# A Method for Dynamic Selection of Optimal Depth Measurements Acquisition with Random Access Range Sensors

Phillip Curtis and Pierre Payeur
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, ON, Canada
[pcurtis, ppayeur]@eecs.uOttawa.ca

***Abstract*—It is well established that acquiring large amount of range data with vision sensors can quickly lead to important data management challenges where processing capabilities become saturated and preempt full usage of the information available for autonomous systems to make educated decisions. While sub-sampling offers a naïve solution for reducing dataset dimension after acquisition, it does not capitalize on the knowledge available in already acquired data to selectively and dynamically drive the acquisition process over the most significant regions in a scene, the latter being generally characterized by variations in depth and surface shape. This paper discusses the development of a formal improvement measure and a method to automatically establish which regions within the field of view of a range sensor would provide the most improvement to a model of the scene if further acquisitions were concentrated in priority over those regions. The proposed algorithm mainly targets applications using random access range sensors, defined as sensors that can acquire depth measurements at specified azimuth and elevation within their field of view. However, the framework is developed to be independent of the range sensing technology used, and is validated with range data acquired from the popular Kinect multi-modal imaging sensor, as well as Neptec's LMS laser random access range sensor.**

***Keywords— improvement map, range measurement, 3D imaging, random access range sensors, smart sensing, selective sensing, Kinect***

## I. INTRODUCTION

With the continuous improvement of range imaging technologies, the overall quality and the amount of data they produce has increased tremendously. This resulted in a larger number of range sensors being used in numerous applications, from entertainments systems with low requirements, to reliable security applications, efficient robotic platforms, and up to high-end ultra-fine artifacts virtual representations. Acquiring, coding, interpreting and transmitting all of this rich and dense information from a wide variety of sensors is a complex task, which contribute to current challenges involved in what is known as the 'Big Data Challenge' [1].

While this work does not deal with the related issues of handling and combining data across multiple sensors and sharing of that data, it addresses the specific issue of efficient 3D data acquisition. One way to tackle the Big Data Challenge for range imaging consists of minimizing the actual amount of data points acquired in a range image. This strategy simultaneously permits to expedite the acquisition process, which proves important with slower imaging technologies, while also resulting in more compact datasets.

A promising approach to achieve this goal aims at identifying regions within the field of view of the sensor which need a higher density of points for a model of the scene to be represented and interpreted accurately, and other areas which do not. In the context of this work, range acquisition is restrained to a single point of view of the sensor in order to ensure a proper experimental evaluation of the performance of the proposed method without dependency on accurate data registration considerations between various viewpoints and related 3D point samples. By performing this analysis, the data can be effectively compressed at acquisition time, while ensuring both an appropriate level of coverage of the overall scene and the quality of the 3D model created. By effectively reducing the quantity of data acquired via data driven approaches, a robotic system can achieve savings by reducing transmission bandwidth requirements, reducing the computational burden of 3D vision processing, especially for embedded decision and navigation systems, and in the case of active sensors such as laser range finders, reducing power consumption, while at the same time maintaining a sufficient level of quality in the acquired representation of the scene.

Firstly, a review is presented about relevant techniques for determining where to scan next in the generic case of a multi-view acquisition approach. As a counterpart, techniques for determining optimal scanning patterns, and approaches for intelligent adaptive sensing from a single point of view are also examined. An introduction to ordinary Kriging technique, which the proposed method is based upon, is then presented. Next, the proposed method is developed for estimating the level of improvement to a 3D representation of a scene that an acquisition made at an unmeasured location provides. Finally, the formal improvement measure and related algorithm are verified experimentally and results are presented before concluding remarks are formulated.

## II. LITERATURE REVIEW

Other works in the past have dealt with the issue of identifying to which location a sensor should be moved in order to improve the coverage and quality of the model of a scene, while minimizing the amount of separate acquisitions required. These approaches are well known as next best view (NBV) algorithms.

Connolly [2], through his previous usage of octrees generated from multiple views [3], realized that determining optimal viewing vectors based on the current knowledge of the scene would improve the overall time required to model a scene. He describes two different methods for determining

978-0-7695-4983-5/13 $26.00 © 2013 IEEE
DOI 10.1109/CRV.2013.24

the NBV: by determining the view which would reveal the most 'unseen' nodes in the octree (the planetarium algorithm), and by summing together the normals of the faces of nodes that are common to both 'unseen' and 'empty' nodes in order to produce a viewing vector which sees the greatest amount of potentially visible 'unseen' nodes (the normal algorithm).

The goal of eliminating occlusions to drive the NBV process was investigated by several researchers [4-7]. Morooka *et al*. [8] define a discretized shell around a region to be modeled in order to limit the number of possible viewing vectors, which allows the use of lookup tables to optimize the entire process. Mackinnon *et al*. [9] use a specialized laser range sensor which provides several additional fields of data in order to derive quality metrics, such as resolvability, planarity, orientation, reflectivity, and spot size, for each acquisition point in order to drive the NBV process to optimize the quality of the overall 3D model.

There has also been works that have looked into optimal fixed scanning patterns for several scenarios. Ho and Saripalli [10] have investigated 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*. [11] 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.

Adaptive and intelligent sensing for range acquisition was previously investigated by Cretu *et al*. [12] who determine regions which require higher resolution acquisition based upon an initial coarse scan, from within a single point of view. Their method uses a neural gas network to determine where features and edges are located by training the network over a short period, which produces clusters of points in regions where there are depth features. The resulting clusters are then analyzed, and regions with higher density of clusters, which correspond to regions in space where there are more potential depth features to finely measure, are re-acquired at a higher resolution.

Shih *et al*. [13] develop three different techniques to guide a non-uniform data acquisition process from a single point of view. In the first two approaches, an initial scan of the object is made. This object is then subdivided in a hierarchal tree-type fashion, 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 need to be acquired. The difference between the first two approaches is that the first one uses plane fitting over regularly shaped sections, such as rectangles and triangles, while the second approach uses curve fitting. The third approach determines the optimal non-uniform scanning pattern for a particular object based upon a CAD model, by first performing a virtual acquisition, and then using a local adjustment algorithm to move points around until an optimal placement occurs where the points cease to move. The resulting point locations correspond to the optimal scanning pattern for that particular object.

In order to predict the optimal location over the scene where to acquire a future range point, estimates of the range at certain locations which have not yet been acquired are needed, as well as a reasonable estimate of the error expected on the estimation. Ordinary Kriging is one such estimation technique that provides both an estimate of a value at a location, and an estimate of the variance on that estimate. Kriging in general is an estimation technique that uses the stochastic properties of the current measurements to estimate the measurements at other locations, while minimizing the estimation variance. Ordinary Kriging is a version of the Kriging technique which assumes that the mean of the measured value follows a trend, but essentially remains unchanged in a local neighborhood. The Kriging family of estimation techniques has been used extensively in geosciences [14-17], as well as to perform crop yield estimations [18] and range image filtering [19], but has not yet been fully exploited and adapted to the context of dynamically selective range acquisition.

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. A semivariogram model has three properties which must be met in order for the semivariogram to be permissible, or valid, for use in Kriging [15], namely that the semivariogram must be non-negative, the semivariogram at zero is equal to zero, and that the semivariogram is conditionally negative definite. 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 [15, 20, 21].

Ordinary Kriging is represented by the system of equations shown in eq. (1)-(6), where $h(p_i, p_j)$ is the 2D Euclidean distance between measured points $p_i$ and $p_j$, $\gamma(h)$ is the semivariance from the semivariogram model, $n$ is the number of known measured points, $z(p_i)$ is the depth at the i[th] measured point $p_i$, $\lambda(\hat{p}_j)$ is the vector containing the Kriging weights for the j[th] unmeasured point $\hat{p}_j$, $\hat{z}(\hat{p}_j)$ is the estimated depth, and $\hat{\sigma}^2(\hat{p}_j)$ is the estimated variance.

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

$$k(\hat{p}_j) = \left[\gamma\left(h(p_1,\hat{p}_j)\right) \quad \gamma\left(h(p_2,\hat{p}_j)\right) \quad \ldots \quad \gamma\left(h(p_n,\hat{p}_j)\right) \quad 1\right]^T \quad (2)$$

$$Z = [z(p_1) \quad z(p_2) \quad \cdots \quad z(p_n) \quad 0]^T \quad (3)$$

$$\lambda(\hat{p}_j) = K^{-1}k(\hat{p}_j) \quad (4)$$

$$\hat{z}(\hat{p}_j) = \lambda^T(\hat{p}_j)Z \quad (5)$$

$$\hat{\sigma}^2(\hat{p}_j) = \lambda^T(\hat{p}_j)k(\hat{p}_j) \quad (6)$$

## III. IMPROVEMENT MEASURE

In order to determine the optimal location to acquire the next point in a range image, a measure of improvement that any particular point can contribute to the overall 3D

representation of the scene must be derived. In order to perform this action 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 selection is inspired from the fact that ordinary Kriging provides both the estimated depth, $\hat{z}(\hat{p}_j)$, as well as 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 in eq. (7).

$$vmr(\hat{p}_j) = \frac{\hat{\sigma}^2(\hat{p}_j)}{\hat{z}(\hat{p}_j)} \tag{7}$$

Now, if in the future, an acquisition is made at 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 $z(\hat{p}_s) = \hat{z}(\hat{p}_s)$. This assumption results in the formulation of eq. (8), which represents the new VMR at unmeasured point, $\hat{p}_j$, given the previous assumption on point $\hat{p}_s$. The difference between these two values results in eq. (9), which contributes a novel formulation for a measure of conditional improvement indicating how much the knowledge acquired on $\hat{p}_s$ via a future range acquisition will improve the estimate of $\hat{p}_j$.

$$vmr(\hat{p}_j|\hat{p}_s) = \frac{\hat{\sigma}^2(\hat{p}_j|\hat{p}_s)}{\hat{z}(\hat{p}_j|\hat{p}_s)} \tag{8}$$

$$imp(\hat{p}_j|\hat{p}_s) = vmr(\hat{p}_j) - vmr(\hat{p}_j|\hat{p}_s) \tag{9}$$

Since any additional information about the depth of a given point contributes to the improvement of knowledge about the scene over a certain area, it is not sufficient to only calculate the improvement that $\hat{p}_s$ imparts on the estimate of a particular point, $\hat{p}_j$, but rather to the total improvement over the entire set of points which $\hat{p}_s$ affects. This must be calculated in order to determine the improvement in VMR that the acquisition of $\hat{p}_s$ makes over the entire acquisition space. Therefore the total estimated improvement due to acquiring point $\hat{p}_s$ is defined in eq. (10), where $m$ is the number of points that $\hat{p}_s$ influences.

$$imp(\hat{p}_s) = \sum_{j=1}^{m} imp(\hat{p}_j|\hat{p}_s) \tag{10}$$

In order to determine the optimal point to acquire next in a range image, all unmeasured points must have this improvement value calculated. As these calculations are fairly computationally expensive by themselves, reuse of calculated values is preferred. After some manipulation, eq. (10) can be reformulated in terms of the previously defined ordinary Kriging system, as shown in eq. (11).

$$imp(\hat{p}_s) = \frac{1}{\hat{\sigma}^2(\hat{p}_s)} \sum_{j=1}^{m} \frac{\left(\lambda^T(\hat{p}_j)k(\hat{p}_s) - \gamma\left(h(\hat{p}_s, \hat{p}_j)\right)\right)^2}{\hat{z}(\hat{p}_j)} \tag{11}$$

Since eq. (11) must be calculated for $m$ different points in the regions of influence, and for each possible unmeasured location, that means that the 2D Euclidean distance between unmeasured points, $\hat{p}_s$ and $\hat{p}_j$, must be calculated many times. As the square root operation has a high computational cost associated with it, a semivariogram model that will allow fewer calculations is desired. The ideal scenario would use the distance squared within the semivariogram model, as this allows for the separation of terms belonging to $\hat{p}_s$ and $\hat{p}_j$. Unfortunately, the power model with the exponent equal to 2 produces an impermissible semivariogram model, which may yield a singular $K$ matrix. To avoid this issue an exponent which is close enough to 2 such that there will be little error introduced in the approximation between the semivariogram model and the squared distance function, as is illustrated in Fig. 1, is chosen. This prevents a singular $K$ matrix and will allow the optimization of the calculations being performed using the squared distance. Therefore, the final semivariogram, $\gamma(h)$, that is being used is the combination of the power and nugget permissible models (see eq. (12)). Note that $h$ is the 2D Euclidean distance, $U(h)$ is the unit step function as shown in eq. (13), and the parameters used for fitting the semivariogram model to the empirically calculated semivariances are $a$ and $b$.

$$\gamma(h) = ah^w + bU(h), 0 \le w < 2 \tag{12}$$

$$U(h) = \begin{cases} 0, & h = 0 \\ 1, & h > 0 \end{cases} \tag{13}$$

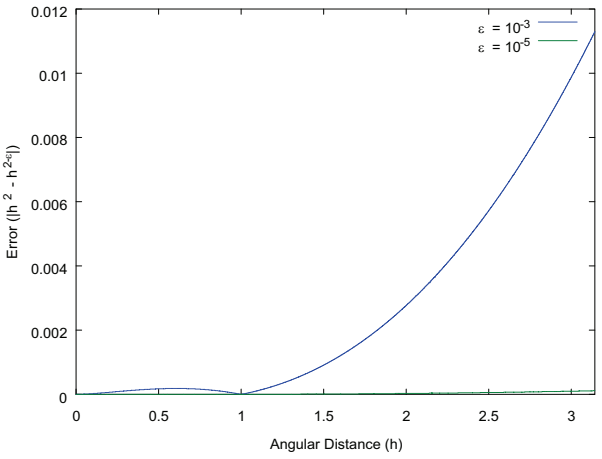



Figure 1. Error between $h^2$ and $h^{(2-\varepsilon)}$ for $\varepsilon=10^{-3}$ and $10^{-5}$.

Combining the semivariogram model from eq. (12) using the distance squared approximation ($w$=2), with the improvement formulation in eq. (11), and the 2D Euclidean distance function from eq. (14) where coordinates of $\hat{p}_s = (\hat{x}_s, \hat{y}_s)$ and $\hat{p}_j = (\hat{x}_j, \hat{y}_j)$, the final 'unrolled' estimated improvement, eq. (15), is developed. Using this formulation, the inner summations of $j$=1 to $m$ can be calculated first, and then those values can be reused for each of the $m$ $imp(\hat{p}_s)$ calculations, thereby cutting down on the number of computations to be linear with respect to $m$ as opposed to quadratic in the approach where squared distance is not used.

$$h(\hat{p}_s, \hat{p}_j) = \sqrt{(\hat{x}_s - \hat{x}_j)^2 + (\hat{y}_s - \hat{y}_j)^2} \tag{14}$$

$$imp(\hat{p}_s) = \frac{1}{\hat{\sigma}^2(\hat{p}_s)}\Bigg(\lambda^T(\hat{p}_s)\left(\sum_{j=1}^{m}\frac{k(\hat{p}_j)k^T(\hat{p}_j)}{\hat{z}(\hat{p}_j)}\right)\lambda(\hat{p}_s) - \left(2a\left(\hat{x}_s{}^2 + \hat{y}_s{}^2\right) + 2b\right)\left(\sum_{j=1}^{m}\frac{k^T(\hat{p}_j)}{\hat{z}(\hat{p}_j)}\right)\lambda(\hat{p}_s) - 2a\left(\sum_{j=1}^{m}\frac{k^T(\hat{p}_j)(\hat{x}_j{}^2+\hat{y}_j{}^2)}{\hat{z}(\hat{p}_j)}\right)\lambda(\hat{p}_s) + 4a\hat{x}_s\left(\sum_{j=1}^{m}\frac{k^T(\hat{p}_j)\hat{x}_j}{\hat{z}(\hat{p}_j)}\right)\lambda(\hat{p}_s) + 4a\hat{y}_s\left(\sum_{j=1}^{m}\frac{k^T(\hat{p}_j)\hat{y}_j}{\hat{z}(\hat{p}_j)}\right)\lambda(\hat{p}_s) + \left(a^2\left(\hat{x}_s{}^2 + \hat{y}_s{}^2\right)^2 + 2ab\left(\hat{x}_s{}^2 + \hat{y}_s{}^2\right) + b^2\right)\left(\sum_{j=1}^{m}\frac{1}{\hat{z}(\hat{p}_j)}\right) + \left(2a^2\left(\hat{x}_s{}^2 + \hat{y}_s{}^2\right) + 2ab\right)\left(\sum_{j=1}^{m}\frac{(\hat{x}_j{}^2+\hat{y}_j{}^2)}{\hat{z}(\hat{p}_j)}\right) - \left(4a^2\left(\hat{x}_s{}^2 + \hat{y}_s{}^2\right)\hat{x}_s + 4ab\hat{x}_s\right)\left(\sum_{j=1}^{m}\frac{\hat{x}_j}{\hat{z}(\hat{p}_j)}\right) - \left(4a^2\left(\hat{x}_s{}^2 + \hat{y}_s{}^2\right)\hat{y}_s + 4ab\hat{y}_s\right)\left(\sum_{j=1}^{m}\frac{\hat{y}_j}{\hat{z}(\hat{p}_j)}\right) + a^2\left(\sum_{j=1}^{m}\frac{(\hat{x}_j{}^2+\hat{y}_j{}^2)^2}{\hat{z}(\hat{p}_j)}\right) - 4a^2\hat{x}_s\left(\sum_{j=1}^{m}\frac{(\hat{x}_j{}^2+\hat{y}_j{}^2)\hat{x}_j}{\hat{z}(\hat{p}_j)}\right) - 4a^2\hat{y}_s\left(\sum_{j=1}^{m}\frac{(\hat{x}_j{}^2+\hat{y}_j{}^2)\hat{y}_j}{\hat{z}(\hat{p}_j)}\right) + 4a^2\hat{x}_s{}^2\left(\sum_{j=1}^{m}\frac{\hat{x}_j{}^2}{\hat{z}(\hat{p}_j)}\right) + 8a^2\hat{x}_s\hat{y}_s\left(\sum_{j=1}^{m}\frac{\hat{x}_j\hat{y}_j}{\hat{z}(\hat{p}_j)}\right) + 4a^2\hat{y}_s{}^2\left(\sum_{j=1}^{m}\frac{\hat{y}_j{}^2}{\hat{z}(\hat{p}_j)}\right)\Bigg) + \frac{2b}{\hat{z}(\hat{p}_s)} - \frac{b^2}{\hat{\sigma}^2(\hat{p}_s)\hat{z}(\hat{p}_s)} \tag{15}$$

## IV. Proposed Algorithm

This section describes the proposed algorithm for creating a map which estimates the predicted improvement, by using eq. (15), at locations that the range sensor can cover within its field of view. The considered scenario of application consists of first acquiring a coarse depth map of the scene by selecting uniformly distributed acquisition azimuth and elevation values from a fixed centric sensor point of view. The dynamic selection process then uses the information available from this limited number of measured samples to predict improvement estimates. The highest improvement areas become the targets for future acquisition with a random access range sensors that can selectively acquire range measurements at specified azimuth and elevation coordinates. The newly acquired 3D points then refine the knowledge about the scene which further drives the progressive refinement of the 3D model of the scene while maximizing the acquisition procedure efficiency in terms of rapid and reliable acquisition of knowledge.

Since calculating the improvement for the whole map after the addition of each future point is computationally expensive, and generally not desired since features tend to have unique local statistics, the algorithm first organizes the available points into local neighborhoods, based on a triangular mesh formed by the points readily available, and then only updating the estimated improvement map in regions where the triangular mesh changes after the addition of a new measurement.

$$\begin{matrix}\theta = atan2(x,z)\\ \varphi = atan2(y,z)\\ \rho = \sqrt{x^2+y^2+z^2}\end{matrix} \overset{Tf}{\leftrightarrow} \begin{matrix}z = \frac{\rho}{\sqrt{1+tan^2(\theta)+tan^2(\varphi)}}\\ x = z\cdot tan(\theta)\\ y = z\cdot tan(\varphi)\end{matrix} \tag{16}$$

The first step in the proposed algorithm is to transform already acquired points from Cartesian space to an angular perspective projection space. This is performed since most range sensors available, such as laser range finders, stereo vision, and structured light systems, are projective based. Consequently, these sensors have a focal point through which their data is collected, allowing the modeling of the collected data in a perspective based coordinate system, as defined by eq. (16), where $z$ coincides with the principal axis of the range sensor.

The next step is to add these points into a 2-D triangular mesh using Delaunay triangulation [22, 23], where the azimuth, $\theta$, and elevation $\varphi$, are used as the 2-D coordinate pair instead of the usual Cartesian coordinates $(x,\ y)$. The use of Delaunay triangulation ensures that the linkages within the mesh between acquired measurements are fairly local, as it prevents long and skinny triangles. The points within the local neighborhood used in the improvement process are found by selecting the three vertices of any triangle, as well as the two or three points which are opposite to the edges in the triangle, as is shown in Fig. 2. By only using a few points in the local neighborhood as opposed to many in a global approach, the computational burden is reduced, and the locality of the results is ensured, at the cost of potential influence of noise.

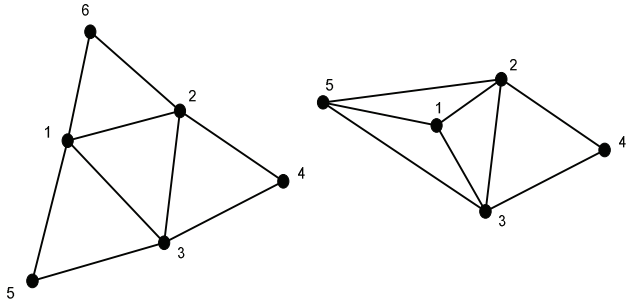



Figure 2. Configuration using 6 points (left), and configuration using 5 points (right).

These five or six points are used to calculate empirical semivariances, which are then used to calculate the parameters $a$ and $b$ using least squares to best fit the semivariogram model in eq. (12) with the points in the local neighborhood. A map of estimated improvement is then calculated, by the application of eq. (15), over the $m$ candidate sampling points that are within the main central triangle. The map in an ideal situation would have a 1:1 mapping from possible sensor acquisition directions to pixels, where the sensor field of view corresponds to the dimensions of the map. For example, for one of the range sensors considered for experimentation, the Microsoft Kinect sensor, a field of view of 57° horizontally and 42° vertically [24] with a depth image resolution of 640x480 is available. This would yield an improvement map of a comparable size with the same resolution covering the same field of view.

A dynamic selection process using this improvement map is implemented by determining the location of the highest improvement value in the map, which has not been yet been visited. Furthermore, as more points are added selectively and progressively into the triangular mesh, using an incremental Delaunay triangulation algorithm, only the triangles which have been modified or added through the acquisition of supplementary points need to have their estimated improvement values recalculated. This further reduces the computational burden of the algorithm.

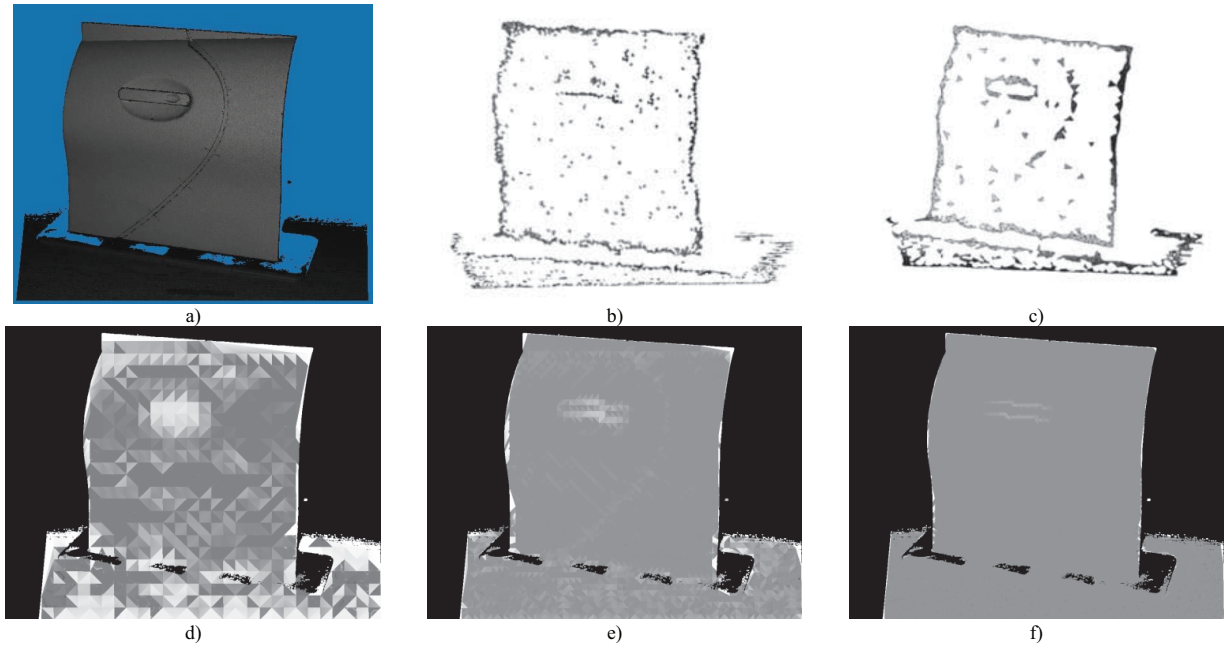


Figure 3. Car door comparison of selective sensing methodologies: a) fully rendered point cloud, b) low resolution neural gas model highlighting features [12], c) medium-low resolution neural gas model, d) extremely low resolution improvement map, e) low resolution improvement map, and f) medium-low resolution improvement map.

## V. Experimental Results

Firstly, a comparison of the improvement map is made against the results obtained by Cretu *et. al.* [12], using the same data which was acquired using a Neptec LMS laser random access range scanner (see Fig. 3a). This is to ensure that similar regions are highlighted. The Neptec LMS has a maximal angular field of view of 30° x 50° [25] and the data set of the car's door region was acquired at 1024x1024 resolution. Fig. 3b shows the neural-gas model from an initial uniform subsampling of the data at 64x64 (low resolution) and Fig. 3c shows the neural-gas model from an initial subsampling of the data at 128x128 (medium-low resolution). Fig. 3d-f show the results of applying the improvement measure algorithm developed in section IV with initial uniform subsamplings of 32x32 (extremely low resolution), 64x64, and 128x128 respectively. The regions of interest in the neural-gas model that highlight features on the object in Fig. 3b and 3c are shown in black, while the regions of interest where the most gain for an acquisition of data is monitored in the improvement maps are shown in white (Fig. 3d-f).

Both methods indicate that the edges of the object, as well as the door handle, and the seam where the door meets the car side panel are the areas where the most improvement can be added. Additionally, it can be qualitatively observed that the results from the improvement map shown in Fig. 3d match closely with that of Fig. 3b, even though only 1/4 of the initial points were used. This trend also exhibits itself when visually comparing Fig. 3e and Fig. 3c. Furthermore the time to generate the results from the subsampled point cloud on an Intel I7-2630 processor operating at a nominal 2.6 GHz was for Fig. 3d of 3.44 seconds, for Fig. 3e of 3.64 seconds, and for Fig. 3f of 4.69 seconds. The results from the neural gas method took 4-5 min on an Intel Pentium IV at 1.3 GHz.

Secondly, the algorithm was experimentally validated on a series of data acquired with the Microsoft Kinect, with the parameters mentioned in the previous section. The value of $w$ in the semivariogram model from eq. (12) was chosen to be 1.99999 (that is $\varepsilon=10^{-5}$), as this provided a sufficient tradeoff between the approximation error and numerical precision for the resolution and field of view of the Kinect sensor. The results of selective sensing using the improvement map are compared to the mean of the results achieved from a random point insertion over 10 runs. Two different scenes are presented, one containing some cones at various depths from within a parking garage (Fig. 6a), and the other containing a cluttered scene of objects acquired within a lab (Fig. 7a). For each, the model was populated with an initial sub-sampling of 10x10 points.

Fig. 4 illustrates the mean estimated improvement over the entire improvement map for both scenes as each acquired 3D point sample is progressively added to the model. In both cases, for each insertion methodology (selective or random), the mean estimated improvement starts off large, and generally decreases to a minimum. Note that when the selection of points is driven using the maximum improvement value within the map, the mean improvement converges to the minimum value faster than what is achieved using random insertion. Fig. 4a does not decay to zero due to the occlusions caused by the stereo setup of the IR projector/camera configuration in the Kinect sensor and the limitations on the types of surfaces the Kinect itself can perceive, and so not all points in the field of view can be acquired. In Fig. 4a and b, the mean estimated improvement decays to be close to its minimum at between 5000 and 10000 points, and does not change significantly beyond that. As a result, the plots have been truncated to better illustrate the earlier behavior.

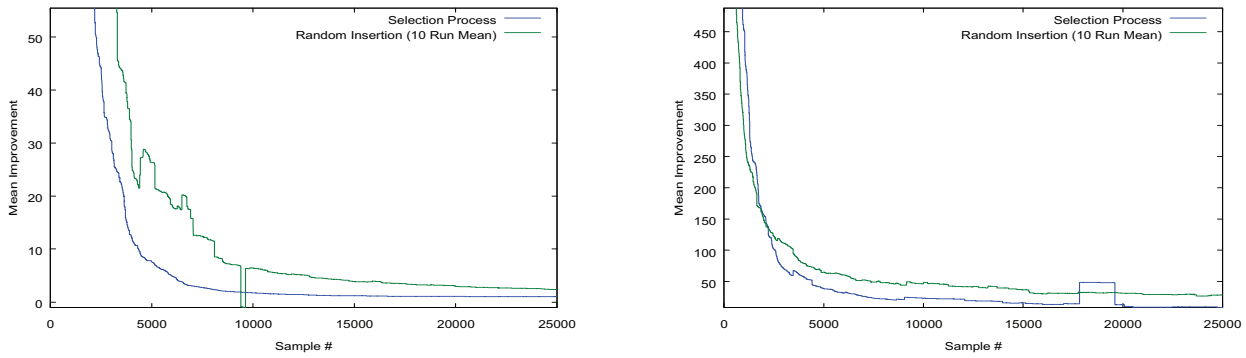



Figure 4. Mean estimated improvement for the scenes containing a) a parking garage with cones and b) a cluttered lab.

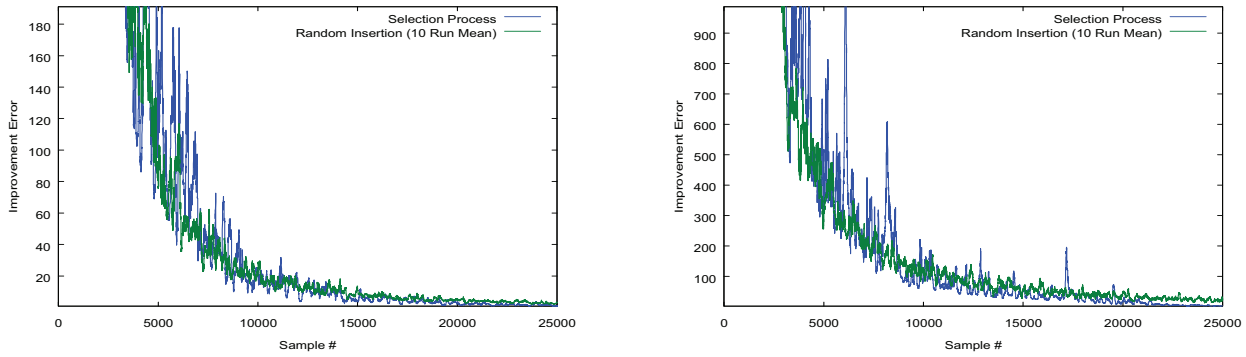



Figure 5. Error of estimated improvement compared to actual improvement for the scenes containing a) a parking garage with cones and b) a cluttered lab.

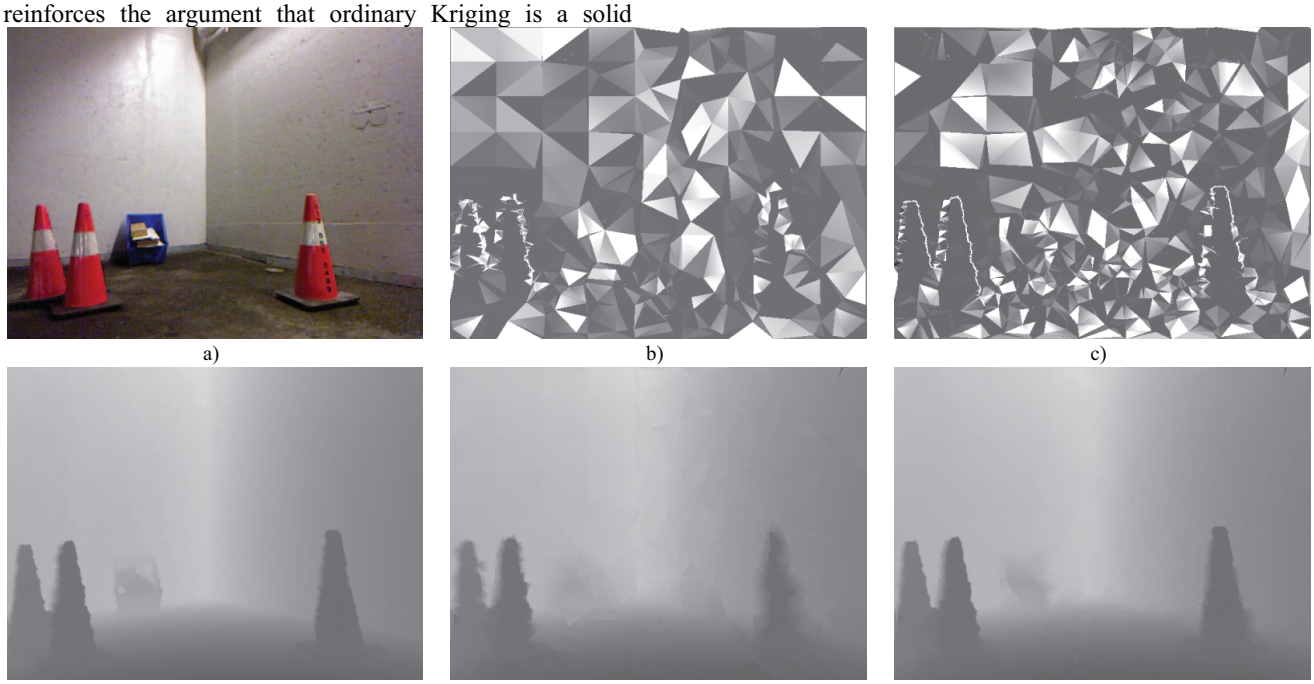
Fig. 5 plots the error between the estimated improvement, and the actual improvement per sample inserted into the 3D model with both insertion methodologies. The actual improvement was determined by calculating the difference in VMR over the entire map before a measurement was added, and after it was added. The plots demonstrate a rapid decrease in error, as the local semivariogram becomes more reflective of the actual underlying statistics of the scene as new points are added. The trend in the improvement error resulting from the dynamic point selection process using the improvement map reaches the minimum prior to that achieved using random point insertion. The plots in Fig. 5 show that the improvement errors for random insertion are less noisy than that of the selection process; this is purely due to the fact that the random insertion curves are calculated from the mean of 10 independent runs. Again, as seen in Fig. 4, curves in Fig. 5 approach a minimum error at between 5000 and 10000 points, and do not change significantly after that, and hence the plots have been truncated to better illustrate the earlier behavior.

Fig. 6b and 7b show the estimated improvement map for both scenes after 2500 points have been inserted, while Fig. 6c and 7c show the estimated improvement map for both scenes after 5000 points have been inserted using the point selection process based on the improvement map. Fig. 6e and 7e present the estimated depth map corresponding to Fig. 6b and 7b, while Fig. 6f and 7f present the estimated depth map corresponding to Fig. 6c and 7c. Fig. 6d and 7d illustrate the estimated depth map after all points have been inserted into the respective model, as quality and completeness comparatives of selective sensing to models achieved from full acquisition.

Whiter colored regions in the improvement map correspond to larger improvement values, and darker correspond to lower improvement values. As can be seen in the figures, regions of higher estimated potential improvement correspond to transition regions, and larger triangles where the sampling is more sparse. The sparse sampling that created the larger triangles in Fig. 7b in the region of the black car body panel are due to the operational limitations of the Kinect sensor itself in acquiring depth values over black surfaces. Notice that between Fig. 6b,e and Fig. 6c,f the selection process advantageously chooses points at the edges of the cones, around where the blue recycling box is located, and at the edges where the two walls meet each other, as well as at the wall/ground interface. This is reflected by the smaller whiter triangles in these zones within the improvement map, in addition to the sharper detail within the depth map. Fig. 7 demonstrates the same enhancement behavior in the selection process around the region of the gray car door, at the edge of the wooden pallet, the blue recycling box, and the brown vertical cardboard tube.

The estimated depth maps calculated from ordinary Kriging in Fig. 6f and 7f for each scene after 5000 points are selectively chosen using the improvement map, while Fig. 6d and 7d show the estimated depth map after all points are

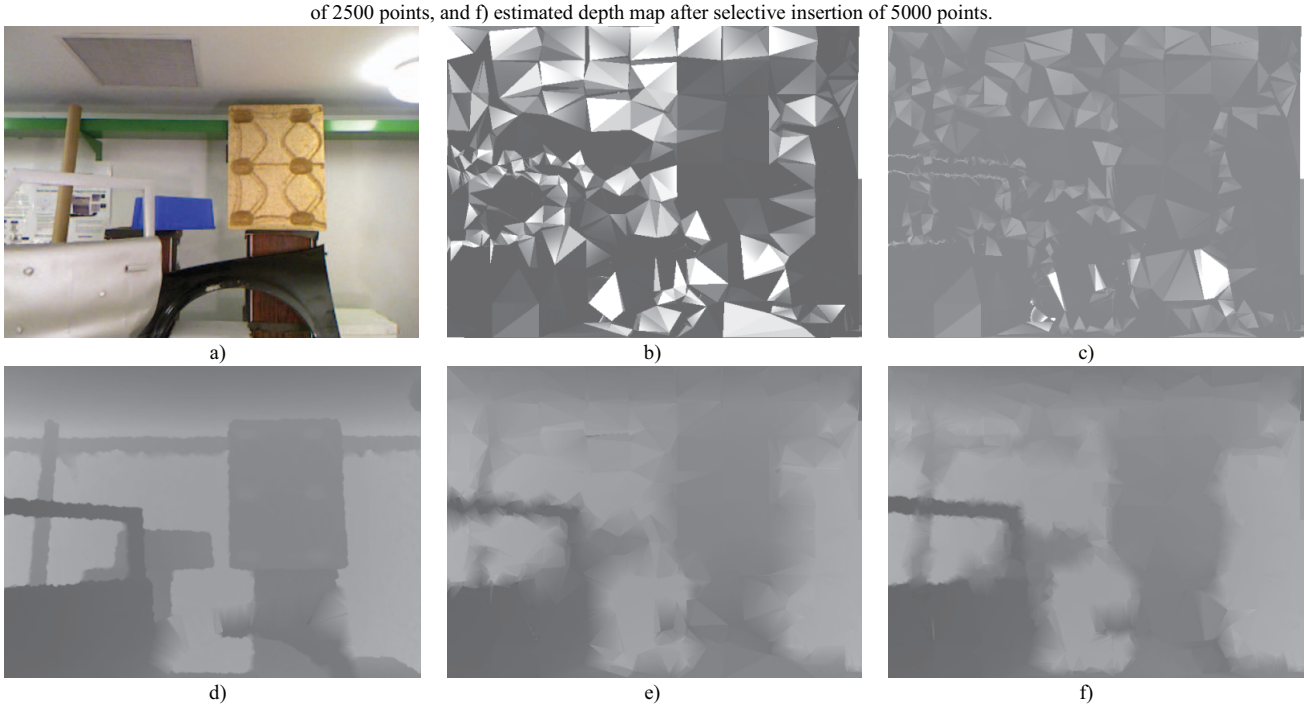
inserted. As can be seen, there is little difference in the completeness of the large scale features in the depth map achieved with only 5000 acquired points compared to full acquisition of 307200 and 277753 points, respectively. This reinforces the argument that ordinary Kriging is a solid foundation for the development of the proposed estimated improvement measure and dynamic selective process detailed in section III, resulting in the formulation of eq. (15).



a) b) c)

d) e) f)

Figure 6. Scene of parking garage with cones: a) RGB image of the scene, b) estimated improvement map after insertion of 2500 points, c) estimated improvement map after insertion of 5000 points, d) estimated depth map after all 307200 points are inserted, e) estimated depth map after selective insertion of 2500 points, and f) estimated depth map after selective insertion of 5000 points.



a) b) c)

d) e) f)

Figure 7. Scene of cluttered lab: a) RGB image of the scene, b) estimated improvement map after insertion of 2500 points, c) estimated improvement map after insertion of 5000 points, d) estimated depth map after all 277753 points are inserted, e) estimated depth map after selective insertion of 2500 points, and f) estimated depth map after selective insertion of 5000 points.

## VI. Conclusion

This research introduces an estimated improvement measure on range data acquisitions that can be used to dynamically select optimal locations where additional depth measurements will provide significant improvement to the knowledge of a scene, as well as determine when the acquisition of more points ceases to provide such significant improvement. The algorithm demonstrates stability, as both the error in the improvement estimation, and the mean estimated improvement level show convergence toward zero as more measurements are progressively added to a 3D representation. This corresponds well to the intuitive notion that as more information about a scene is acquired, the less effect that any one particular point will have on the overall model of the scene. Additionally, ordinary Kriging, which the improvement measure was built upon, is shown to provide a good basis for estimation. The contribution of the dynamic selection of depth measurements acquisition using the improvement map has been demonstrated to effectively reduce the number of points required to adequately represent a scene. The selective sensing algorithm additionally was shown to outperform random sampling in both the mean improvement metric, as well as the improvement error metric. Furthermore, with the reduction of acquired points that is made possible by using the proposed algorithm, slow random access sensors can acquire more scenes, while at the same time reducing bandwidth and storage requirements of the raw data, and reducing the power required for operating active range sensors.

## Acknowledgments

The authors would like to acknowledge the support of Ontario Centres of Excellence (OCE) and Natural Sciences and Engineering Research Council of Canada (NSERC) toward this research, as well as the collaboration of Neptec Design Group.

## References

[1] L. G. Weiss, "Autonomous Robots in the Fog of War," *IEEE Spectrum,* pp. 30-34, 56-57, Aug. 2011.

[2] C. I. Connolly, "The Determination of Next Best Views," *IEEE International Conference on Robotics and Automation*, pp. 432-435, Mar. 1985.

[3] C. I. Connolly, "Cumulative Generation of Octree Models from Range Data," *IEEE International Conference on Robotics and Automation*, pp. 25-32, Mar. 1984.

[4] K. Klein and V. Sequeira, "The View-Cube: An Efficient Method of View Planning for 3D Modelling from Range Data," *IEEE Workshop on Applications of Computer Vision*, pp. 186-191, Palm Spring, CA, Dec. 2000.

[5] M. K. Reed, P. K. Allen and I. Stamos, "3-D Modeling from Range Imagery," in *IEEE International Conference on Recent Advances in 3-D Digital Imaging and Modeling*, pp. 76-83, May 1997.

[6] V. Sequeira, J. G. M. Goncalves and M. I. Ribeiro, "Active View Selection for Efficient 3D Scene Reconstruction," *IEEE Proc. of the Internation Conference on Pattern Recognition*, Vienna, Austria, vol. 1, pp. 815-819, Aug 1996.

[7] J. Maver and R. Bajcsy, "Occlusions as a Guide for Planning the Next View," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 15, no. 5, pp. 417-433, May 1993.

[8] K. Morooka, H. Zha and T. Hasegawa, "Computations on a Spherical View Space for Efficient Planning of Viewpoints in 3-D Object Modeling," *IEEE International Conference on 3-D Digital Imaging and Modeling*, pp. 138-147, Oct. 1999.

[9] D. MacKinnon, V. Aitken and F. Blais, "Adaptive Laser Range Scanning using Quality Metrics," *IEEE Instrumentation and Measurement Technology Conference*, Victoria, BC, pp. 348-353, May 2008.

[10] C. Ho and S. Saripalli, "Where Do You Sample? - An Autonomous Underwater Vehicle Story," *IEEE International Symposium on Robotic and Sensors Environments*, Montreal, QC, pp. 119-124, Sep. 2011.

[11] 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," *International Journal of Intelligent Control and Systems,* vol. 16, no. 2, pp. 79-96, 2011.

[12] A.-M. Cretu, P. Payeur and E.-M. Petriu, "Selective Range Data Acquisition Driven by Neural Gas Networks," *IEEE Transactions on Instrumentation and Measurement,* vol. 58, no. 8, pp. 2634-2642, 2009.

[13] 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.

[14] G. Bohling, "Kriging", University of Kansas, 2005.

[15] P. Goovaerts, *Geostatistics for Natural Resources Evaluation*, New York, New York: Oxford University Press, 1997.

[16] I. Clark, "Practical Geostatistics", 13 July 2001. [Online]. Available: http://www.kriging.com/PG1979/PG1979_pdf.html.

[17] L.E. Dunlap and J.M. Spinazola, "Interpolating Water-Table Altitudes in West-Central Kansas Using Kriging Techniques," United States Government Printing Office, Washington DC, 1984.

[18] Grains Research & Development Corporation, "Making Yield Maps: A guide for the Australian Grains Industry,". [Online]. Available: www.usyd.edu.au/agriculture/acpa/documents/YieldMapping.pdf.

[19] B. Grinstead, A. Koshchan and M.A. Abidi, "Geometry Refinement of 3D Surfaces Using Kriging", in *IEEE Third International Symposium on 3D Data Processing, Visualization, and Transmission*, Chapel Hill, USA, pp. 719-726, June 2006.

[20] D. Nordman, "An Empirical Likelihood Method for Variogram Estimation", in *International Workshop on Spatio-Temporal Modelling*, Pamplona, Spain, 2006.

[21] E. Pardo-Iguzquiza, "MLREML4: A Program for the Inference of the Power Variogram Model by Maximum Likelihood and Restricted Maximum Likelihood", *Computers & Geosciences*, vol. 24, no. 6, pp. 537-543, 1998.

[22] M. de Berg, O. Cheong, M. van Kreveld and M. Overmars, "Delaunay Triangulations: Height Interpolation", in *Computational Geometry: Algorithms and Applications Third Edition*, Springer, 2008, pp. 191-218.

[23] S. Fomel, "Incremental Delaunay Triangulation and Related Problems", 09 Dec 2000. [Online] Available: http://sepwww/stamfprd.edu/oldsep/sergey/sepsergey/fmeiko/paper_html/node9.html. [Accessed 19/06/2012].

[24] Microsoft, "Kinect for Windows Sensor Components and Specifications," Microsoft, [Online]. Available: http://msdn.microsoft.com/en-us/library/jj131033(d=printer).aspx. [Accessed 19 07 2012].

[25] Neptec, "Neptec Laser Metrology System," Neptec, [Online], Available: http://www.neptectechnologies.com/Media/brochures/NTC-Laser-Metrology-System(LMS).pdf. [Accessed 11 02 2013].