

Computational Methods for Selective Acquisition of Depth Measurements in Machine Perception

Pierre Payeur, Phillip Curtis

School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, ON, Canada
[ppayeur, pcurtis]@eecs.uottawa.ca

Ana-Maria Cretu

Department of Computer Science and Engineering
Université du Québec en Outaouais
Gatineau, QC, Canada
ana-maria.cretu@uqo.ca

Abstract—Simultaneous acquisition of depth and texture information, such as that provided by RGB-D sensors, finds an ever increasing number of applications, including objects modeling, human-machine interfaces, and robot navigation. One of the challenges resulting from the use of densely populated 3D datasets originates from the massive acquisition, management and processing of the data generated. This reality often preempts full usage of the information available for autonomous systems to make educated decisions. Current methods for reducing dataset's dimension remain independent from the content of the model and therefore do not optimize the balance between the richness of the measurements and their compression. This paper presents two computational methods to selectively drive the selection of depth measurements over the most significant regions of a scene, characterized by their 3D features distribution, while capitalizing on the knowledge readily available in previously acquired depth data. One of the methods builds on self-organizing neural networks, namely neural gas, while the second one computes an empirical improvement metric. Both techniques are adapted to automatically establish which subset of depth measurements within a range sensor's field of view contribute most to the representation of the scene, and therefore streamline the depth measurements acquisition process.

Index Terms—3D imaging, depth measurement, selective sensing, computational intelligence, neural gas, machine perception.

I. INTRODUCTION

The ever increasing 3D acquisition capabilities of vision sensors now provide advanced possibilities to generate textured 3D models of an environment or specific objects. However, a large fraction of the data acquired by sensors such as RGB-D cameras, laser range finders, LIDARs or stereo-cameras contain substantial correlation, which leads to redundant information, large model size, lengthy acquisition, and heavy data processing. Reducing the complexity of such datasets proves essential for autonomous machines to perform subsequent decisions on the resulting data at a reasonable computational cost. Current solutions for dimensionality reduction in range data rely either on predefined pattern-based or random subsampling, where user provided input is expected as to the desired density and regularity of sampling, or the minimum distance between samples. This proves difficult as the user is not always aware of the appropriate level of accuracy required for a given model to be further processed adequately, which leads to a lengthy trial-and-error process.

However, a reduction of the redundancy in the data, immediately upon acquisition, can be also accomplished by initiating the acquisition with only a coarse collection of depth measurements, and then selecting regions of interest, characterized by rich depth features, to focus on for further refinement. In order to perform such selective sensing, regions of similar stochastic properties and continuity must be separated from each other in order to determine what areas need to be enhanced in the model. This research focuses on the design of innovative approaches to achieve automatic selection of regions of observation for range and RGB-D sensors. Identified regions guide a sensor to collect only the most relevant measurements, without human guidance, expedite the acquisition process, and prevent a data avalanche.

Two original and different computational methods for automated selective depth acquisition procedures are detailed and experimentally evaluated. Both methods begin with an initial sparse and rapidly acquired subset on 3D points over the surface of a scene. In the first method, the regression process of a neural gas network in the training phase is used to adaptively identify areas of interest for further scanning in order to improve the accuracy of the model with only a reduced number of depth measurements. In the second method, a formal improvement metric, which expands on the classical interpolation technique of ordinary Kriging [1], is applied to automatically establish which regions within the field of view of a depth camera would provide the most improvement to a model of the scene if further acquisitions were concentrated in priority over those regions. Both methods are evaluated on datasets acquired with the popular Kinect multi-modal imaging sensor, but are designed to be inherently independent of the depth sensing technology used.

II. LITERATURE REVIEW

Various sampling policies have been exploited in the literature in relation with 3D point clouds [2, 3, 4]. Uniform sampling favors a sample distribution where the probability of a surface point to be sampled is equal for all. In random sampling, each point over an 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 and eventually reaches that of uniform sampling. Stratified sampling subdivides the sampling domain into non-overlapping partitions and generates evenly spaced samples by sampling independently from each partition.

Alternatively, Kalaiah and Varshney [5] propose a scheme to compactly decimate and represent point clouds using Principal Component Analysis (PCA). Coherent regions exhibit similar PCA parameters (orientation, frame, mean, variance) and can therefore be classified using clustering and quantization. These methods are not meant to be part of the actual sampling procedure, but rather post-process already collected data.

Pai *et al.* [6, 7] merge the sampling procedure into the measurement process, for modeling deformable objects. The probing procedure considers a known mesh of the object along with parameters such as the maximum force on the object, the maximum probing depth and the number of steps for measuring the deformation. An algorithm generates the next position and orientation for the probe based on the specifications and the object mesh. However, the procedure is not selective and reaches similar complexity as collecting data for all points over the mesh. Shih *et al.* [8] develop different techniques to guide a non-uniform data acquisition process based on a hierarchal tree representation, with error between actual values at leaf nodes and estimated values at those points calculated from the next layer up being used to determine if new points within each sub-division are worthwhile to acquire. The resulting point locations correspond to an optimal scanning pattern for that particular object.

In a different perspective, numerous publications addressed the next best view (NBV) problem which consists of dynamically defining a configuration where a sensor should be placed to maximize the coverage and quality of the model of a scene, while minimizing the amount of separate acquisitions. Connolly [9] proposes a method based on octrees generated from multiple views to determine optimal viewing vectors from the current knowledge of the scene. Several researchers also investigated active view selection [10, 11]. Morooka *et al.* [12] define a discretized shell around a region to limit the number of possible viewing vectors and optimize the entire process. Mackinnon *et al.* [13] rely on several additional fields of data provided by a laser range sensor to derive quality metrics for each acquisition point in order to drive the NBV process and optimize the quality of the overall model.

Other research analyzed optimal fixed scanning patterns for specific scenarios. Ho and Saripalli [14] investigate scanning patterns for autonomous underwater vehicles (AUV) which attempt to maximize coverage and quality, while minimizing energy use from the AUV propulsion system. English *et al.* [15] use a Lissajous, a rosette, and a spiral scanning pattern, along with an adaptive algorithm to swap between them depending on the characteristics and objects detected in the scene, with the goal of optimizing the estimation of position and orientation for automated space docking operations.

III. MEASUREMENTS SELECTION WITH NEURAL GAS

An adaptive computational approach for intelligent depth data acquisition is developed to support active machine perception. The proposed automated selective scanning scheme builds upon a self-organizing neural network to select regions of interest for further refinement. 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, leading to an efficient way to compress data. The neural gas network is selected over other self-organizing architectures due to its capability to capture fine details, unlike others that tend to smooth them. The neural gas algorithm can be described as follows [16]: A set S of network nodes is initialized to contain N units c_i with the corresponding reference vectors $w_{c_i} \in \mathfrak{R}^n$ (each unit c has an associated n -dimensional reference vector that indicates its position in the input space) chosen randomly according to a probability density function $p(x)$ or from a set $D = \{x_1, x_2, \dots, x_M | x_i \in \mathfrak{R}^n\}$. The winning neuron, namely the one that best matches an input vector x is identified using the minimum Euclidean distance criterion:

$$s(x) = \operatorname{argmin}_{c \in S} \|x - w_c\| \quad (1)$$

where $\|\cdot\|$ denotes the Euclidean vector norm. The neurons to be adapted during the learning procedure are selected according to their rank in an ordered list of distances between their weights and the input vector. When a new input vector x is presented to the network, a neighborhood ranking indices list is built (j_0, \dots, j_{N-1}) , where w_{j_0} is the weight of the closest neuron to x , w_{j_1} the weight of the second-closest neuron, and w_{j_k} is the reference vector such that k vectors w_i exist with: $\|x - w_i\| \leq \|x - w_{j_k}\|$. The weights of the neurons to be updated are calculated as follows:

$$w_j(t+1) = w_j(t) + \alpha(t)h_\lambda(k_j(x, w_j))[x(t) - w_j(t)] \quad (2)$$

where $\alpha(t) \in [0, 1]$ describes the overall extent of the modification, and h_λ is 1 for $k_j(x, w_j) = 0$ and decays to zero for higher values according to:

$$h_\lambda(k_j(x, w_j)) = \exp(-k_j(x, w_j)/\lambda(t)) \quad (3)$$

where $k_j(x, w_j)$ is a function that represents the ranking of each weight vector w_j . If j is the closest to input x then $k = 0$, for the second closest $k = 1$ and so on. The learning rate $\alpha(t)$ and the function $\lambda(t)$ are both time-dependent. These parameters are decreased slowly during the learning process in order to ensure that the algorithm converges. The following time dependencies are used, as in [17]:

$$\alpha(t) = \alpha_o(\alpha_T/\alpha_o)^{t/T}, \lambda(t) = \lambda_o(\lambda_T/\lambda_o)^{t/T} \quad (4)$$

where the constants α_o and λ_o are the initial values for $\alpha(t)$ and $\lambda(t)$, α_T and λ_T are the final values, t is the time step and T the training length. The algorithm continues to generate random input signals x while $t < T$.

Starting from an initial sparsely scanned sample of 3D points over an object, the neural gas network with a predefined number of nodes is trained to adapt its nodes to the point cloud. The number of nodes is chosen according to the size of the initial scan as described in [17]. In the context of the current work, it varies from 1400 to 2800 for the different objects. The other parameters are set as follows: $\alpha_o = 0.5$ and λ_o is set equal to half the number of neurons in the initial map. Through this process, the nodes in the neural gas map converge toward regions where varying features and edges are located, which produces clusters of points in regions where more pronounced depth variations are present. The process is illustrated in Fig. 1 for the Stanford bunny [18].

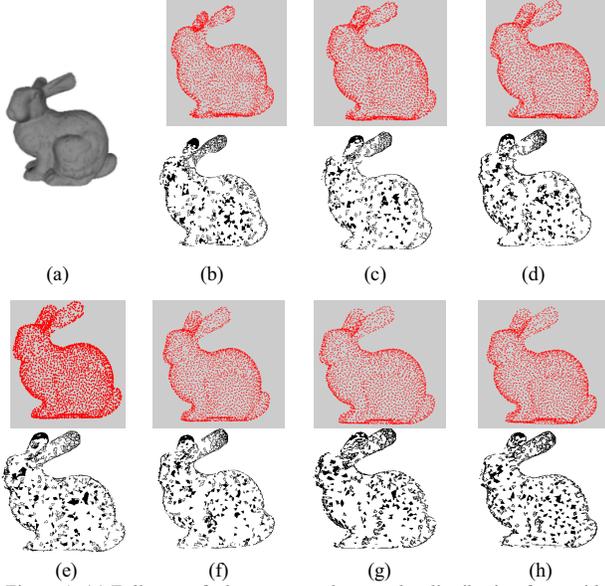


Figure 1. (a) Full scan of a bunny, neural gas nodes distribution for a grid of 45×50 (up), and identified areas of interest (down) for training (T) of (b) 5 epochs, (d) 10, (e) 25, (f) 40, (g) 60, and (h) 100 epochs.

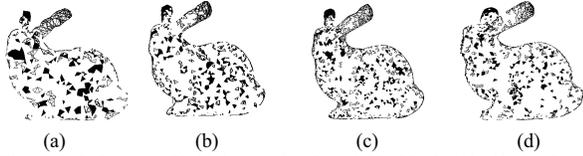


Figure 2. Influence of initial scan size for (a) 32×32 , (b) 40×40 , (c) 64×64 , and (d) 75×75 uniformly sampled points.

Fig. 1a shows the full 3D point cloud of the bunny, containing 35947 points, and the results obtained when using an initial sparse scan of approximately 64×64 (4200 points), and an initial number of nodes in the neural gas map of 45×50 . The upper rows show in red the distribution of nodes in the neural gas map after different training epochs. Regions that require additional sampling to refine the model are detected by finding higher density areas in the neural gas map. This is done by first applying a Delaunay triangulation to the neural gas output map. Areas of high density of nodes are represented by smaller triangles in the tessellation. The mean value of the length of vertices between every pair of nodes for every triangle is set as a threshold. Subsequently, all the edges of triangles that are larger than this threshold are removed from the tessellation. The removal of the edges longer than the threshold ensures the identification of close points and, therefore, dense areas of features. Such features are marked by darker areas in the lower rows of Fig. 1. The subset of remaining nodes extracted from the neural gas map drives the rescanning over the regions of interest to acquire extra samples of depth measurements. A denser multi-resolution model can then be constructed by selectively augmenting the initial sparse low-resolution point cloud with the extra data.

Fig. 1 also shows that it is beneficial to stop the training after a reduced number of epochs to ensure that the nodes capture details rather than becoming uniformly distributed.

This effect is visible in Fig. 1, where the best extraction of featured areas is obtained for 40-60 epochs (Fig. 1f and 1g), after which an increase in the number of epochs leads to more uniformly spread of nodes and the loss of details around the eye area and the paws (Fig. 1h). Also, the selection of areas improves as expected with an increased density in the sparse scan, as shown in Fig. 2 that illustrates the identified areas of interest for various numbers of points in the initial scan.

IV. MEASUREMENTS SELECTION WITH IMPROVEMENT METRIC

An alternative computational method is introduced that extends on the interpolation formalism of Kriging [1] to formulate an original and computationally efficient improvement metric which serves to dynamically guide further acquisition of depth measurements over regions of interest. By monitoring a relative improvement map which gets computed solely on the basis of data acquired at any given stage in the acquisition process, the data can be effectively compressed at acquisition time, while ensuring both an appropriate level of coverage of the overall scene and a sufficient level of quality in the 3D model created.

Kriging is an estimation technique that uses the stochastic properties of current measurements to estimate the measurements at other locations, while minimizing the estimation variance. Its advantage to the context of selective sampling of depth measurements is that it provides both an estimate of a value at a location, and an estimate of the variance on that estimate. Ordinary Kriging, defined by eq. (5)-(10), relies on the estimation of a semivariogram model, $\gamma(h)$, which is a graph that relates how much variation to expect over a given distance, h . In order to have the semivariogram be related to measured data, and as a result to the Kriging system, the semivariogram model is fit to the empirical semivariance of the measured data.

$$K = \begin{bmatrix} \gamma(h(p_1, p_1)) & \gamma(h(p_1, p_2)) & \cdots & \gamma(h(p_1, p_n)) & 1 \\ \gamma(h(p_2, p_1)) & \gamma(h(p_2, p_2)) & \cdots & \gamma(h(p_2, p_n)) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(h(p_n, p_1)) & \gamma(h(p_n, p_2)) & \cdots & \gamma(h(p_n, p_n)) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \quad (5)$$

$$k(\hat{p}_j) = [\gamma(h(p_1, \hat{p}_j)) \quad \gamma(h(p_2, \hat{p}_j)) \quad \cdots \quad \gamma(h(p_n, \hat{p}_j)) \quad 1]^T \quad (6)$$

$$Z = [z(p_1) \quad z(p_2) \quad \cdots \quad z(p_n) \quad 0]^T \quad (7)$$

$$\lambda(\hat{p}_j) = K^{-1}k(\hat{p}_j) \quad (8)$$

$$\hat{z}(\hat{p}_j) = \lambda^T(\hat{p}_j)Z \quad (9)$$

$$\hat{\sigma}^2(\hat{p}_j) = \lambda^T(\hat{p}_j)k(\hat{p}_j) \quad (10)$$

Capitalizing on this framework, and in order to determine optimal locations to acquire future range measurements, a formal measure of potential improvement that any particular point can contribute to the overall 3D representation of the scene is derived. Since it is desired to have an estimation of how the error in the estimation is reduced when a previously unknown point is acquired, the measure of error that is used for determining the estimation of improvement measure is the variance to mean ratio (VMR), $vmr(\hat{p}_j)$. This takes advantage of the fact that ordinary Kriging provides both the estimated depth, $\hat{z}(\hat{p}_j)$, and the estimated variance of the estimation, $\hat{\sigma}^2(\hat{p}_j)$, for an unmeasured point, \hat{p}_j . The VMR also reflects the fact that typically, and for most range sensors, as a depth

measurement is located further from the sensor, the error on the measurement increases, and is inherently normalized in the formulation of the VMR, defined as follows.

$$vmr(\hat{p}_j) = \frac{\hat{\sigma}^2(\hat{p}_j)}{\hat{z}(\hat{p}_j)} \quad (11)$$

Now, if in the future, an acquisition is made at a point, p_s , it will result in a depth measurement, $z(p_s)$. In order to predict the effects of this acquisition before it occurs, the assumption is made that the estimated depth value for that point is the actual value, namely that $p_s = \hat{p}_s$ and $z(p_s) = \hat{z}(\hat{p}_s)$. This assumption results in the formulation of eq. (12), which represents the new VMR at unmeasured point, \hat{p}_j , given the previous assumption on point p_s . The difference between these two values contributes to the novel formulation for a measure of improvement, eq. (13), indicating how much the overall knowledge acquired on \hat{p}_s via a future range acquisition will improve the estimates of all \hat{p}_j in the neighborhood of \hat{p}_s , or in other words, how much improvement in the representation of the scene is estimated to be gained by the acquisition of \hat{p}_s .

$$vmr(\hat{p}_j|\hat{p}_s) = \frac{\hat{\sigma}^2(\hat{p}_j|\hat{p}_s)}{\hat{z}(\hat{p}_j|\hat{p}_s)} \quad (12)$$

$$imp(\hat{p}_s) = \sum_{j=1}^m vmr(\hat{p}_j) - vmr(\hat{p}_j|\hat{p}_s) \quad (13)$$

By substituting the formulations of the vmr , eq. (11) and (12), into eq. (13), eq. (14) can be developed.

$$imp_{SM}(\hat{p}_s) = \sum_{j=1}^m \frac{\hat{\sigma}^2(\hat{p}_j)}{\hat{z}(\hat{p}_j)} - \frac{\hat{\sigma}^2(\hat{p}_j|\hat{p}_s)}{\hat{z}(\hat{p}_j|\hat{p}_s)} \quad (14)$$

As $\hat{\sigma}^2(\hat{p}_j)$ and $\hat{z}(\hat{p}_j)$ are calculated directly from a Kriging system, eq. (9) and (10), only the relationships for $\hat{\sigma}^2(\hat{p}_j|\hat{p}_s)$ and $\hat{z}(\hat{p}_j|\hat{p}_s)$ need to be developed, which are the results of the Kriging system when \hat{p}_s is included among the measured points, namely the system that can be formed by eq. (15)-(20). Furthermore by comparing the two systems represented by eq. (5)-(10) and eq. (15)-(20), it is observed that the former contains the later, with the addition of new terms, which is determined using the original Kriging system where $\hat{p}_j = \hat{p}_s$.

$$\hat{R}(\hat{p}_s) = \begin{bmatrix} \gamma(h(p_1, p_1)) & \gamma(h(p_1, p_2)) & \dots & \gamma(h(p_1, p_n)) & \gamma(h(p_1, \hat{p}_s)) & 1 \\ \gamma(h(p_2, p_1)) & \gamma(h(p_2, p_2)) & \dots & \gamma(h(p_2, p_n)) & \gamma(h(p_2, \hat{p}_s)) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma(h(p_n, p_1)) & \gamma(h(p_n, p_2)) & \dots & \gamma(h(p_n, p_n)) & \gamma(h(p_n, \hat{p}_s)) & 1 \\ \gamma(h(\hat{p}_s, p_1)) & \gamma(h(\hat{p}_s, p_2)) & \dots & \gamma(h(\hat{p}_s, p_n)) & \gamma(h(\hat{p}_s, \hat{p}_s)) & 1 \\ 1 & 1 & \dots & 1 & 1 & 0 \end{bmatrix} \quad (15)$$

$$\hat{k}(\hat{p}_j|\hat{p}_s) = [\gamma(h(p_1, \hat{p}_j)) \quad \dots \quad \gamma(h(p_n, \hat{p}_j)) \quad \gamma(h(\hat{p}_s, \hat{p}_j)) \quad 1]^T \quad (16)$$

$$\hat{Z}(\hat{p}_s) = [z(p_1) \quad \dots \quad z(p_n) \quad z(\hat{p}_s) \quad 0]^T \quad (17)$$

$$\lambda(\hat{p}_j|\hat{p}_s) = \hat{R}(\hat{p}_s)^{-1} \hat{k}(\hat{p}_j|\hat{p}_s) \quad (18)$$

$$\hat{z}(\hat{p}_j|\hat{p}_s) = \lambda^T(\hat{p}_j|\hat{p}_s) \hat{Z}(\hat{p}_s) \quad (19)$$

$$\hat{\sigma}^2(\hat{p}_j|\hat{p}_s) = \lambda^T(\hat{p}_j|\hat{p}_s) \hat{k}(\hat{p}_j|\hat{p}_s) \quad (20)$$

Hence, eq. (21)-(22) can now be developed.

$$\hat{\sigma}^2(\hat{p}_j|\hat{p}_s) = \hat{\sigma}^2(\hat{p}_j) + \frac{\left(k^T(\hat{p}_j)\lambda(\hat{p}_s) - \gamma(h(\hat{p}_s, \hat{p}_j))\right)^2}{\gamma(h(\hat{p}_s, \hat{p}_s)) - \hat{\sigma}^2(\hat{p}_s)} \quad (21)$$

$$\hat{z}(\hat{p}_j|\hat{p}_s) = \hat{z}(\hat{p}_j) + \frac{\left(k^T(\hat{p}_j)\lambda(\hat{p}_s) - \gamma(h(\hat{p}_s, \hat{p}_j))\right) \left(\hat{z}(\hat{p}_s) - z(\hat{p}_s)\right)}{\gamma(h(\hat{p}_s, \hat{p}_s)) - \hat{\sigma}^2(\hat{p}_s)} \quad (22)$$

After substituting these results into the improvement equation, eq. (13), and by using the properties of the semivariogram, namely that the semivariance over zero distance is zero, the improvement equation in terms of the original Kriging system can be formed, as follows.

$$imp(\hat{p}_s) = \frac{1}{\hat{\sigma}^2(\hat{p}_s)} \sum_{j=1}^m \frac{\left(k^T(\hat{p}_j)\lambda(\hat{p}_s) - \gamma(h(\hat{p}_s, \hat{p}_j))\right)^2}{\hat{z}(\hat{p}_j)} \quad (23)$$

From this improvement equation it can be observed that there will be m terms summed together which depend on both \hat{p}_s and \hat{p}_j , and furthermore, the improvement value is to be computed at each of these points. In the current formulation, each of these terms needs to be calculated for each \hat{p}_s and then summed. But by careful selection of the semivariogram model, $\gamma(h)$, an improvement formulation is created where the summation does not depend on \hat{p}_s , and hence the summation depending on \hat{p}_j only needs to be performed once. Selecting the semivariogram model defined in eq. (24) meets this criteria, when $w = 2$. Unfortunately this case yields an impermissible semivariogram, where the resulting Kriging system may have singular matrices. To solve this issue, an approximation where $w = 2 - \varepsilon$ is used in the Kriging system when determining the terms of the Kriging system to prevent the use of singular matrices.

$$\gamma(h(\hat{p}_s, \hat{p}_j)) = ah^w(\hat{p}_s, \hat{p}_j) + bU(h(\hat{p}_s, \hat{p}_j)) \quad (24)$$

$$h(\hat{p}_s, \hat{p}_j) = \sqrt{(\hat{x}_s - \hat{x}_j)^2 + (\hat{y}_s - \hat{y}_j)^2} \quad (25)$$

$$U(h) = \begin{cases} 0, & h = 0 \\ 1, & h > 0 \end{cases} \quad (26)$$

where \hat{p}_s and \hat{p}_j are located at the coordinates (\hat{x}_s, \hat{y}_s) and (\hat{x}_j, \hat{y}_j) respectively, while a and b are the fitting parameters of the semivariogram model. Combining the semivariogram model where $w = 2$, eq. (24), fitted on the local neighborhood of \hat{p}_s with the improvement measure, eq. (23), and expanding the formulation, a final estimated improvement, eq. (27), is developed for all locations within the field of view of a sensor.

$$\begin{aligned} imp(\hat{p}_s) = & \frac{1}{\hat{\sigma}^2(\hat{p}_s)} \left(\lambda^T(\hat{p}_s) \left(\sum_{j=1}^m \frac{k(\hat{p}_j)k^T(\hat{p}_j)}{\hat{z}(\hat{p}_j)} \right) \lambda(\hat{p}_s) - (2a(\hat{x}_s^2 + \right. \\ & \left. \hat{y}_s^2) + 2b) \left(\sum_{j=1}^m \frac{k^T(\hat{p}_j)}{\hat{z}(\hat{p}_j)} \right) \lambda(\hat{p}_s) - 2a \left(\sum_{j=1}^m \frac{k^T(\hat{p}_j)(\hat{x}_j^2 + \hat{y}_j^2)}{\hat{z}(\hat{p}_j)} \right) \lambda(\hat{p}_s) + \right. \\ & 4a\hat{x}_s \left(\sum_{j=1}^m \frac{k^T(\hat{p}_j)\hat{x}_j}{\hat{z}(\hat{p}_j)} \right) \lambda(\hat{p}_s) + 4a\hat{y}_s \left(\sum_{j=1}^m \frac{k^T(\hat{p}_j)\hat{y}_j}{\hat{z}(\hat{p}_j)} \right) \lambda(\hat{p}_s) + \\ & \left. (a^2(\hat{x}_s^2 + \hat{y}_s^2)^2 + 2ab(\hat{x}_s^2 + \hat{y}_s^2) + b^2) \left(\sum_{j=1}^m \frac{1}{\hat{z}(\hat{p}_j)} \right) + \right. \\ & \left. (2a^2(\hat{x}_s^2 + \hat{y}_s^2) + 2ab) \left(\sum_{j=1}^m \frac{(\hat{x}_j^2 + \hat{y}_j^2)}{\hat{z}(\hat{p}_j)} \right) - (4a^2(\hat{x}_s^2 + \hat{y}_s^2)\hat{x}_s + \right. \\ & 4ab\hat{x}_s) \left(\sum_{j=1}^m \frac{\hat{x}_j}{\hat{z}(\hat{p}_j)} \right) - (4a^2(\hat{x}_s^2 + \hat{y}_s^2)\hat{y}_s + 4ab\hat{y}_s) \left(\sum_{j=1}^m \frac{\hat{y}_j}{\hat{z}(\hat{p}_j)} \right) + \\ & a^2 \left(\sum_{j=1}^m \frac{(\hat{x}_j^2 + \hat{y}_j^2)^2}{\hat{z}(\hat{p}_j)} \right) - 4a^2\hat{x}_s \left(\sum_{j=1}^m \frac{(\hat{x}_j^2 + \hat{y}_j^2)\hat{x}_j}{\hat{z}(\hat{p}_j)} \right) - \\ & 4a^2\hat{y}_s \left(\sum_{j=1}^m \frac{(\hat{x}_j^2 + \hat{y}_j^2)\hat{y}_j}{\hat{z}(\hat{p}_j)} \right) + 4a^2\hat{x}_s^2 \left(\sum_{j=1}^m \frac{\hat{x}_j^2}{\hat{z}(\hat{p}_j)} \right) + \\ & \left. 8a^2\hat{x}_s\hat{y}_s \left(\sum_{j=1}^m \frac{\hat{x}_j\hat{y}_j}{\hat{z}(\hat{p}_j)} \right) + 4a^2\hat{y}_s^2 \left(\sum_{j=1}^m \frac{\hat{y}_j^2}{\hat{z}(\hat{p}_j)} \right) \right) + \frac{2b}{\hat{\sigma}^2(\hat{p}_s)} - \frac{b^2}{\hat{\sigma}^2(\hat{p}_s)\hat{z}(\hat{p}_s)} \end{aligned} \quad (27)$$

This formulation leads to a bi-dimensional improvement map where areas of higher potential improvement are put in evidence, similarly to the clusters of nodes obtained with the neural gas approach described in section III.

V. EXPERIMENTAL EVALUATION

The evaluation of the proposed computational methods is performed using range images acquired with the Microsoft Kinect for Xbox 360 sensor which uses a structured light approach to efficiently generate a dense distribution of depth measurements. Data acquisition was accomplished with the open source OpenNI drivers, the depth sensor resolution set at 640x480, over scenes at about 2 m depth which provides a spatial resolution of 1 cm along the depth axis.

Three different scenes are considered here to support the experimental evaluation. The first case consists of a standard computer workstation exhibiting various planar surfaces with different reflectance characteristics, as shown in Fig. 3a. The second scene is that of a large exercise ball, shown in Fig. 3b, which is selected for its curved and smooth surface. Finally, a more elaborate scene, composed of a fire hose station surrounded by pipes over a flat wall, shown in Fig. 3c, supports the validation of the computational methods over complex shapes and a wider range of depth values. All scenes are initially acquired with the Kinect sensor in order to provide datasets, shown in Fig. 4, from which a collection of sparse depth measurements is extracted via uniform subsampling to initialize the selective sensing procedure. As reported in section III, the selection of depth measurements to acquire in the future tends to improve with an increase in the density of the initial sparse scan. Therefore a compromise between the quality, the compactness, and the speed of acquisition of the resulting model was experimentally determined. It involves a uniform subsampling to be performed over the raw data to extract uniformly distributed 3D point clouds composed of 128x128 depth measurements over each of the three scenes. This subsampling provides the initial rough acquisition to initialize the measurements selection procedures, which are both based on previously acquired knowledge about a scene rather than on user selected parameters.

The approach based on neural gas is applied on every dataset, initially subsampled at a 128x128 density, and the resulting location of dense neural gas nodes highlights the regions of interest where further acquisitions are worthwhile to be conducted to refine the definition of the scene. In this case the regions identified for further exploration are marked by dark triangles in the center column of Fig. 5. The improvement metric method is similarly applied on the datasets and an improvement map is computed, following the methodology described in section IV. The resulting improvement maps are displayed in the right column of Fig. 5. Brighter (white) areas represent those with the highest potential for contributing to increase the knowledge about a scene, and darker regions (black) are those where further time and energy spent at acquiring depth measurements is not likely to contribute significantly to knowledge and accurate modeling of the scene. Gray pixels map intermediate improvement potential on a continuous 0-1 (black-to-white) scale.

One can notice in the set of comparative figures that the two methods succeed to identify, in spite of their different approach, most of the areas that require additional scanning to

improve the model. In the current implementations, only depth information is used to monitor regions of interest over which further acquisition should be prioritized. This is motivated by the fact that the methods were developed to accommodate a diversity of range sensors, including laser triangulation and LIDAR sensors that do not provide color or texture information. Close examination of Fig. 5 also confirms the sharpness of the regions of interest identified by both measurements selection methods while relying initially only on a very low percentage (about 5%) of the information about the scene to further drive depth measurements collection.

For the neural gas nodes distributions, the smaller the density of the initial scan, the smaller number of nodes is needed to extract the topology of the scene, but more training epochs are required to ensure the correct identification of regions. This approach also calls for a number of internal thresholds to be tuned in order to optimize the outcome, which is not the case for the formal improvement metric approach.

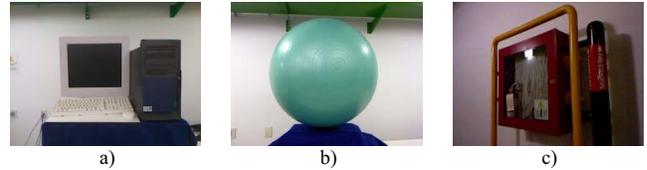


Figure 3. Three scenes supporting the experimental evaluation: computer workstation, exercise ball, and fire hose station.

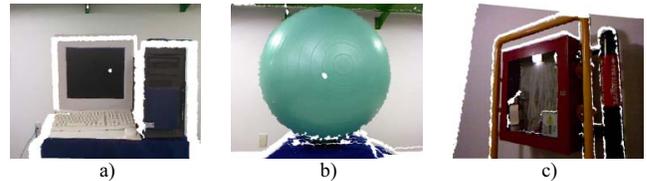


Figure 4. RGB-D datasets acquired with Kinect sensor.

Scene	Measurements selection with neural gas	Measurements selection with improvement metric
Computer		
Exercise ball		
Fire hose station		

Figure 5. Depth measurements selection computational methods applied on three different scenes acquired with Kinect sensor.

The correspondence between regions of interest identified by both methods is evidenced in all sets of results. However, the improvement metric method tends to highlight the edges and contours of components of the scene, where depth transitions occur, as denoted by white pixels in all improvement maps. The method therefore concentrates in the areas of transition between the shape of the object and the background, or between various components of the scene at different depths, resulting in a clean definition of the object boundaries. On the other hand, the neural gas method concentrates clusters of nodes over sections of the surface of the objects. The complex fire hose station scene exemplifies this behavior. The neural gas nodes tend to obtain regions that are overall more uniformly spread, resulting in the identification of regions over the surface of the object. As a result, the improvement metric method appears as a very efficient technique for edge detection in depth maps or 3D models. Alternatively, the neural gas measurement selector provides an efficient approach to rapidly acquire a compact representation of a scene from only a very sparse set of measurements. Both methods can therefore find application in rapid scene understanding and object recognition, beyond their suitability to dynamically drive the acquisition process with random access range or RGB-D sensors.

TABLE 1. Computing time for obtaining the neural gas nodes distribution (NG), and the improvement map (ImpMap) on objects acquired with Kinect.

Computer		Exercise ball		Fire hose	
NG	ImpMap	NG	ImpMap	NG	ImpMap
153.0 s	1.39 s	153.9 s	1.41 s	150.0 s	1.41 s

Tables 1 summarizes the computation time required to obtain the neural gas nodes distribution and the improvement map, both marking the regions of interest. A significant difference is observed in between the computing time required to obtain neural gas nodes distribution and improvement maps. However, these are reported only to demonstrate independence to the scene complexity of both methods, and not as absolute values given that the two computational methods are implemented in different manners and on different platforms. While the improvement metric method is coded in optimized C++ language and executed on an Intel I7-2630 processor operating at 2.6 GHz, the neural gas method currently runs on the Matlab platform and is not optimized.

VI. CONCLUSION

Two computational methods for the selective acquisition of depth measurements with range sensors in autonomous machine perception have been detailed. An experimental validation demonstrated the effectiveness of the proposed techniques to selectively and automatically determine which regions of a scene should be prioritized for the acquisition of supplementary data to progressively enhance knowledge about that scene while minimizing the load of the acquisition process and further data processing. Such a capability proves essential when operating slower range sensors or random access laser range sensors, as the acquisition can be significantly

accelerated. The methods also find application with faster sensors, such as Kinect devices, to efficiently detect the boundaries, location and shape of objects, and therefore support the operation of objects recognition processes.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support of Ontario Centres of Excellence, and Natural Sciences and Engineering Research Council of Canada to this research.

REFERENCES

- [1] G. Bohling, "Kriging", University of Kansas, 2005, [online]. Available: <http://people.ku.edu/~gbohling/cpe940/Kriging.pdf> [Accessed 12 4 2013].
- [2] M. Pauly, M. Gross, and L.P. Kobbelt, "Efficient Simplification of Point-Sampled Surfaces", *IEEE Conf. Visualization*, pp. 163-170, 2002.
- [3] D. Uesu, L. Bavoil, S. Fleishman, J. Shepherd, and C.T. Silva, "Simplification of Unstructured Tetrahedral Meshes by Point Sampling", *IEEE Int. Workshop on Volume Graphics*, E. Groller, I Fujishio (Eds.), pp. 157-238, 2005.
- [4] D. Nehab and P. Shilane, "Stratified Point Sampling of 3D Models", *Eurographics Symp. on Point-Based Graphics*, M. Alexa, S. Rusinkiewicz (Eds.), pp. 49-56, 2004.
- [5] A. Kaliah and A. Varshney, "Statistical Point Geometry", *Eurographics Symp. on Geometry Processing*, K. Kobbely, P. Schroder and H. Hoppe (Eds.), pp. 107-115, 2003.
- [6] D.K. Pai, K. van der Doel, D.L. James, J. Lang, J.E. Lloyd, J.L. Richmond, and S.H. Yau, "Scanning Physical Interaction Behavior of 3D Objects", *Computer Graphics and Interactive Techniques*, pp. 87-96, 2001.
- [7] 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.
- [8] C. S. Shih, L. A. Gerhardt, C.-C. Williams, C. Lin, C.-H. Chang, C.-H. Wan, and C.-S. Koong, "Non-uniform Surface Sampling Techniques for Three-dimensional Object Inspection," *Optical Engineering*, vol. 47, no. 5, 053606, May 2008.
- [9] C. I. Connolly, "The Determination of Next Best Views", *IEEE Int. Conf. on Robotics and Automation*, pp. 432-435, Mar. 1985.
- [10] V. Sequeira, J.G.M. Goncalves, and M.I. Ribeiro, "Active View Selection for Efficient 3D Scene Reconstruction", *IEEE Int. Conf. on Pattern Recognition*, Vienna, Austria, vol. 1, p. 815-819, Aug. 1996.
- [11] J. Maver and R. Bajcsy, "Occlusions as a Guide for Planning the Next View", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no 5, pp. 417-433, May 1993.
- [12] K. Morooka, H. Zha, and T. Hasegawa, "Computations on a Spherical View Space for Efficient Planning of Viewpoints in 3-D Object Modeling", *IEEE Int. Conf. on 3-D Digital Imaging and Modeling*, pp. 138-147, Oct. 1999.
- [13] D. MacKinnon, V. Aitken, and F. Blais, "Adaptive Laser Range Scanning using Quality Metrics", *IEEE Instrumentation and Measurement Technology Conf.*, Victoria, BC, pp. 348-353, May 2008.
- [14] C. Ho and S. Saripalli, "Where Do You Sample? - An Autonomous Underwater Vehicle Story", *IEEE Int. Symp. on Robotic and Sensors Environments*, Montreal, QC, pp. 119-124, Sep. 2011.
- [15] C. English, G. Okouneva, P. Saint-Cyr, A. Choudhuri, and T. Luu, "Real-Time Dynamic Pose Estimation Systems in Space: Lessons Learned for System Design and Performance Evaluation", *Int. Journal of Intelligent Control and Systems*, vol. 16, no 2, pp. 79-96, 2011.
- [16] T.M. Martinetz, S.G. Berkovich, and K.J. Schulten, "Neural-Gas Network for Vector Quantization and its Application to Time-Series Prediction", *IEEE Trans. Neural Networks*, vol.4, no 4, pp. 558-568, 1993.
- [17] A.-M. Cretu, P. Payeur, and E.M. Petriu, "Selective Range Data Acquisition Driven by Neural Gas Networks", *IEEE Trans. on Instrumentation and Measurement*, vol. 58, no 8, pp. 2634-2642, 2009.
- [18] ---, The Stanford 3D Scanning Repository, [online]. Available: <http://graphics.stanford.edu/data/3Dscanrep/> [Accessed 12 5 2013].