of “extreme surface” as suggested by Amenta et al. [Amenta 2004]. We implement the data projection by a hill-climbing line search method. More specifically, given a point  $x$ , it will move along a direction  $l$  that is (approximately) orthogonal to the real surface, and will stop when a local maximum of the density value is reached. Since the real surface is unknown, the moving direction  $l$  is approximated by the eigenvector  $u_3$  (Equation (21)) that is associated with the smallest eigenvalue  $\sigma_3^2$  (Equation (21)) of the covariance matrix  $H$  (Equation (11)), calculated by the local neighbourhood of point  $x$ . The orientation

of the moving direction is determined by the directional derivative of the density value along  $l$ , i.e. point  $x$  will always move uphill (i.e. move towards the direction that will increase the density value). Figure 4(c) shows the result of projecting data points by line search. A 2D slice view of the point projection is shown in Figure 3(g). Note that the outlier removal and the point projection procedure can be iterated one or two times to further improve the result.

## 3. Conclusions and Future Work

In this paper, we propose a nonparametric approach for noisy point data preprocessing. In particular, we proposed an anisotropic kernel based nonparametric density estimation for outlier removal, and a hill-climbing line search approach for projecting data points onto the real surface boundary. Our method is very robust with highly noisy dataset, as can be seen from the examples shown. In addition, our approach is computationally very efficient. For example, it only takes 3 to 4 minutes in a middle-range desktop PC to process (outlier removal and point projection) the multi-view stereo point data (Figure 4) which contains more than 600,000 points. The majority of the running time actually spends on computing the density. The line search based projection is done in less than 10 seconds for the whole dataset.

The main limitation of our approach is that currently a fixed bandwidth of the kernel is used across the whole dataset, which can cause some of the “good” data being removed during the outlier removal process (as can be seen in Figure 3(e)). We would like to develop a scheme that can automatically select the optimal kernel bandwidth based on the local neighborhood. We would also like to extend our approach so that it can handle hole fillings automatically.

## 4. Acknowledgments

We are very grateful to the Stanford Graphics Archive for the bunny dataset, and to Steve Seitz et al. [Seitz 2006] for the multi-view stereo image datasets.

Research was sponsored in part by the Leonard Wood Institute in cooperation with the U.S. Army Research Laboratory and was accomplished under Cooperative Agreement # LWI-281074, by the U. S. Army Research Office under contract/grant number #W911NF-07-1-0185, and by the NSF grant CMMI-0856206. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Leonard Wood Institute, the Army Research Laboratory, the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

## 5. Appendix: Decomposition of the Covariance Matrix $H$

Since the covariance matrix  $H$  is defined as:  $H = DD^T$  and  $D = (x_1 - x, x_2 - x, \dots, x_n - x)$ . Using Single Value Decomposition (SVD),  $D$  can be decomposed as:

$$D = U\Sigma V^T \quad (16)$$

$\Sigma$  is the diagonal matrix:

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} \quad (17)$$

with  $\sigma_1 \geq \sigma_2 \geq \sigma_3$ .  $U$  and  $V$  are orthonormal matrixes, i.e.  $U^T U = V^T V = I$  is the identity matrix. So,

$$H = DD^T = U\Sigma V^T V \Sigma U^T = U\Sigma^2 U^T \quad (18)$$

and let's define matrix  $A$  as:

$$A = \Sigma^2 \quad (19)$$

which is also a diagonal matrix. Thus we have

$$H = UAU^T \quad (20)$$

Furthermore, if we denote  $U = (u_1, u_2, u_3)$ , since  $U$  is an orthonormal matrix, we will have

$$Hu_i = \sigma_i^2 u_i, \quad (21)$$

which means  $u_i$  is the eigenvector of  $H$ , associated with its  $i^{\text{th}}$  largest eigenvalue  $\sigma_i^2$ .

## References

- [1] Alexa, M., Behr, J., Cohen-Or, D., Fleishman, S., Levin, D., Silva, C.T., Computing and rendering point set surfaces. *IEEE Transactions on Visualization and Computer Graphics* 9, 1, 3–15. 2003.
- [2] Amenta, N and Bern, M, Surface reconstruction by voronoi filtering, *Discrete and Comput. Geometry*, vol. 22, pp. 481-504, 1999.
- [3] Amenta, N., Bern, M., and Eppstein, D., The crust and the -skeleton: Combinational curve reconstruction, 14th ACM Symposium on Computational Geometry, 1998.
- [4] Amenta, N., Bern, M., and Kamvyselis, M., A new voronoi-based surface reconstruction algorithm, in *SIGGRAPH '98: Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, (New York, NY, USA), pp. 415-421, ACM Press, 1998.
- [5] Amenta, N., Kil, Y. J., Defining point-set surfaces. *ACM Trans. Graph.* 23(3): 264-270 (2004)
- [6] Bajaj, C. L., Bernardini, F., and Xu, G, Automatic reconstruction of surfaces and scalar fields from 3d scans, in *SIGGRAPH '95: Proceedings of the 22<sup>nd</sup> annual conference on Computer graphics and interactive techniques*, (New York, NY, USA), pp. 109-118, ACM Press, 1995.
- [7] Boissonnat, J. D., Geometric structures for three dimensional shape reconstruction, *ACM Trans. Graphics* 3, pp. 266-286, 1984.
- [8] Carr, J. C., Beatson, R. K., Cherie, J. B., Mitchell, T. J., Fright, W. R., McCallum, B. C., and Evans, T. R., Reconstruction and representation of 3d objects with radial basis functions, in *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, (New York, NY, USA), pp. 67-76, ACM Press, 2001.
- [9] Curless, B., and Levoy, M., A volumetric method for building complex models from range images, in *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, (New York, NY, USA), pp. 303-312, ACM Press, 1996.
- [10] Duda, R. O., Hart, P. E., Stork, D. G., *Pattern Classification*, October 2000, Wiley-Interscience; 2 edition.
- [11] Edelsbrunner, H., Shape reconstruction with delaunay complex, in *LATIN '98: Proceedings of the Third Latin American Symposium on Theoretical Informatics*, (London, UK), pp. 119-132, Springer-Verlag, 1998.
- [12] Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., and Stuetzle, W., Surface reconstruction from unorganized points, in *SIGGRAPH '92: Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, (New York, NY, USA), pp. 71-78, ACM Press, 1992.
- [13] Hilton, A., Stoddart, A. J., Illingworth, J., and Windeatt, T., Implicit surface-based geometric fusion, *Comput. Vis. Image Underst.*, vol. 69, no. 3, pp. 273-291, 1998.
- [14] Levin, D., Mesh-independent surface interpolation. In *Geometric Modeling for Scientific Visualization*, G. Brunnett, B. Hamann, K. Mueller, and L. Linsen, Eds. Springer-Verlag, 2003.
- [15] Medioni, G, Lee, M.-S., and Tang, C.-K., A computational framework for segmentation and grouping. Elsevier, 2000.
- [16] Ohtake, Y., Belyaev, A., Alexa, M., Turk, G., and Seidel, H., Multi-level partition of unity implicits, *ACM Trans. Graph.* 2003, Pages: 463-470.
- [17] Seitz, S., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R., A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms, *CVPR 2006*, vol. 1, pages 519-526.
- [18] Whitaker, R., A level set approach to 3D reconstruction from range data, *International journal of Computer Vision*, 1997.
- [19] Xie, H., Wang, J., Hua, J., Qin, H., Kaufman, A., Piecewise c1 continuous surface reconstruction of noisy point clouds via local implicit quadric regression. *IEEE Visualization 2003*, 91–98.
- [20] Zhao, H., Osher, S., Merriman, B., and Kang, M., Implicit and non-parametric shape reconstruction from unorganized data using a variational level set method, *Computer Vision and Image Understanding*, vol. 80, pp. 295-319, 2000.
- [21] Zhao, H., Osher, S., and Fedkiw, R., Fast surface reconstruction using the level set method, in *VLSM '01: Proceedings of the IEEE Workshop on Variational and Level Set Methods (VLSM'01)*, (Washington, DC, USA), p. 194, IEEE Computer Society, 2001.

---

**Ye Duan** is an Associate Professor of Computer Science at University of Missouri-Columbia. He received his B.A. degree in Mathematics from Peking University in 1991. He received his M.S. degree in Mathematics from Utah State University in 1996. He received his M.S. and Ph.D. degree in Computer Science from the State University of New York at Stony Brook in 1998 and 2003. From September 2003 to August 2009, he was an Assistant Professor of Computer Science at University of Missouri-Columbia. His research interests include Computer Graphics and Visualization, Biomedical Imaging and Computer Vision, Geometric and Physics-based Modeling, Virtual Reality and Human-Computer Interaction, Computer Animation and Simulation.

---

---

**Yongjian Xi** received B.S. and M.E. degrees in electrical engineering from Tsinghua University, China, in 1994 and 1997, received M.S. degree in computer science from Polytechnic University, New York, USA, in 2001. He is presently pursuing a Ph.D. degree in computer science in the University of Missouri-Columbia. His research interests includes 3D model reconstruction and volumetric data visualization.

---

---

**Hongkai Zhao** his B.Sc in Mathematics in 1990 from Peking University, China. He gets his Ph.D in Mathematics in 1996 from UCLA. He was Gabor Szego assistant professor in Mathematics at Stanford since 1996 before he moved to University of California Irvine in 1999, where he is a professor in Mathematics and Computer Science now.

---



Ye Duan



Yongjian Xi



Hongkai Zhao