# Computational Methods for Selective Acquisition of Depth Measurements: An Experimental Evaluation

Pierre Payeur[1], Phillip Curtis[1], and Ana-Maria Cretu[2]

[1] School of Electrical Engineering and Computer Science,
University of Ottawa 800 King Edward, Ottawa, ON, Canada
{ppayeur,pcurtis}@eecs.uOttawa.ca
[2] Department of Computer Science and Engineering,
Université du Québec en Outaouais 101 Saint-Jean-Bosco, Gatineau, QC, Canada
ana-maria.cretu@uqo.ca

**Abstract.** Acquisition of depth and texture with vision sensors finds numerous applications for objects modeling, man-machine interfaces, or robot navigation. One challenge resulting from rich textured 3D datasets resides in the acquisition, management and processing of the large amount of data generated, which often preempts full usage of the information available for autonomous systems to make educated decisions. Most subsampling solutions to reduce 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 experimentally evaluates the performance achieved with two computational methods that selectively drive the acquisition of depth measurements over regions of a scene characterized by higher 3D features density, while capitalizing on the knowledge readily available in previously acquired data. Both techniques automatically establish which subsets of measurements contribute most to the representation of the scene, and prioritize their acquisition. The algorithms are validated on datasets acquired from two different RGB-D sensors.

**Keywords:** 3D imaging, depth measurement, RGB-D cameras, computational intelligence, selective sampling, neural gas.

## 1 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. Acquiring, coding, interpreting and transmitting all of this information is a complex task, which contributes to what is known as the 'Big Data Challenge' [1]. Reducing the complexity of datasets proves essential to perform subsequent decisions on the resulting data at a reasonable

J. Blanc-Talon et al. (Eds.): ACIVS 2013, LNCS 8192, pp. 389–401, 2013.

computational cost. Current solutions for dimensionality reduction in range data rely either on predefined pattern-based or random subsampling, where a user input is expected as to the desired sampling density, 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.

However, a reduction of the redundancy in the data, immediately upon acquisition, can also be accomplished by initiating the acquisition with only a coarse collection of depth measurements, and then selecting regions of interest, characterized by rich depth features, within this acquisition 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 and evaluation of innovative approaches to achieve automatic selection of regions of acquisition for range and RGB-D sensors, in order for a sensor to collect only the most relevant measurements without human guidance, and as a result, expedite the acquisition process. The relevant regions of interest are extracted from 3D point clouds during the acquisition procedure to prevent an avalanche of data.

Two original and different computational methods recently introduced by the authors in [2, 3] are reported and experimentally compared in the context of RGB-D imaging to determine their relative performance and to develop guidelines for the implementation of automated selective depth acquisition procedures. 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. In the second method, a formal improvement measure, which expands on the classical interpolation technique of ordinary Kriging [4], 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 from datasets acquired with the popular Kinect multi-modal imaging sensor and a custom RGB-D structured light sensor, but are designed to be inherently independent of the depth sensing technology used.

## 2 Literature Review

Three sampling policies have been largely explored in the literature in relation with 3D point clouds [5, 6, 7]. 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 [8] 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 operate as post-processing on collected data.

Pai *et al*. [9, 10] 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 exerted on the object, the probing depth and the number of steps for the deformation measurement. An algorithm generates the next pose for the probe based on the specifications and the object mesh. However, the procedure is not selective and therefore reaches similar complexity as collecting data for all points over the mesh. Shih *et al.* [11] develop different techniques to guide a non-uniform data acquisition process based on a hierarchal tree representation, with error between actual values at the leaf nodes and the 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 define the optimal scanning pattern for that particular object.

In a different perspective, numerous publications have addressed the next best view (NBV) problem which consists of dynamically defining a configuration where a sensor should be positioned and oriented in order to maximize the coverage and quality of the model of a scene, while minimizing the amount of separate acquisitions required. Connolly [12] proposed a method based on octrees generated from multiple views to determine optimal viewing vectors based on the current knowledge of the scene. Active view selection was investigated by several researchers [13, 14]. Morooka *et al.* [15] define a discretized shell around a region to limit the number of possible viewing vectors, which allows the use of lookup tables to optimize the entire process. Mackinnon *et al.* [16] rely on several additional fields of data provided by a laser range sensor to derive a quality metric for each acquisition point in order to drive the NBV process and optimize the quality of the overall model.

There has also been research that looked into optimal fixed scanning patterns for various scenarios. Ho and Saripalli [17] 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*. [18] use three different patterns, 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.

## 3 Measurement Selection with Neural Gas

An adaptive computational approach for intelligent depth acquisition was developed by the authors in [2]. Meant to be an active part of the sampling procedure, the automated selective scanning scheme builds upon a self-organizing neural network to select regions of interest for further refinement. A self-organizing architecture is chosen for its ability to quantize a 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 other architectures that tend to smooth them. The neural gas algorithm can be described as follows [19]: A set $S$ of network nodes is initialized to contain $N$ units $c_i$ with the corresponding reference vectors $w_{c_i} \in \Re^n$ (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 $D = \{x_1, x_2, \ldots, x_M | \, x_i \in \Re^n\}$. The winning neuron, namely the one that best matches an input vector $x$ is identified using the minimum Euclidean distance:

$$s(x) = argmin_{c \in S} \|x - w_c\| , \tag{1}$$

where ||.|| 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 ($j_0, \ldots, j_{N-1}$), 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:

$$w_j(t+1) = w_j(t) + \alpha(t) h_\lambda \left(k_j(x, w_j)\right) [x(t) - w_j(t)] , \tag{2}$$

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

$$h_\lambda \left(k_j(x, w_j)\right) = exp\left(-k_j(x, w_j)/\lambda(t)\right) , \tag{3}$$

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 [19]:

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

where the constants $\alpha_o$ and $\lambda_o$ are the initial values for $\alpha(t)$ and $\lambda(t)$, $\alpha_T$ and $\lambda_T$ 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 [2]. In the current work, it varies from 1400 to 3000 for the different objects. Through this process, the nodes in the neural gas map converge toward regions where features and edges are located, which produces clusters of points in regions where more pronounced variations are present in the geometric shape. The training is stopped early by reducing the number of training epochs, to ensure that the nodes capture details rather than becoming uniformly distributed.

Regions that require additional sampling to ensure an accurate model are detected by finding higher density areas in the neural gas output map. A Delaunay

triangulation is first applied to the neural gas map. Areas of high density of nodes are represented by small 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 these edges ensures the identification of close points and, therefore, dense areas of features. The subset of remaining nodes extracted from the neural gas map drives the rescanning over the regions of interest to acquire extra samples of 3D points. A model can then be constructed by selectively augmenting the initial sparse point cloud with the extra data samples.

## 4 Measurement Selection with Improvement Metric

More recently, an alternative computational method was introduced by the authors [3] that extends on the interpolation formalism of Kriging [4] 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 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 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 relies on the estimation of a semivariogram model, which is a graph that relates how much variation to expect over a given distance. In order to have the semivariogram be related to measured data, the semivariogram model is fit to the empirical semivariance of the measured data.

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 as the basis in 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 appropriately 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:

$$vmr(\hat{p}_j) = \frac{\hat{\sigma}^2(\hat{p}_j)}{\hat{z}(\hat{p}_j)}. \tag{5}$$

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 leads to the formulation of eq. (6), which represents the new VMR at unmeasured point, $\hat{p}_j$, given the previous assumption on point $p_s$. The difference between the former and the new VMR values leads to the formulation of a measure of improvement, eq. (7), indicating how much the knowledge acquired on $\hat{p}_s$ via a future range acquisition will improve the estimates of all points, $\hat{p}_j$, in the neighborhood of $\hat{p}_s$, or how much improvement in the model of the scene is estimated to be achieved by the acquisition of $\hat{p}_s$.

$$vmr(\hat{p}_j|\hat{p}_s) = \frac{\hat{\sigma}^2(\hat{p}_j|\hat{p}_s)}{\hat{z}(\hat{p}_j|\hat{p}_s)}. \tag{6}$$

$$imp(\hat{p}_s) = \sum_{j=1}^{m} vmr(\hat{p}_j) - vmr(\hat{p}_j|\hat{p}_s)\,. \tag{7}$$

Combining the semivariogram model fitted on readily available data with the improvement measure based on the variance to mean ratio, a final 'unrolled' estimated improvement, eq. (8), is developed [3] for all locations in the field of view of a range sensor, which 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 3:

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

where $m$ is the number of points in the neighborhood of $\hat{p}_s$, $\hat{p}_s$ and $\hat{p}_j$ are located at the coordinates $(\hat{x}_s, \hat{y}_s)$ and $(\hat{x}_j, \hat{y}_j)$ respectively, $a$ and $b$ are the fitting parameters of the semivariogram model, $\lambda(\hat{p}_j)$ is the ordinary Kriging weight vector corresponding to point $\hat{p}_j$, and $k(\hat{p}_j)$ is the ordinary Kriging measured-points-to-estimated-point semivariance vector corresponding to point $\hat{p}_j$.

## 5 Depth Sensing Technologies and Datasets

The evaluation of the proposed computational methods is performed here using a series of range images acquired, on one side, from the popular Microsoft Kinect for Xbox 360 platform, and on the other side, from a custom RGB-D sensor called Adaptive Structured Light Sensor (ASLS) developed in our laboratory that supports a larger

depth of field. The Kinect RGB-D camera uses an IR camera and an IR projector to generate a structured light pattern. Data acquisition was accomplished using the open source OpenNI drivers, with the depth sensor resolution set at 640x480. The Kinect sensor has a 57° horizontal, and a 43° vertical field of view and depth sensing provides reliable data between 0.8m and 3.5m, with a depth resolution at 2m being about 10mm [20, 21]. The ASLS [22, 23] capitalizes on adaptive structured light sensing with a visible marching pseudo-random pattern projected onto a scene to generate features that are imaged by a stereoscopic pair of cameras. A time-domain multiplexing strategy projects a three-color pattern, where any 3x3 code block is unique and supports reliable stereo matching. Multi-focal capability is also integrated to further increase the operational range of the sensor. The configuration of the ASLS creates a maximum field of view of approximately 41.4°x31.7°, and a theoretical quantization error of 39.5mm at 10m depth. Due to its adaptive characteristics, the sensor can provide depth readings over a wide variety of surfaces, but takes longer to acquire a scene in high detail, which further substantiates the need for selective sensing.

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. 1a. The second scene is that of a large exercise ball, shown in Fig. 1b, 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. 1c, supports the validation of the computational methods over complex shapes and a wider range of depth values. All scenes are initially acquired with both sensors in order to provide datasets from which a coarse collection of depth measurements is extracted via uniform subsampling, at various densities, to initialize the selective sensing procedure. The datasets for the three scenes are also displayed in Fig. 1, respectively for the Kinect sensor and for the adaptive structured light sensor (ASLS).

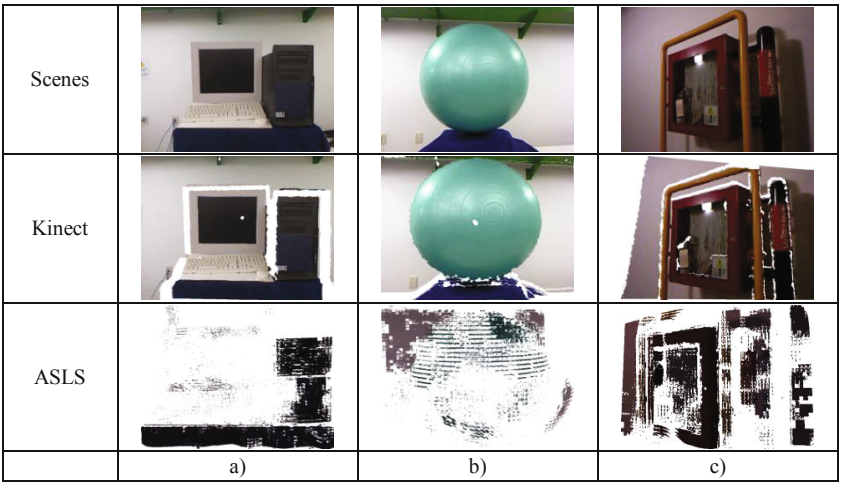


**Fig. 1.** Three scenes supporting the experimental evaluation: a) computer workstation, b) exercise ball, c) fire hose station, and RGB-D data acquired with Kinect and ASLS sensors

Comparing the two datasets, one easily notices the completeness and sharpness of the RGB-D information generated by the Kinect sensor. In comparison, the ASLS did not provide a similar density of depth measurements. White regions in the second and third rows of Fig. 1 correspond to locations where RGB-D measurements were not acquired over the scenes. The ASLS sensor also generated a large number of outliers in its datasets which have been removed here to better support the comparative evaluation, as they would otherwise have appeared as features and erroneously attracted the attention of the measurement selectors. Nevertheless it is interesting to study the performance of the two measurement selection techniques over datasets with different characteristics in order to monitor their ability to accommodate various means of acquisition.

## 6 Experimental Evaluation

This section examines the behavior and performance of both measurement selection techniques, while assuming an initial coarse scan of depth measurements is available. For comparison purposes, initial uniform subsampling is performed over the raw data to extract uniformly distributed 3D point clouds composed of 32x32, 64x64, and 128x128 depth measurements over each of the three scenes, and for both the Kinect and ASLS datasets respectively. This subsampling plays the role of an initial rough acquisition of a few measurements to initialize the measurements selection procedure, given that the methods rely on a priori acquired knowledge about a scene and not on user selected parameters to drive the acquisition. This makes the computational approaches fully automated and adaptive to the contents and nature of any scene.

In the case of the improvement metric method, an improvement map is computed for each of the three initial subsampling densities, following the methodology described in section 4 and the resulting improvement maps are displayed in the second and third columns of Fig. 2-4, respectively for the computer, exercise ball, and fire hose station datasets acquired with the Kinect and ASLS sensors. 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.

The approach based on neural gas is similarly applied on every dataset, initially subsampled at the same densities, and the resulting location of dense neural gas nodes highlights the regions of interest where further acquisitions are worthwhile to be performed to refine the definition of the scene. In this case the regions identified for further exploration are marked by dark triangles in the two last columns of Fig. 2-4.

One can notice in the set of comparative figures that the two methods succeed to consistently identify, in spite of their different approaches, 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. In the special case where full RGB-D content is available, such as with the Kinect and ASLS

sensors, this extra dimensionality of the data space can be taken advantage of to further refine the clustering of regions of interest. This aspect is not considered in the experimental tests reported in this section.

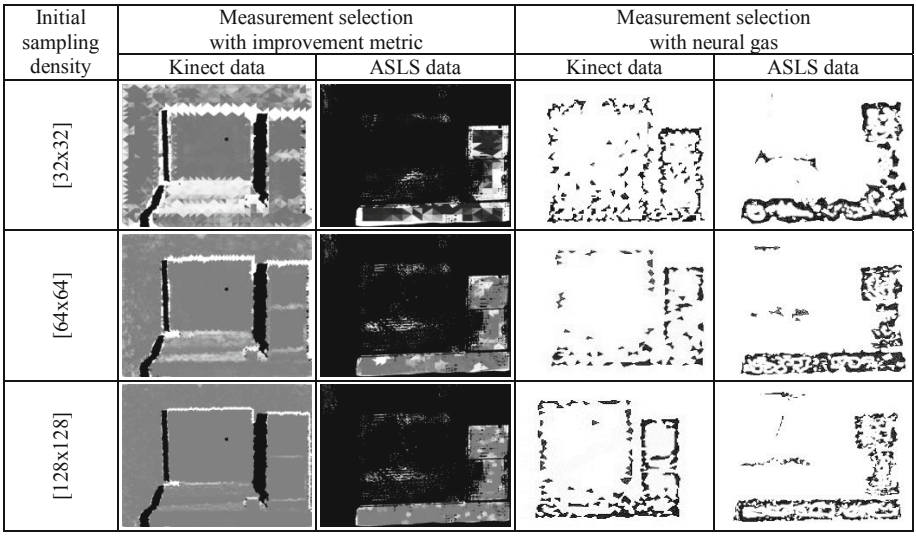


**Fig. 2.** Measurement selection computational methods applied on computer workstation acquired with Kinect and ASLS sensors

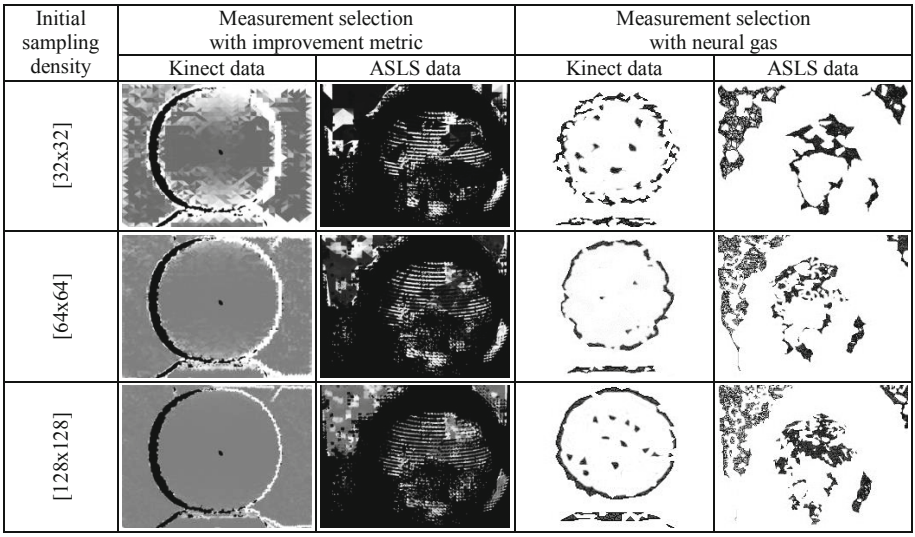


**Fig. 3.** Measurement selection computational methods applied on exercise ball acquired with Kinect and ASLS sensors

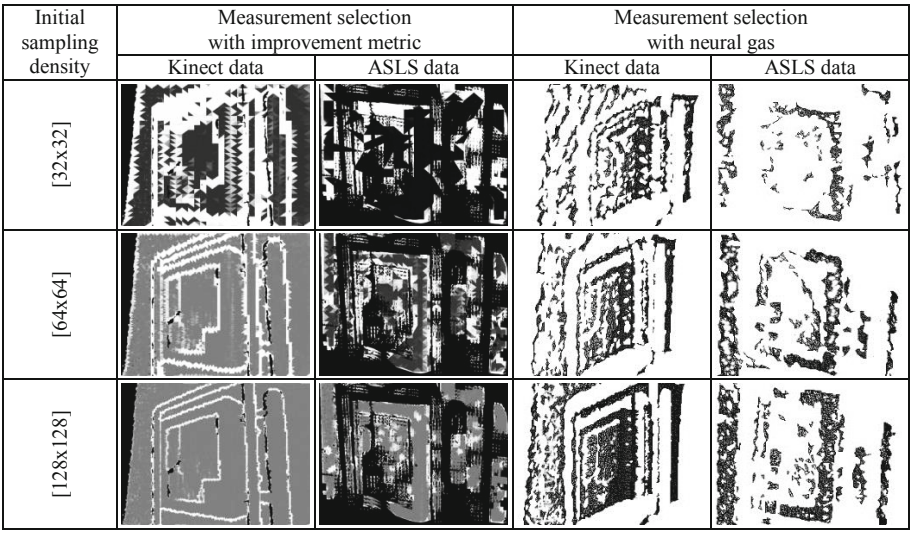


**Fig. 4.** Measurement selection computational methods applied on fire hose station acquired with Kinect and ASLS sensors

Close examination of Fig. 2-4 confirms that the sharpness of the regions of interest identified by both measurement selection methods improves with the density of the coarse scan of the scene used to initialize the process. The results also demonstrate that the techniques adapt well to the datasets, independently from their completeness. In the case of the Kinect sensor, the original data is cleaner and denser than with the ASLS, ensuring for both methods a sharper definition of areas of interest. For the neural gas nodes distributions, the smaller the density of the initial scan, a smaller number of nodes is needed to extract the topology of the scene, but more training epochs are required in general to ensure the correct identification of regions.

In cases where a larger number of areas are not properly acquired by the sensor, as can be observed for the computer and the ball scenes acquired with the ASLS, both measurement selection methods focus their mapping over regions where knowledge is available. This behavior is expected given that depth features are only detectable over those areas. A modification to the improvement map method that is currently under development aims at addressing this issue by introducing a mechanism to force a balance between accuracy (improving knowledge over already acquired areas) and coverage (improving knowledge over missed areas).

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, especially those with finer initialization provided by 128x128 initial subsampling density. 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 points 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 improvement map (ImpMap) and neural gas nodes distribution (NG) from various initial sampling densities on objects acquired with each sensor

| | Sensor | Computer | | Exercise ball | | Fire hose | |
|---|---|---|---|---|---|---|---|
| | | ImpMap | NG | ImpMap | NG | ImpMap | NG |
| [32x32] | Kinect | 0.68 s | 9.5 s | 0.66 s | 9.6 s | 0.66 s | 9.3 s |
| [64x64] | Kinect | 0.82 s | 37.2 s | 0.83 s | 37.5 s | 0.85 s | 36.5 s |
| [128x128] | Kinect | 1.39 s | 153.0 s | 1.41 s | 153.9 s | 1.41 s | 150.0 s |
| [32x32] | ASLS | 13.4 s | 9.8 s | 16.7 s | 8.9 s | 19.5 s | 8.8 s |
| [64x64] | ASLS | 25.5 s | 36.8 s | 19.8 s | 35.8 s | 40.0 s | 32.3 s |
| [128x128] | ASLS | 42.4 s | 161.6s | 15.2 s | 142.3s | 81.7 s | 140.2s |

Table 1 summarizes the computation time required to obtain respectively the improvement map and neural gas node distribution that mark regions of interest. A significant difference is observed in between the computing time required to obtain improvement maps and neural gas nodes distribution. As can be observed from Table 1, the NG method scales near linearly with the number of points acquired in the subsampling, while the ImpMap method scales sub-linearly, although the NG method provides more consistent timing results regardless of the dataset and source processed. When acquisition is performed with slower range scanners, the methods are efficient enough to be embedded in the sensor and dynamically drive the acquisition process to collect measurements in priority over regions that contribute the most to increase the knowledge about the scene, that is, focus on regions that are rich in depth features. On the contrary, when used in conjunction with rapid RGB-D sensors, like the Kinect technology, advantage can be taken of the proposed computational methods to rapidly acquire an understanding of the content of a scene.

## 7 Conclusion

This experimental evaluation of two computational methods for the selective acquisition of measurements with RGB-D sensors demonstrates the effectiveness of the proposed techniques to selectively and automatically determine which regions of a scene best support the acquisition of supplementary data to progressively enhance knowledge about that scene while reducing the amount of data required to understand the nature of a scene. Such a capability proves essential when operating slower RGB-D

sensors, such as the ASLS, or random access laser range sensors, as the acquisition can be interrupted at an earlier stage, when points that truly contribute to knowledge about the scene are already acquired. The methods also find applications with faster range sensors to efficiently detect the location and shape of objects, and support the operation of recognition processes when used as contour extractors from depth data.

## References


1. Weiss, L.G.: Autonomous Robots in the Fog of War. IEEE Spectrum, 30-34 & 56-57 (2011)
2. Cretu, A.-M., Payeur, P., Petriu, E.M.: Selective Range Data Acquisition Driven by Neural Gas Networks. IEEE Trans. on Instrumentation and Measurement 58(8), 2634–2642 (2009)
3. Curtis, P., Payeur, P.: A Method for Dynamic Selection of Optimal Depth Measurements Acquisition with Random Access Range Sensors. In: Canadian Conf. on Computer and Robot Vision, pp. 311–318 (2013)
4. Bohling, G.: Kriging, University of Kansas, http://people.ku.edu/~gbohling/cpe940/Kriging.pdf
5. Pauly, M., Gross, M., Kobbelt, L.P.: Efficient Simplification of Point-Sampled Surfaces. IEEE Conf. on Vizualization, 163–170 (2002)
6. Uesu, D., Bavoil, L., Fleishman, S., Shepherd, J., Silva, C.T.: Simplification of Unstructured Tetrahedral Meshes by Point Sampling. In: Groller, E., Fujishio, I. (eds.) IEEE Intl. Workshop on Volume Graphics, pp. 157–238 (2005)
7. Nehab, D., Shilane, P.: Stratified Point Sampling of 3D Models, Eurographics. In: Alexa, M., Rusinkiewicz, S. (eds.) Symp. on Point-Based Graphics, pp. 49–56 (2004)
8. Kalaiah, A., Varshney, A.: Statistical Point Geometry.Eurographics. In: Kobbely, K., Schroder, P., Hoppe, H. (eds.) Symp. on Geometry Processing, pp. 107–115 (2003)
9. Pai, D.K., van der Doel, K., James, D.L., Lang, J., Lloyd, J.E., Richmond, J.L., Yau, S.H.: Scanning Physical Interaction Behavior of 3D Objects. Computer Graphics and Interactive Techniques, 87-96 (2001)
10. Lang, J., Pai, D.K., Woodham, R.J.: Acquisition of Elastic Models for Interactive Simulation. Intl. Journal of Robotics Research 21(8), 713–733 (2002)
11. Shih, C.S., Gerhardt, L.A., Williams, C.-C., Lin, C., Chang, C.-H., Wan, C.-H., Koong, C.-S.: Non-uniform Surface Sampling Techniques for Three-dimensional Object Inspection. Optical Engineering 47(5), 053606 (2008)
12. Connolly, C.I.: The Determination of Next Best Views. In: IEEE Intl. Conf. on Robotics and Automation, pp. 432–435 (1985)
13. Sequeira, V., Goncalves, J.G.M., Ribeiro, M.I.: Active View Selection for Efficient 3D Scene Reconstruction. In: IEEE Intl. Conf. on Pattern Recognition, vol. 1, pp. 815–819 (1996)
14. Maver, J., Bajcsy, R.: Occlusions as a Guide for Planning the Next View. IEEE Trans. on Pattern Analysis and Machine Intelligence 15(5), 417–433 (1993)
15. Morooka, K., Zha, H., Hasegawa, T.: Computations on a Spherical View Space for Efficient Planning of Viewpoints in 3-D Object Modeling. In: IEEE Intl Conf. on 3-D Digital Imaging and Modeling, pp. 138–147 (1999)
16. MacKinnon, D., Aitken, V., Blais, F.: Adaptive Laser Range Scanning using Quality Metrics. In: IEEE Instrumentation and Measurement Technology Conf., pp. 348–353 (2008)

17. Ho, C., Saripalli, S.: Where Do You Sample? - An Autonomous Underwater Vehicle Story. In: IEEE Intl. Symposium on Robotic and Sensors Environments, pp. 119–124 (2011)
18. English, C., Okouneva, G., Saint-Cyr, P., Choudhuri, A., Luu, T.: Real-Time Dynamic Pose Estimation Systems in Space: Lessons Learned for System Design and Performance Evaluation. Intl. Journal of Intelligent Control and Systems 16(2), 79–96 (2011)
19. Martinetz, T.M., Berkovich, S.G., Schulten, K.J.: Neural-Gas Network for Vector Quantization and its Application to Time-Series Prediction. IEEE Trans. on Neural Networks 4(4), 558–568 (1993)
20. Khoshelham, K.: Accuracy Analysis of Kinect Depth Data. In: ISPRS Workshop Laser Scanning (2011)
21. Macknojia, R., Chávez-Aragón, A., Payeur, P., Laganière, R.: Experimental Characterization of Two Generations of Kinect's Depth Sensors. In: IEEE Intl. Symposium on Robotic and Sensors Environments, pp. 150–155 (2012)
22. Boyer, A., Curtis, P., Payeur, P.: 3D Modeling from Multiple Views with Integrated Registration and Data Fusion. In: Canadian Conf. on Computer and Robot Vision, pp. 252–259 (2009)
23. Boyer, A.: Adaptive Structured Light Imaging for 3D Reconstruction and Autonomous Robotic Exploration, University of Ottawa, Thesis (2009)