A Neuro-Fuzzy Approach for 3D Object Detection and Extraction

Kousik Dasgupta, Siddhartha Bhattacharyya and Paramartha Dutta

Abstract—**A novel neuro-fuzzy approach for the detection and extraction of three-dimensional (3D) objects is presented in this article. The basic philosophy of the 3D object detection technique lies in the detection of contours of the objects under consideration. A fuzzy hostility index based thresholding on the volume neighborhood geometry contribution of the 3D objects is applied for the purpose of detection of the object contours.**

A multilayer self organizing neural network architecture characterized by a volume neighborhood geometry is used to extract 3D objects from a noisy 3D image scene.

Applications of the proposed 3D object detection and extraction technique are demonstrated with a 3D sphere image and a 3D cone image. The results are found to be encouraging.

Index Terms--**3D image processing, Contour Detection, Fuzzy Hostility, Object extraction, Multilayer Self Organizing Neural Network, Volume Graphics.**

I. Introduction

OBJECT extraction has been at the top of the research agenda of the computer vision community for more than a decade. The task of object extraction entails the segmentation of an image into foreground regions corresponding to different localized homogeneous object regions and background regions corresponding to non-object regions. This classification of the image pixels based on some local features as color, texture and position, is followed by merging the localized regions together. Typical applications include removal of noise artifacts, deblurring and segmentation/clustering of image data. The fields of defense, GIS, and space research rely heavily on these types of image processing application.

To date, most of the research efforts as well as commercial developments have been focused on two-dimensional (2D) approaches. This focus on monocular imaging has partly been motivated by costs but to a certain extent also by the need to retrieve objects from existing 2D image and video database. Extraction of three-dimensional (3D) volume images assumes greater complexity as 3D information about the subject is not always available. Moreover, the 3D sensing capability of the human perception system cannot be incorporated in the implementation phases. Another drawback is the unavailability of well known 3D image formats, as in 2D like bmp® or jpeg® etc. This also leads to unavailability of sufficient benchmark images and data to work on.

Several classical approaches exist in the literature to deal with the task of extraction of object centric features from an image [1], [2]. These approaches rely on identifying 4 or 8 connected components in an image scene merging and labeling tem appropriately as the object or background class. Another approach focuses on filtering out regions composed by less than a given number of pixels, assuming these pixels to be a part of the inherent noise content [3], [4]. Among other approaches, techniques involving standard morphological operations to retrieve object regions out of an image scene can also be found in [3], [5] and [6]. Neural networks have also been employed for dealing with object extraction, classification of relevant object specific information from redundant image information bases and recognition of objects from an image [7]-[9].

It may be noted that most of the aforesaid object extraction techniques should be independent of data representation, when applied to 3D images. Moreover, these methods also should be invariant under different object transformations like translation, rotation and scaling. This is the most important challenge for 3D object extraction as the 3D objects are usually saved in various poses and scales.

A score of approaches to deal with the 3D object extraction paradigm have been reported in the literature [10], [11]. Iso-surface techniques of 3D object extraction are effective for extracting bone, skin, respiratory organs, or the colon, which are all relatively easy to classify through intensity differences in the volume data. However, these techniques are not effective when applied to muscle or other soft tissues, for which contours are ill-defined in their neighborhoods. Content based 3D shape retrieval techniques classify objects in terms of image content. These methods are further classified as feature based methods such as Snakes [12], 3D surface

K. Dasgupta is with the Department of Computer Science and Engineering, Kalyani Government Engineering College, Kalyani 741235, India (e-mail: kousik.dasgupta@gmail.com).

S. Bhattacharyya is with the Department of Computer Science and Information Technology, University Institute of Technology, The University of Burdwan, Burdwan 713104, India (e-mail: siddhartha.bhattacharyya@gmail.com).

P. Dutta is with the Department of Computer and System Sciences, Visva-Bharati University, Bolpur, Santiniketan 731235, India (e-mail: paramartha.dutta@gmail.com).

deformable models [13], and a dynamic finite element surface model [14]. Graph based methods extract a geometric meaning from a 3D shape using model and Reeb graphs showing how shape components are linked together. Skeleton based methods [15] are sensitive to noise but they are time complex in nature and are restricted to specific 3D models. Sundar *et al.* [16] used a skeletal graph for encoding the geometric and topological information of 3D images. Neural networks have also been employed for dealing with 3D object extraction [17], [18].

II. Volume Graphics

Arie Kaufman, Daniel Cohen and Roni Yagel are the pioneers in the field of volume graphics [19]. Advancement in hardware technology leading to faster, larger and cheaper processor, memory and graphics card coupled with the development of volume visualization software, has helped in transforming revolutionary approaches in computer graphics to reality.

Volume graphics involves volume visualization - a method of extracting meaningful information from the available volumetric datasets by using interactive graphics and imaging techniques. It is mainly concerned with the representation, manipulation, and rendering of volumetric datasets [20]. Its objective is to provide mechanisms for peering inside volumetric datasets and for probing into voluminous and complex data structures and dynamics. It encompasses an array of techniques for projecting and shading a volumetric dataset or properties thereof and for retrieving volume information using transformations, cuts, segmentation, translucency control, measurements etc. Typically, the volumetric dataset is represented as a 3D discrete regular grid (i.e., a 3D raster) of volume elements (in short, voxels) and is commonly stored in a volume buffer (also called cubic frame buffer), which is a large 3D array of voxels. Alternatively, other data structures and formats have been employed for storing and manipulating the dataset, such as cell decomposition as in octrees [21], sparse voxel matrices, semi-boundaries, voxel runs, irregular grids, and surfaces of objects.

A voxel is the cubic unit of volume centered at any grid point. Representing a unit of volume, the voxel is the 3D counterpart of the 2D pixel which represents a unit of area, and thus the volume buffer of voxels can be regarded as the 3D counterpart of the 2D frame buffer of pixels. Each voxel has numeric values associated with it, which represent some measurable properties or independent variables (e.g., color, opacity, density, material, coverage proportion, refractive index, velocity, strength, time) of the real phenomenon or object residing in the unit volume represented by that voxel. The aggregate of voxels tessellating the volume buffer forms the volumetric dataset [19].

III. Review of current status of research

Most of the 2D object extraction techniques are affected strongly by variations in pose and illumination. A robust object/feature detection procedure of 2D images is still an open difficulty. But, the use of 3D information has gained much attention [22]-[24] especially in the field of face recognition, since 3D data is not affected by translation and rotation. Besides, 3D data is also immune to the effect of illumination variation. Since the 3D object extraction techniques should remain invariant under the different object transformations, such a task is a challenging thoroughfare. 3D object recognition research is, however, still weakly reported in the published literatures. A main reason baffling the development lies in that 3D image data capturing usually requires special expensive equipments.

But the recent advancements in the field of modeling, digitizing and visualizing techniques for 3D shapes has resulted in an increasing amount of 3D models, both on the Internet and in domain-specific databases. These have led to the development of different experimental benchmark search engines for 3D shapes, such as the 3D model search engine at Princeton University [25], the 3D model retrieval system at the National Taiwan University [26], the Ogden IV system at the National Institute of Multimedia Education, Japan [27], the 3D retrieval engine at Utrecht University [28]. But all these techniques suffer from the inherent limitation in that they do not deal with volume graphics in form of voxel but are mostly responsible for surface models.

For the purpose of visualization, 3D shapes are often represented as a surface model using particular polygonal meshes, for example in VRML format. These models often contain holes and intersecting polygons, and do not enclose any volume information. On the contrary, 3D volume models, such as solid models produced by CAD systems, or voxel models, enclose a volume properly. The paucity of reported literature is also partly due to the unavailability of standard 3D volume image or benchmark data. Some of the image formats which support voxel are ANALYZE 7.5®, DICOM®, binvox®, 3fd® etc [29]-[32].

In this article, a voxel based 3D object detection and extraction approach is presented. The binary voxel information retrieved from the binvox [32] 3D volume images, are used for the task of processing. Similar to the 2D counterpart, a fuzzy hostility index is computed to reflect the heterogeneity in the 3D volume images. This heterogeneity information enables the detection of the edges and surfaces of the 3D volume images. In addition, a multilayer self organizing neural network (MLSONN) architecture entailing an augmented neighborhood topology-based interconnectivity is used to extract 3D volume objects from noisy versions of the images.

IV. Mathematical Prerequisites

In this section, relevant fuzzy set theoretic concepts are discussed. A fuzzy hostility index reflective of the heterogeneity of the 3D volume images is also discussed in this section.

Fuzzy set theory was developed by L.A. Zadeh in 1965 [33] to explain the varied nature of ambiguity often encountered in real life situation. A fuzzy set [33], [34] is said to be a collection of elements, $A = \{x_1, x_2, x_3, ..., x_n\}$, characterized by a membership function $\mu_A(x_i)$, where x_i refers to the i^{th} element in the set. This membership function is an indication of the fuzziness in the set. It lies in [0, 1]. A membership value of 1 indicates strict containment of an element within the fuzzy set, while a membership value of 0 indicates weak containment of an element within the set.

The linear index of fuzziness of a fuzzy set A is the measure of the degree of fuzziness of the fuzzy set. It is given by the Hamming distance between the fuzzy set A and its nearest ordinary set $\underline{A}$. It is given by [35]

$$\nu_l = \min[\mu_A(x), 1-\mu_A(x)] \tag{1}$$

A. Fuzzy hostility index

Similar to 2D images which represent maps of pixels, 3D volume images represent maps of voxels. The ambiguity in the volume information in a 3D image can be regarded as a fuzzy set of voxel distribution therein. If a 3D volume image is represented as a collection of voxel neighborhood regions, where each voxel is surrounded by a number of neighboring voxels, such a fuzzy set of voxel presence/absence is equivalent to a superset of several neighborhood fuzzy subsets of voxel distributions. Each such candidate voxel in a particular neighborhood voxel distribution fuzzy subset is surrounded by several orders of neighboring entities, starting from a first order neighborhood (comprising seven immediate neighbors) or a second order neighborhood (comprising eleven neighbors) to some other higher order neighborhoods (comprising of seventeen and twenty six neighborhoods) as shown in Fig. 1.

The degree of ambiguity in these voxel distribution fuzzy subset neighborhoods is indicative of the degree of homogeneity/heterogeneity in that neighborhood. The closer are the representative membership values in the voxel distribution fuzzy subset neighborhoods, the higher is the homogeneity in that neighborhood and less is a candidate voxel hostile to its neighbors. On the contrary, a heterogeneous voxel distribution fuzzy subset neighborhood arises out of sharp contrasting membership values of the constituent voxels in the neighborhood and hence such a neighborhood is a hostile one.

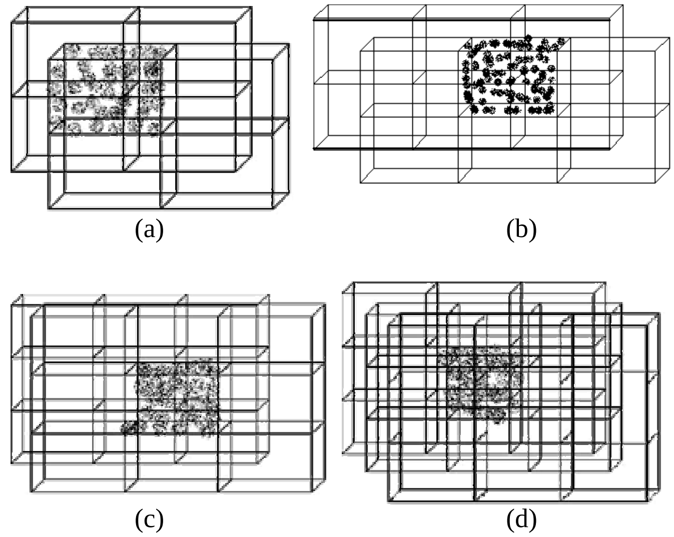

Fig. 1. Voxel neighborhood geometry (a) corner voxel with seven neighboring voxels (b) edge voxel with eleven neighboring voxels (c) surface voxel with seventeen neighboring voxels and (d) interior voxel with twenty six neighboring voxels

This homogeneity/heterogeneity in a voxel distribution fuzzy subset neighborhood can be accounted for by a fuzzy hostility index defined over the neighborhood as [36],

$$\varsigma = \frac{3}{N}\sum_{i=1}^{N}\frac{|\mu_p - \mu_{q_i}|}{|\mu_p + 1| + |\mu_{q_i} + 1|} \tag{2}$$

where, μ_p is the membership value of the candidate voxel and μ_{qi} , $i = 1, 2, 3, ..., N$ are the membership values of its neighbors in the neighborhood. The value of the fuzzy hostility index (ζ) lies in [0, 1].

V. 3D Object Extraction Using A Multilayer Self Organizing Neural Network (MLSONN) Architecture

The multilayer self organizing neural network (MLSONN) [35] introduced by Ghosh *et al.*, is efficient in extracting objects from a noisy 2D image. It is a feedforward neural network architecture with feedback of inputs. The network architecture comprising an input layer, any number of hidden layers and an output layer, operates in a self supervised manner. The number of neurons in each layer of the architecture corresponds to the number of pixels in the input noisy image. The neurons of the different network layers enjoy neighborhood topology-based interconnectivity among themselves. The output layer neurons are connected to the input layer neurons on a one-to-one basis for feedback of processed information. A schematic of the three layer version of the MLSONN architecture employing second order interconnectivity is shown in Fig. 2.

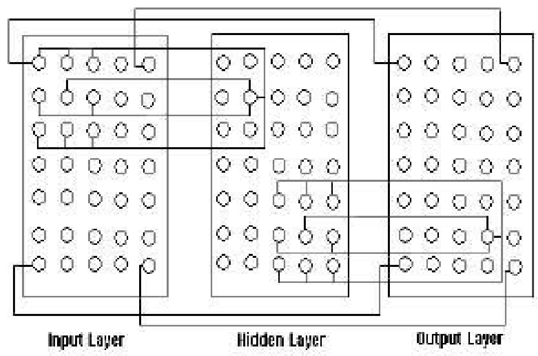

Fig. 2. Three layer MLSONN architecture employing second order neighborhood topology-based interconnectivity

The input layer neurons of the network architecture accept external world input image pixel intensity information in [0, 1]. The input at the j^{th} hidden layer neuron of the network architecture is given by

$$H_j = \sum_j I_i w_{ij} \tag{3}$$

where, I_i are the image information at the i^{th} input layer neuron and w_{ij} are the interconnection weights. The hidden layer processes this information as

$$O_j = f(H_j) \tag{4}$$

where, f is the standard sigmoidal activation function.

This information is further propagated to the output layer for further processing. As the network operates in self supervised nature, the system errors are determined from the linear indices of fuzziness in the outputs obtained. These systems errors are used to adjust the neighborhood topology-based interconnection weights between the different network layers by the standard backpropagation algorithm.

The output layer outputs are fed back to the input layer for further processing. This self supervised mode of processing is carried on until the interconnection weights stabilize or the system errors fall below some tolerable limits.

Thus, an MLSONN architecture with an augmented neighborhood topology-based interconnectivity corresponding to a volume neighborhood geometry can be used to extract 3D objects as well, when the neighborhood information of the 3D image are fed as inputs to the input layer of the MLSONN architecture. Each neuron of the different layers of the resultant MLSONN architecture would then correspond to the candidate voxels of the input 3D volume image information. The neurons of a particular layer of the augmented architecture would enjoy connectivity with twenty six neighboring neurons of the preceding layer neurons as shown in Fig. 1(d).

VI. Results of 3D Object Detection and Extraction

The entire process of 3D object detection and extraction has been carried out in four phases. Initially, the constituent voxel information are retrieved from 3D images. These information represent the presence/absence of voxels in the 3D input image. Subsequently, the neighborhood information of each of the constituent voxels are computed from the voxel neighborhood geometry. These voxel neighborhood information are the inputs to the 3D contour detection/3D object extraction procedure. For the purpose of detection of the contours of the input 3D images, a fuzzy hostility index based thresholding is applied on the voxel neighborhood information. The thresholded voxels represent the contour voxels of the detected images, reconstructed thereafter.

The computed neighborhood voxel information of noisy 3D images are fed to the input layer of an MLSONN architecture. The MLSONN architecture self organizes this input information into extracted voxels. The corresponding 3D objects extracted from the input noisy images are obtained by reconstructing the 3D images with the processed voxel information.

The proposed methodology has been applied for the detection of the contours as well as for the extraction of 3D objects from noisy/noise-free 3D images. The test 3D Cone and Sphere images are shown in Fig. 3. Figs. 4-6 show the detected surface and edges of the test 3D images using fuzzy hostility index based thresholding. A noisy 3D Cone image is shown in Fig. 7. Fig. 7 also show the extracted noise-free 3D Cone image obtained by the MLSONN architecture.

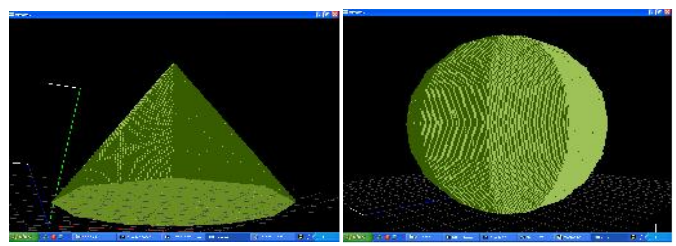

Fig. 3. Original 3D Cone and Sphere image

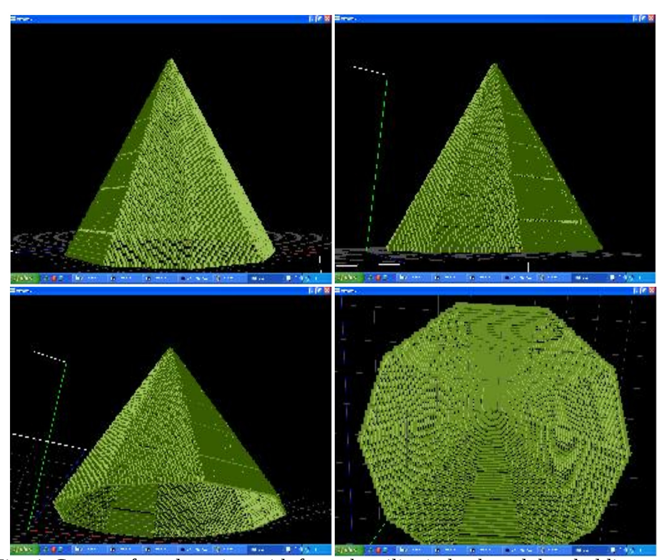

Fig. 4. Cone surface detection with fuzzy hostility index based thresholding

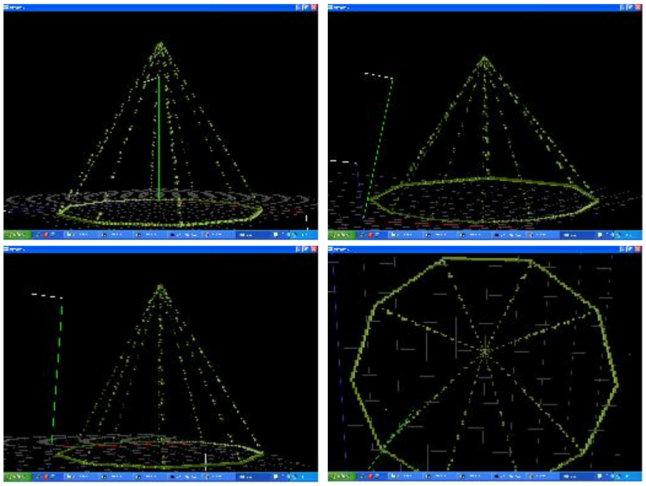

Fig. 5. Cone edge detection with fuzzy hostility index based thresholding

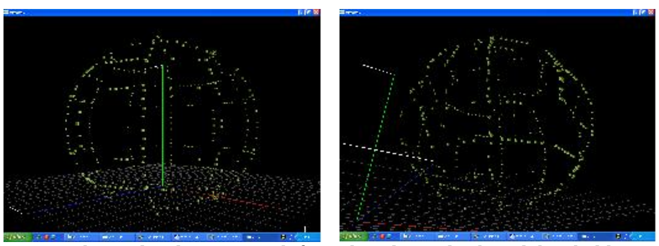

Fig. 6. Sphere edge detection with fuzzy hostility index based thresholding

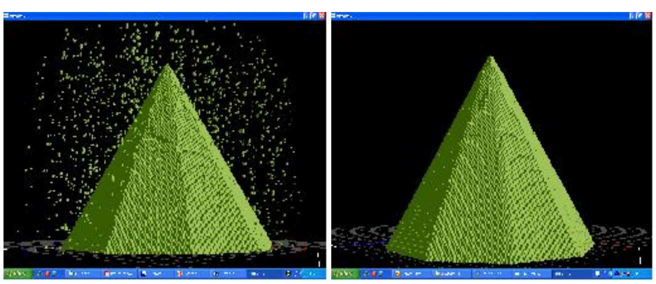

Fig. 7. Noisy 3D Cone image and MLSONN extracted counterpart

VII. References

[1] K. Suzuki, I. Horiba, and N. Sugie, "Fast connected-component labeling based on sequential local operations in the course of forward-raster followed by background–raster scan," in *Proc. 2000 International Conference on Pattern Recognition*, vol. II, pp. 434-437.

[2] F. Chang, C.-J. Chen, and C.-J. Lu, "A linear-time component-labeling algorithm using contour tracing technique," *Computer Vision and Image Understanding*, vol. 93 (2), pp. 206-220, 2004.

[3] R. C. Gonzalez and P. Wintz, *Digital Image Processing*, Addison-Wesley, MA, 1977.

[4] L.-H. Chen and J.-R. Chen, "Object segmentation for video coding," in *Proc. 2000 International Conference on Pattern Recognition*, vol. III, pp. 383-386.

[5] A. Rosenfield and A. C. Kak, *Digital Image Processing Techniques*, Academic, New York, 1982.

[6] F. Meyer and S. Beucher, "Morphological, Segmentation," *Journal of Visual Communication and Image Representation*, vol. 1 (1) pp. 21-46, 1990.

[7] G. L. Bilbro, M. White, and W. Synder, "Image segmentation with neurocomputers, " in R. Eckmiller and C. V. D. Malsburg (eds), *Neural Computers*, Springer, New York, 1988.

[8] M. A. Abdallah, T. I. Samu, and W. A. Grission, "Automatic target identification using neural networks, " *SPIE Proc. Intell Robots Comput Vis*, vol. XIV 2588, pp. 556-565, 1988.

[9] H. W. Tang, V. Srinivasan, and S. H. Ong, "Invariant object recognition using a neural template classifier," *Image Vis Comput*, vol. 14 (7), pp. 473-483, 1996.

[10] H. Tomas and S. Vaclav, "A survey of methods for 3D Model Feature Extraction," Project MSM 23500005, supported by the Ministry of Education of the Czech Republic.

[11] W. H. Tangelder John and C. Veltkamp Remco, "A Survey of Content Based 3D Shape Retrieval Methods" in *Proc. of 2004 Shape Modeling International Conference (SMI'04)*.

[12] M. Kass, A. Witkin, and D. Telzopoulos, "Snakes: Active Contour Models, " *International Journal of Computer Vision*, vol. 1 (4), pp. 321–331, 1988.

[13] D. Terzopoulos, A.Witkin, and M. Kass, "Constraints on Deformable Models: Recovering 3D Shape and Nonrigid Motion, " *Artificial Intelligence*, vol. 36, pp. 91–123, 1988.

[14] T. McInerney and D. Tezopoulos, "A Dynamic Finite Element Surface Model for Segmentation and Tracking in Multidimensional Medical Images with applications to Cardiac 4D Image Analysis," *Computerized Medical Imaging and Graphics*, vol. 19 (1), pp. 69–83, 1995.

[15] W.-C. Ma, F.-C. Wu, M. Ouhyoung, "Skeleton Extraction of 3D Objects with Radial Basis Functions" in *Proc. of 2003 Shape Modeling International Conference (SMI'03)*.

[16] H. Sundar, D. Silver, N. Gagvani, and S. Dickenson, "Skeleton based shape matching and retrieval," in *Proc. of 2003 Shape Modeling International Conference (SMI'03)*, pp. 130–139.

[17] W. C. Lin, F. Y. Liao, C. K. Tsao, "A hierarchical multi view approaches to three dimensional object recognition," in *IEEE Transactions on Neural Networks*, vol. 2, pp. 84-92, 1991.

[18] L. Spirkovska and M. B. Reid, "Higher-Order Neural Networks Applied to 2D and 3D Object Recognition" in *Proc. Machine Learning*, vol. 15, pp. 169-199, 1994.

[19] A. Kaufman, D. Cohen, and R. Yogel, "Volume Graphics, " *IEEE Computer Society Press*, pp. 51-64, July 1993.

[20] A. Kaufman, "Volume Visualization," *IEEE Computer Society Press*, 1990.

[21] D.J. Meagher, "Geometric Modeling Using Octree Encoding," *Computer Graphics and Image Processing*, vol.19 (2), pp. 129-147, 1982.

[22] X. Lu and A. K. Jain, "Deformation analysis for 3D face matching," *Proc. 2005 IEEE Workshop on Applications of Computer Vision*, pp. 99–104.

[23] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25 (9), pp. 1063–1074, 2003.

[24] C. Beumier and M. Acheroy, "Automatic 3D face authentication," Image and Vision Computing, vol. 18, no. 4, pp. 315–321, 2000.

[25] 3D model search engine, http://shape.cs.princeton.edu.

[26] 3D model retrieval system, http://3d.csie.ntu.edu.tw/~dynamic/.

[27] Ogden IV system, http://www.nime.ac.jp/~motofumi/Ogden/.

[28] 3D shape retrieval engine, http://www.cs.uu.nl/centers/give/imaging/3Drecog/3Dmatching.html.

[29] http://www.amiravis.com/usersguide41/hximio/HxFileFormat_Analyze75.html

[30] http://www.amiravis.com/usersguide41/hxfield/HxFileFormat_RawData.html

[31] http://www.amiravis.com/usersguide41/hxdicom/HxFileFormat_DICOM.html

[32] http://www.cs.princeton.edu/~min/binvox/binvox.html

[33] L. A. Zadeh, "Fuzzy Sets," *Inform and Control*, vol. 8, pp. 338–353, 1965.

[34] T. J. Ross, T. Ross, *Fuzzy Logic with Engineering Applications*, McGraw Hill College Div, 1995.

[35] A. Ghosh, N. R. Pal, and S. K. Pal, "Self-organization for object extraction using a multilaye rneural network and fuzziness measures," *IEEE Transactions on Fuzzy Systems*, vol. 1 (1), pp. 54-68, 1993.

[36] S. Bhattacharyya, P. Dutta, and U. Maulik, "Binary object extraction using bi-directional self-organizing neural network (BDSONN) architecture," *Pattern Analysis and Applications*, 2007.

VIII. BIOGRAPHIES

Kousik Dasgupta did his Bachelors in Engineering in Electronics and Power Engineering from Nagpur University, Nagpur, India in 1993. Subsequently, he did his Masters in Computer Science & Engineering in 2007 from West Bengal University of Technology, Kolkata, India.

He is currently a Lecturer in the Department of Computer Science and Engineering of Kalyani Government Engineering College, Kalyani, India. He served industries like ABB and L & T during 1993-1996 and as a Technical Assistant in Kalyani Government Engineering College, Kalyani, India before joining the same institute as lecturer in 2001. He is co-author of two books and about 10 research publications. His research interests include soft computing, computer vision and image processing.

Dasgupta is a Life Member of ISTE, India, Associate Member of The Institute of Engineers, India and Chartered Engineer [India] of The Institute of Engineers, India. He is a Fellow of OSI, India,

Siddhartha Bhattacharyya did his Bachelors in Physics and Optics & Optoelectronics from Calcutta University, Kolkata, India in 1995 and 1998, respectively. Subsequently, he did his Masters in Optics & Optoelectronics in 2000 from Calcutta University, Kolkata, India.

He is currently a Lecturer in the Department of Computer Science and Information Technology of University Institute of Technology, The University of Burdwan, Burdwan, India. He served as a Lecturer in the Department of Information technology of Kalyani Government Engineering College, Kalyani, India during 2001-2005. He has served as a Project Fellow in Bengal Engineering and Science University, Shibpore, India. He is a coauthor of one book and about 35 research publications. His research interests include soft computing, pattern recognition and image processing.

Bhattacharyya is a Life Member of ISTE, India and Member of IAENG, Hong Kong. He is a Fellow of OSI, India.

Paramartha Dutta did his Bachelors and Masters in Statistics from Indian Statistical Institute, Kolkata, India in 1988 and 1990, respectively. Subsequently, he did his Masters in Computer Science in 1993 from Indian Statistical Institute, Kolkata, India. He did his Ph. D. in 2005 from Bengal Engineering and Science University, Shibpore, India.

He is currently a Professor in the Department of Computer and System Sciences of Visva-Bharati University, Bolpur, India. He served as a Professor and Head of the Department of Computer Science and Engineering of Kalyani Government Engineering College, Kalyani, India in 2005. Prior to that, he was an Assistant Professor in the same Department during 2001-2005 and in the Department of Computer Science and Engineering of College of Engineering and Management, Kokaghat, India during 1998-2001. He has served as a Research Scholar in Indian Statistical Institute, Kolkata, India and in Bengal Engineering and Science University, Shibpore, India. He is a coauthor of four books and about 50 research publications. His research interests include evolutionary computing, soft computing, pattern recognition, multi-objective optimization and mobile computing.

Dutta is a Life Member of ISTE, India. He is a Member of ISCA, CSI, IETE, India and IAENG, Hong Kong. He is a Fellow of OSI, India.