I²MTC 2008 – IEEE International Instrumentation and
Measurement Technology Conference
Victoria, Vancouver Island, Canada, May 12–15, 2008

Selective Vision Sensing with Neural Gas Networks

Ana-Maria Cretu, Pierre Payeur, Emil M. Petriu
Sensing and Modeling Research Laboratory
School of Information Technology and Engineering
University of Ottawa
Ottawa, ON, CANADA, K1N 6N5
[acretu, ppayeur, petriu]@site.uottawa.ca

Abstract *– Vision sensing systems are experiencing an unprecedented growth in numerous applications. The collection of such a rich flow of information has brought a new challenge in the selection of only relevant features out of the avalanche of data generated by the sensors. This paper presents some aspects of our research work on intelligent sensing for advanced robotic applications. The main objective of the research is to design innovative approaches for automatic selection of regions of observation for fixed and mobile sensors to collect only relevant measurements without human guidance. A solution using neural gas networks has been investigated to adaptively select regions of interest that require further sampling from a cloud of 3D measurements sparsely collected. The technique automatically determines bounded areas where sensing is required at high resolution to accurately map 3D surfaces. It provides significant benefits over brute force strategies as scanning time is reduced and datasets size is kept manageable. Experimental evaluation of this technology is presented for 3D surface sampling/sensing.*

Keywords *– Selective sensing, 3D vision, neural networks, neural gas, feature detection, surface modeling.*

I. INTRODUCTION

Due to the high measurement speed of 3D data acquisition devices (e.g. laser scanners) on one side and to the lack of knowledge on appropriate accuracy levels for correct description of shape and geometry on the other side, the data acquisition process is elaborate and often produces too many sample points. Reducing the complexity of such datasets is one of the key techniques required in order to operate subsequent applications on the resulting data at a reasonable computational cost. To tackle this issue, the most widely exploited trend in literature implies the post-processing of large datasets obtained by acquisition devices. Frequently, the proposed algorithms rely on the user input for providing parameters such as the desired density of sampling, the regularity of sampling, or the minimum distance between samples. This is a difficult task as the user is not always aware of the appropriate level of accuracy required for a model in order to be further processed and the adjustment of such parameters can be a lengthy trial-and-error process. Further research on adaptive sampling would benefit from automated selective procedures to determine regions of interest and collect only relevant measurements for modeling applications.

The main objective of this research is the design of innovative approaches to achieve automatic selection of regions of observation for vision sensors to collect only relevant measurements without human guidance. The relevant regions of interest are extracted from 3D point clouds during the acquisition procedure to prevent an avalanche of data and the related excessive processing load. Starting from an initial, fast, and sparse scan of an object, a neural gas network map is used to adaptively select areas of interest for further scanning in order to improve the accuracy of the model. The final model is a multi-resolution model with a higher resolution in areas rich in features.

The paper is structured as follows: we start by showing the state-of-the-art in post-processing and sampling of large datasets in form of point clouds in Section II. We then detail our proposed solution for selective sensing in Section III and show experimental results for data sampling using vision sensors in Section IV. Finally, we present future research directions and draw the conclusions.

II. LITERATURE REVIEW

In general, there are three sampling policies proposed in the literature: uniform sampling, random sampling and stratified sampling. In uniform (regular or grid) sampling, samples are spread such that the probability of a surface point to be sampled is equal for all surface points. The method is popular because it can be easily implemented and ensures complete coverage of the surface within the sensor's field of view. However the cost is high, since in order to achieve adequate sampling density over those regions requiring the highest resolution, the sampling density must be uniformly high everywhere. In random sampling, each point of the object has an equal chance of being selected, but only a lower number of points are collected. As the percentage of sampled points increases, the cost gets higher to eventually reach the one of uniform sampling. The risk here is that samples randomly collected can miss important features.

Stratified sampling is a technique that generates evenly spaced samples by subdividing the sampling domain into non-overlapping partitions (clusters) and by sampling independently from each partition. The method ensures that an adequate sampling is applied to all partitions. This idea is often exploited in the context of post-processing of large point clouds or meshes [1, 2, 3, 4, 5, 6, 7], where a subdivision of models into grid cells occurs and sample points that fall into the same cell are replaced by a common representative.

The 3D model proposed by Nehab *et al.* [1] is first voxelized with an octree and one sample is outputed for each

voxel. The common representative for a voxel is selected according to a probability that decays as the distance of the sample to the center of the voxel increases. The user controls the sampling resolution, the regularity of the sampling, and the minimum distance between samples. In [2] the representative point is the measured point that is closest to the average of points that fall into the same voxel.

The 3D grids proposed by Lee *et al.* [3] are octrees that voxelize 3D points constructed by registration and integration of multiple scanned datasets of an object or scene of objects. The method uses point normal values on the surface of the object (computed based on the knowledge that scanned lines and points are ordered due to raster scanning and based on a triangulation performed on two neighbouring scan lines) from which non-uniform grids are generated using the standard deviation of normal values. The representative point for each grid is the point whose normal is closest to the average of the points in the same voxel.

The standard volumetric subdivision strategy cannot adapt to non-uniformities in the sampling distribution and sometimes joins unconnected parts of a surface if the grid cells are too large. To alleviate these problems, Pauly *et al.* [4] perform surface-based clustering, where clusters are built by collecting neighbouring samples while taking into account local sampling density. Points are incrementally added to a cluster until a maximum size and/or a maximum allowed variation is reached. The sample points of those clusters that do not reach the minimum size or variation bound are distributed to the neighbouring clusters. Alternatively, clusters can be computed by recursively splitting the point cloud using binary space partitions.

Uesu *et al.* [5] simplify large unstructured meshes by segmenting them into two parts: the boundary of the original domain and the interior samples and then simplifying each part separately, considering proper error bounds. For the boundaries, a modified surface-simplification algorithm that takes into account the scalar field defined at the vertices is employed, while the interior points are sampled using a *k-d* tree partition of the mesh from which the samples that are outside the boundary, or closer than a certain minimum distance to the boundary are removed. Finally, the simplified domain boundary and scalar field are combined into a complete, simplified mesh using a Delaunay tetrahedralization.

A similar idea is employed by Song and Feng [6] whose point cloud simplification algorithm starts by identifying and retaining edge boundaries (using a measure of the deviation of the normal vectors in the Voronoi neighbourhood) and then removing less important points from the remaining data based on their contribution to the local surface geometry. If the local geometry cannot be reliably reflected by the neighbouring points of a point and their associated properties, that point is considered important for defining geometry, otherwise it can be removed. The removal procedure ends once the specified data reduction ratio is reached.

Kalaiah and Varshney [7] propose a scheme to compactly decimate and represent point clouds using Principal Component Analysis (PCA). The input is pre-processed using an octree and PCA analysis is performed for each cell. Due to the fact that PCA parameters (orientation, frame, mean, variance) tend to be similar for coherent regions, the node parameters can be classified using clustering and quantization. At run-time, based on the viewpoint, a cut in the octree is determined and each node of the cut is visualized using a Gaussian random generator. Attributes like normals and color are also generated in the same manner.

Based on the fact that sampling algorithms are usually implying a measure of error, the idea of using error propagation neural networks that minimize an error measure seems a good choice to Fiori *et al.* [8], who propose a multilayer-feedforward network for non-uniform image sampling for robot motion control. The neural algorithm allows a controller to determine the robot's location within a structured environment based on a digital image sequence coming from a camera. The proposed sampling starts with a number of points (pixels in the image) uniformly distributed over the whole image that all become input points in a neural network. Network pruning is then performed on the basis of inputs relevance and therefore the number of inputs is reduced while the mean-squared error is kept below a pre-defined threshold. After pruning, the remaining sampling points are moved toward the high relevance areas based on a radial-basis-function sampling operator.

All these methods are not meant to be incorporated in the actual sampling procedure, but they rather post-process collected data. An approach to integrate the sampling procedure into the measurement procedure is proposed by Pai *et al.* [9, 10] in the context of deformable object modeling. The authors use a probing procedure that considers a known mesh of the object under study, as well as a set of parameters such as the maximum force exerted on the object, the maximum probing depth and the number of steps for the deformation measurement. During probing, an algorithm generates the next position and orientation for the probe based on the specifications and the mesh of the object under test. It performs at the same time proximity checks and verifies the expected contact location of the probe with the mesh based on line intersection. However, the procedure is not selective and therefore is similar to collecting data for all the points of the mesh and can take very long.

III. PROPOSED FRAMEWORK

Meant to be incorporated directly in the sampling procedure, the proposed framework to achieve automated selective scanning over large workspaces uses a self-organizing neural network architecture that adaptively selects regions of interest for further refinement from a cloud of 3D sparsely collected measurements. The framework is depicted in Fig. 1.

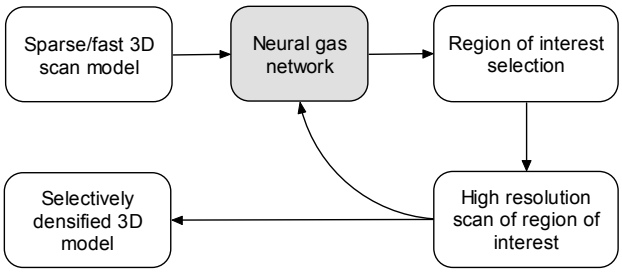

Fig. 1. Proposed framework for selective sampling.

Starting from an initial low resolution scan of an object, a neural gas network is employed to model the resulting point cloud. Those regions that are worth further sampling in order to ensure an accurate model are detected by finding higher density areas in the neural gas map. This is done by applying a Delaunay tessellation to the resulting neural gas output map and by subsequently removing all the triangles in the tessellation that are larger than a set threshold. The threshold is automatically computed based on the length of vertices for each triangle in the tessellation. Rescanning at higher resolution is performed for each identified region and a multi-resolution model is then built using the initial sparse model and augmenting it with the high resolution regions of interest.

A. Neural Gas Network

The use of a self-organizing architecture is justified by its ability to quantize the given input space into clusters of points with similar properties. As it was presented in the literature review, clustering is an efficient way to compress data. In our previous research we proved that a neural gas network is able to cluster both geometric and elastic properties of the objects embedded in a modeled point cloud [11, 12]. The neural gas network is selected instead of other self-organizing architectures due to its capability to capture fine details, unlike other architectures that tend to smooth them, as the Kohonen self-organizing map for example [11].

Starting from the points collected during a fast scan of an object via an active range finder and an initial configuration of unconnected nodes, the latter move freely over the data space while learning and the model contracts asymptotically towards the points in the input space, respecting their density and thus taking the shape of the objects encoded in the point cloud. This ensures that density of the probing points will be higher in the regions with more pronounced variations in the geometric shape.

The training is stopped early in order to avoid that the nodes become uniformly distributed instead of capturing details. As the point clouds that we obtain using our sensors are raster-like models and their density is uniform, the neural gas, that respects the density in the point cloud, after a long training, will tend to build uniformly dense maps as opposed to keeping nodes in the regions rich in features. Therefore the need for a shortened training period.

B. Selective Sampling

Since we know that the higher density regions in the neural gas are the ones of interest, a simple technique is employed to detect them. First, a Delaunay triangulation is applied over the output map in order to connect the nodes of the neural gas map. The triangulation is traversed and the length of vertices is estimated between every pair of points for every triangle. The average value of all these lengths is computed and a threshold is set equal to this value. All the triangles that contain vertices larger than the threshold are then removed from the model. The points of the remaining triangles identify those regions that require additional sampling. Data is collected over these regions and the resulting selectively sampled multi-resolution model is constructed by augmenting the initial sparse low resolution scan with the higher resolution data samples. The procedure can be repeated to improve accuracy in successive steps.

IV. EXPERIMENTAL RESULTS

The proposed method is tested in the context of range imaging where selective sampling of points is performed with range vision sensors in order to accurately describe a 3D object. Two objects are used for experimentation: a foam armchair and a mockup car door, each presenting different sorts of regions of interest, as depicted in Fig. 2a and Fig. 2c. The armchair should have as regions of interest the edges, while for the door model, regions of interest should be depicted around the door knob and the door opening gap. The model of the armchair depicted in Fig. 2a is obtained using an automated Jupiter laser scanner mounted on a robot arm [13], as shown in Fig. 2b. The full resolution point cloud shown in Fig. 2c is collected with the Neptec Design Group Inc.'s Laser Metrology System (LMS) range sensor, depicted in Fig. 2d.

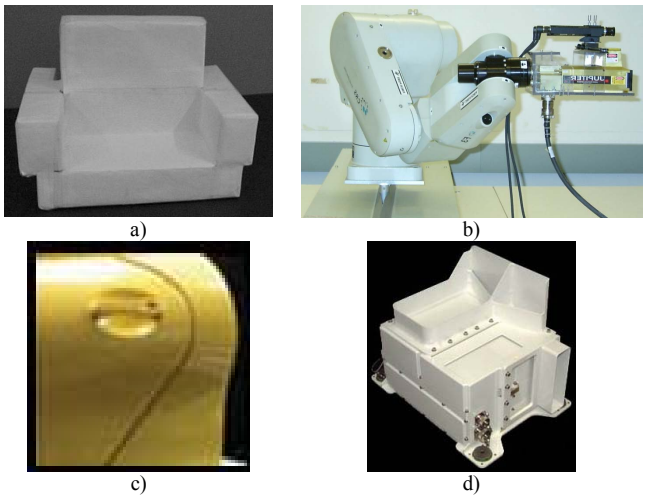

Fig. 2. a) Foam armchair, b) Jupiter laser scanner used to collect the data, and c) rendered point cloud of car door collected using d) Neptec's full image LMS range sensor.

Starting from an initial fast, sparse scan of each object, a neural gas network is employed to model the data in the point cloud. The data is normalized prior to the neural gas mapping such that its variance is unity in order to improve the learning procedure.

The results for the model of the armchair are presented in Fig. 3. The 3851 points of the initial low resolution scan and 16384 points of the initial medium resolution scan are presented in Fig. 3a and 3b respectively. The data is used as provided by the Jupiter laser scanner, without any filtering procedure, therefore some noise is present in the scans. These point clouds are provided each to a neural gas network with the map size of 45×40 and 35×45 respectively (implying a size of about 40% and about 10% respectively from the point cloud of the initial scan). The training using Matlab code, on a Pentium IV 1.3GHz machine with 512MB of memory takes about 3.5 min and 2 min respectively for 40 training epochs. The relative error is 0.04 and 0.011 respectively. The obtained output maps are compressed models for the dataset in which the weight vectors consist of the 3D coordinates of the object's points and are shown in Fig. 3c and 3d respectively. The artifacts in neural gas map are due to the fact that the training is stopped early, as explained in Section IIIA.

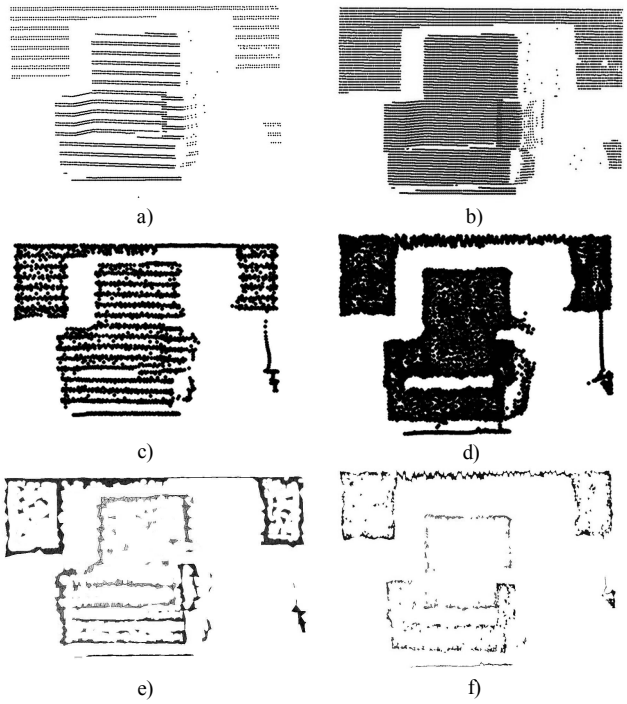

Fig. 3. Initial scan of an armchair at (a) low resolution and (b) medium resolution; neural gas model for a map size of (c) 45×40 for low resolution model and (d) 35×45 for medium resolution; and detected regions of interest for further sampling from (e) low resolution and (f) medium resolution models.

Since there are no other significant details in the model apart from the edges and the back plane, these are selected as regions of interest for additional sampling in order to improve the accuracy of the model. Even in the case of a low resolution initial dataset and in spite of some noise and artifacts visible in the modeling result (Fig. 3e), the proposed algorithm succeeds to identify the edges. When the initial dataset is denser, those features are more accurately located (Fig. 3f). But it is interesting to observe that the training time of the neural gas does not grow proportionally with the increase in the dataset size.

The more complex case of the car door which presents finer details is shown in Fig. 4 and Fig. 5. A fast scan is initially performed to obtain a very low resolution point cloud of 4096 points shown in Fig. 4a (0.39% of the high-resolution full scan of 1048576 points) and a medium low resolution of 16384 points depicted in Fig. 5a (1.56% of the high-resolution full scan). The data provided by the Neptec's full image LMS range sensor is less noisy, due to its higher resolution when compared to the Jupiter laser scanner. Fig. 4b and 5b show the rendered mesh models embedded in each point cloud. The region representing the door opening gap is very small in comparison with the whole model and so is the number of points representing it. This requirement poses additional challenges to the modeling. In case of the very low resolution model, a denser neural gas map is required in order to capture the details, again stopped early in the training as in the case of the armchair. The normalized point cloud is provided as input to a neural gas network with a map size of 45×40 (1800 points which represents a map size of 43% of the number of points in the initial very low resolution sparse scan) and the resulting model is depicted in Fig. 4c. The training procedure takes about 5 min for 30 epochs and the relative error is 0.011. A map with lower density provides good results both in terms of accuracy and length of training for the medium low resolution model. The training for a map size of 35×45 (roughly 10% of the number of points in the initial medium low resolution sparse scan), for 30 epochs takes about 4 min. The relative error reached is 0.009 and the resulting model is shown in Fig. 5c.

Those regions that are worth further sampling in order to ensure an accurate model are detected by finding high density areas in the neural gas model as described in Section IIIB. The detected regions of high density in the neural gas output map are identified in Fig. 4d and 5d for the two low resolution point clouds, by building the Delaunay triangulation and removing large triangles from it. The regions of interest superimposed on the rendered model are depicted in Fig. 4e and 5e respectively. The resulting augmented multi-resolution models for the sample of regions of interest are presented in Fig. 4f and 5f respectively. The selectively densified model in Fig. 4f contains 111596 points and it represents a reduction of about 90% in the number of points when compared to the full scan, while the one in Fig. 5f contains 173884 points (about 83% reduction of the full resolution scan).

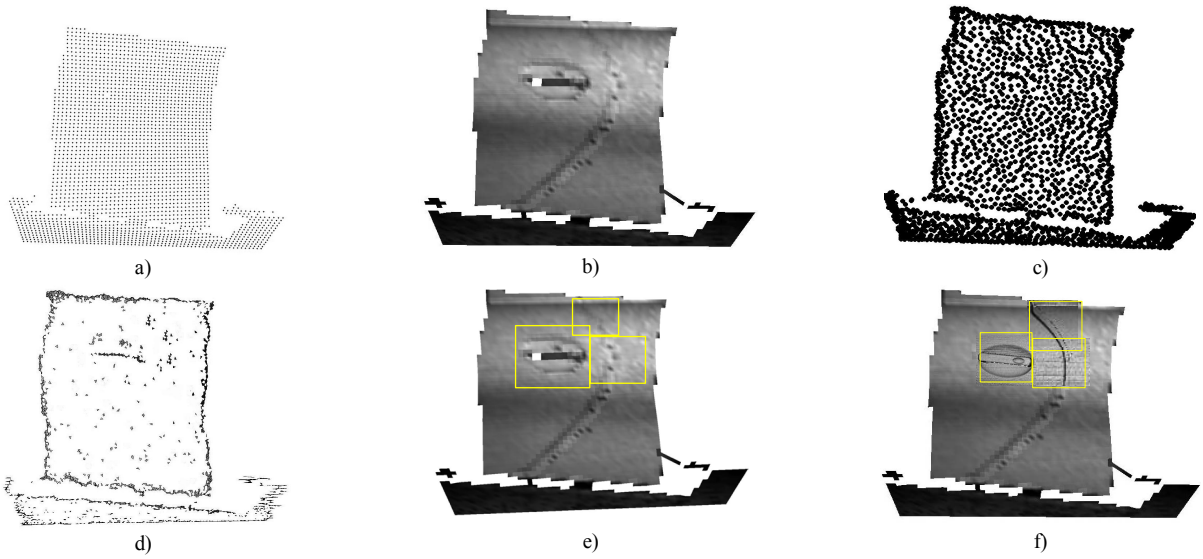

Fig. 4. Modeling steps from a very low resolution mockup car door point cloud: a) initial point cloud, b) rendered mesh model, c) neural gas model, d) higher density areas in neural gas model, e) identified regions of interest, and f) selectively densified models.

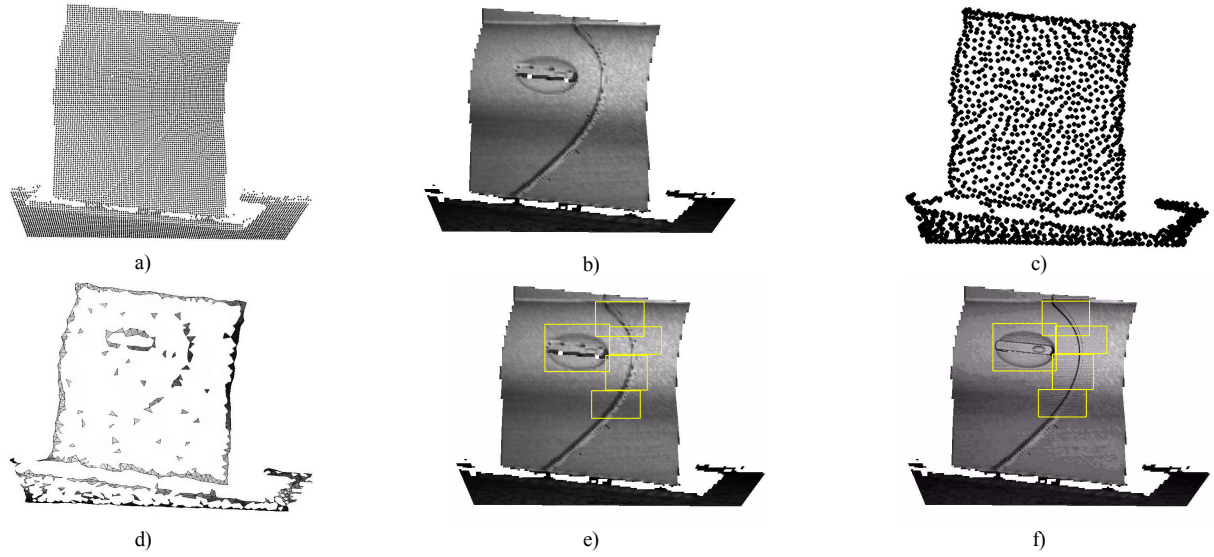

Fig. 5. Modeling steps from a medium low resolution mockup car door point cloud: a) initial point cloud, b) rendered mesh model, c) neural gas model, d) higher density areas in neural gas model, e) identified regions of interest, and f) selectively densified models.

Starting from the medium low resolution scan, the model is able to determine finer regions of interest and therefore saves from larger amounts of less relevant data in the scan. We would expect that even higher resolution initial scans would give better results in terms of regions of interest identification. However, this is not necessarily true. Additional points in the model make the door gap almost invisible, since the relative number of points representing it is significantly reduced for point clouds with higher number of points than 40% of the highest resolution scan available.

The same procedure can be repeated recursively for each of the regions of interest detected in the previous step. Each region can be provided as input to a neural gas network in order to further detect fine details that are worth to be scanned at a higher resolution. Fig. 6 presents the details of the high resolution rendered model of the door, with the selected regions from the very low resolution scan in the previous step. For each of the regions it also shows the neural gas model for a map size of about 20% of the total number of the points in the region and the selected sub-regions for additional sampling, identified as high density areas in the neural gas model. The training time is about 6 min and the relative error is below 0.01 for each of the regions.

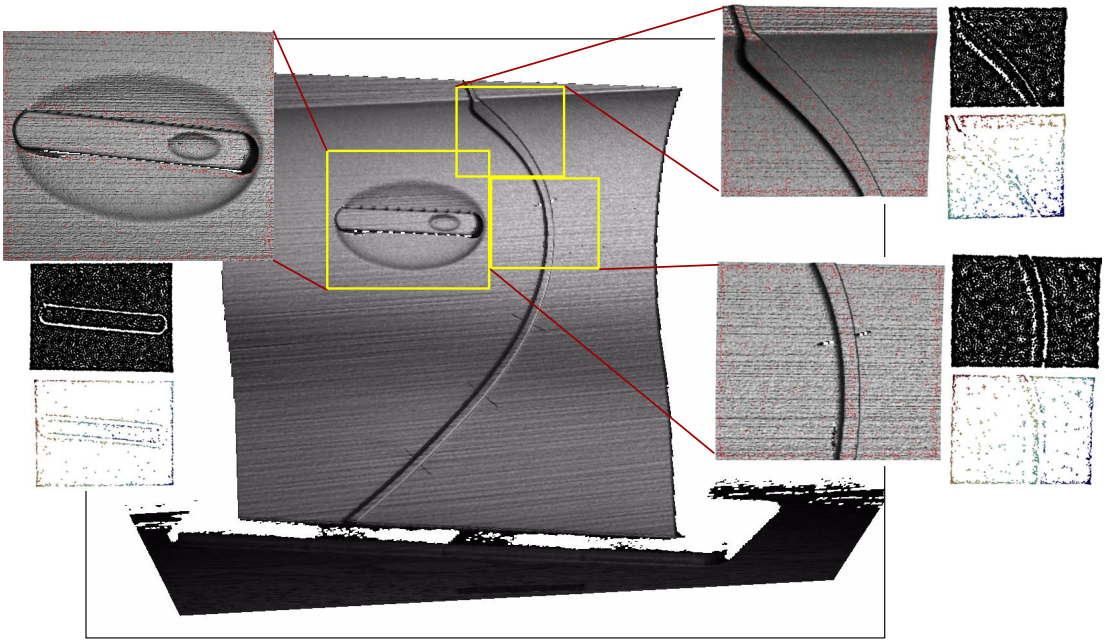

Fig. 6. Model of a car door showing a sample of selected regions, their enlargement, their local neural gas model and the selected sub-areas for sampling.

V. CONCLUSIONS

Both examples show the ability of the neural gas map to capture the fine details in the sparsely collected point clouds of the objects under study. By finding the high density areas in the neural gas map, the proposed selective sampling procedure is able to identify and guide the vision sensors to collect only measurements in those regions that are of interest for the improvement of accuracy of the obtained models, saving large amount of less relevant data in the scans.

In a wider context, the same procedure can be extended to sample elastic behavior by tactile probing since changes in geometry are usually correlated with changes in the elastic behavior. The fact that similar architectures can be used for adaptive tactile sensing and selective vision sensing opens the door to the development of multi-resolution composite geometric and elastic models based on the proposed approach.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the support from the Communications and Information Technology Ontario Centre of Excellence, and from the Natural Sciences and Engineering Research Council of Canada. The authors also want to thank Dr. Chad English from Neptec Design Group Inc. for his collaboration and for providing access to high resolution range data.

REFERENCES

[1] D. Nehab and P. Shilane, "Stratified Point Sampling of 3D Models", *Eurographics Symposium on Point-Based Graphics*, M. Alexa, S. Rusinkiewicz (Eds.), pp. 49-56, 2004.

[2] R. Dillmann, S. Vogt, and A. Zilker, "Data Reduction for Optical 3D Inspection in Automotive Applications", *Proc. of IEEE Int. Conf. Multisensor Fusion and Integration for Intelligent Systems,* pp. 159-164, 1999.

[3] K.H. Lee, H. Woo, and T. Suk, "Point Data Reduction Using 3D Grids", *Int. Journal of Adv. Manuf. Tech.,* vol. 18, pp. 201-210, Springer, 2001.

[4] M. Pauly, M. Gross, and L.P. Kobbelt, "Efficient Simplification of Point-Sampled Surfaces", *Proc. of IEEE Conf. Visualization*, pp. 163-170, 2002.

[5] D. Uesu, L. Bavoil, S. Fleishman, J. Shepherd, and C. T. Silva, "Simplification of Unstructured Tetrahedral Meshes by Point Sampling", *Proc. of IEEE Int. Workshop Volume Graphics*, E. Gröller, I. Fujishiro (Eds.), pp. 157- 238, 2005.

[6] H. Song and H.-Y. Feng, "A Point Cloud Simplification Algorithm for Mechanical Part Inspection", *Information Technology for Balanced Manufacting Systems*, W. Shen (Ed.), Springer, pp. 461-468, 2006.

[7] A. Kalaiah and A. Varshney, "Statistical Point Geometry", *Proc. of Eurographics Symp. Geometry Processing*, K. Kobbelt, P. Schroder and H. Hoppe (Eds.), pp. 107 - 115, 2003.

[8] S. Fiori, A. Faustini, and P. Burrascano, "Non-Uniform Sampling for Robot Motion Control by the GFS Neural Algorithm", *Proc. of IEEE Int. Conf. Neural Networks*, pp. 2057-2060, 1999.

[9] D. K. Pai, K. van den Doel, D. L. James, J. Lang, J. E. Lloyd, J. L. Richmond, and S. H. Yau, "Scanning physical interaction behavior of 3D objects", *Proc. of Computer Graphics and Interactive Techniques,* pp. 87-96, 2001.

[10] J. Lang, D.K. Pai, and R. J. Woodham, "Acquisition of Elastic Models for Interactive Simulation", *Int. Journal of Robotics Research*, vol. 21, no.8, pp.713-733, 2002.

[11] A.-M. Cretu and E.M. Petriu, "Neural Network-Based Adaptive Sampling of 3D Object Surface Elastic Properties", *IEEE Trans. Instr. and Meas.*, vol. 55, no. 2, pp. 483-492, 2006.

[12] A.-M. Cretu, E.M. Petriu, and P. Payeur, "Neural Network Mapping and Clustering of Elastic Behavior from Tactile and Range Imaging for Virtualized Reality Applications", *Proc. of Int. Workshop Imaging Systems and Techniques*, pp. 17-22, Minori, Italy, Apr. 2006.

[13] P. Curtis, C.S. Yang, and P. Payeur, "An Integrated Robotic Multi-Modal Range Sensing System", *Proc. of IEEE Intl Instr. and Meas. Tech. Conf.*, pp. 1991-1996, Ottawa, ON, May 2005.