Determining Canonical Views of 3D Object Using Minimum Description Length Criterion and Compressive Sensing Method

Ping-Feng Chen, Hamid Krim, *Fellow, IEEE*

North Carolina State University

ABSTRACT

In this paper, we propose using two methods to determine the canonical views of 3D objects: minimum description length (MDL) criterion and compressive sensing method. MDL criterion searches for the description length that achieves the balance between model accuracy and parsimony. It takes the form of the sum of a likelihood and a penalizing term, where the likelihood is in favor of model accuracy such that more views assists the description of an object, while the second term penalizes lengthy description to prevent overfitting of the model. In order to devise the likelihood term, we propose a model to represent a 3D object as the weighted sum of multiple range images, which is used in the second method to determine the canonical views as well.

In compressive sensing method, an intelligent way of parsimoniously sampling an object is presented. We make direct inference from Donoho[1] and Candès'[2] work, and adapt it to our model. Each range image is viewed as a projection, or a sample, of a 3D model, and by using compressive sensing theory, we are able to reconstruct the object with an overwhelming probability by scarcely sensing the object in a random manner. Compressive sensing is different from traditional compressing method in the sense that the former compress things in the sampling stage while the later collects a large number of samples and then compressing mechanism is carried out thereafter. Compressive sensing scheme is particularly useful when the number of sensors are limited or the sampling machinery cost much resource or time.

Keywords: minimum description length, compressive sampling, range image, 3D model reconstruction, canonical views

1. INTRODUCTION

The problem of determining canonical views is not new and has been explored in some literatures.[3–6] Most of them are in the 2D domain that the *canonical images* were searched. This problem can also be classified into either psychological[4,6] or engineering[3,6] domain, where our efforts in this paper oriented from the later point of view. Furthermore, from the engineering point of view, the canonical views can be determined by how they were defined. Some define them as the most representative, or the most featurewise unique views, some define them as the most information abundant views, and others define them as the views that can best help reconstruct the original object. Our definition is in accordance with the last one, but in 3D domain.

The candidates of canonical views are *range images* (i.e. pixel value denoting distance from the scanner to the corresponding point on the object surface) acquired by laser scanner, which captures partial shape of the object from certain viewing angle (ex: Fig. 1). From now on, we use the term partial view and range image interchangeably without confusion being made. If enough range images of one object are collected, reconstruction of the object's outer surface is possible. Traditionally, to reconstruct a full 3D model, several range images were acquired with a fair amount of overlaps in viewing angles, and then some sort of fusion technique was carried out to stitch or merge them into one entity.[7–9] It is known as a fact that when the laser ray illuminates the object surface with a *grazing angle* the reading is inaccurate,[9] which happens most prominently at the boundaries of the object. Therefore when acquiring data the overlap in viewing angle is necessary for one partial view's reliable readings to compensate for the erroneous readings of the other. However in most works the amount of overlap, or the canonical views, is not studied. The amount of overlap is taken at will, or arbitrary, and different techniques were carried out to sum the overlapped parts.[8,9] The focus of this paper is therefore to find out how much overlap is enough, or what the canonical views are.

Computational Imaging VI, edited by Charles A. Bouman, Eric L. Miller, Ilya Pollak,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6814, 68140R, © 2008 SPIE-IS&T · 0277-786X/08/$18

In this paper, we propose two novel methods to determine the canonical views of an object in 3D domain, Minimum Description Length (MDL) criterion[10] and Compressive Sensing method,[1,2,11] for reconstruction purpose. Through the development of these methods, a new model to represent a closed surface by the weighted sum of multiple range images is proposed as well, which is motivated by Hoppe[12] and Whitaker's[13] signed distance function. The paper is organized as follows: in section 2 we illustrate the model of representing the sum of range images, in section 3 and 4 we illustrate determining canonical views using MDL criterion and Compressive Sensing method, together with their simulation results, and at last section 5 is the conclusion.

2. A MODEL OF SUM OF MULTIPLE RANGE IMAGES

We propose a new model to represent the sum of range images from different viewing angles. A common assumption made in the fusion of multiple range image literature is that the corresponding position of each view is known,[9,13] such that no pre-registration process is needed; even in the registration of multiple range images literature, the "roughly aligned" assumption is also made.[7,8] Therefore we adopt the same assumption that the exact position of each range image is known. We further assume that we have a full 3D model of the object of interest, and the position and viewing angles of the scanner is also known, such that the exact 3D position of each partial view can be located. Practically, the viewing angles can be easily determined when scanning small scale object, and can be calculated by the corresponding positions acquired by GPS (Global Positioning System) when doing large scale scan, for example the airborne scan on the battle field.

A full 3D object is represented by triangle meshes which consist of vertices and edges. Each range image can also be triangle meshed similarly as follows: a range image can be imagined as the result of throwing a silk on top of the object such that self-occlusions are not captured, and triangulation mesh can be easily done using Delaunay triangulation in 2D, since each range image is essentially a 2D image except for that the pixel values denoting distances. The meshes of range image and 3D model are constructed differently though, i.e. the vertices of them do not correspond to the same points. Fig. 1 shows a simulated range image overlaid on top of the 3D model. Notice that at the boundary of the object, the range values are more erroneous than the central readings as we described in previous section.

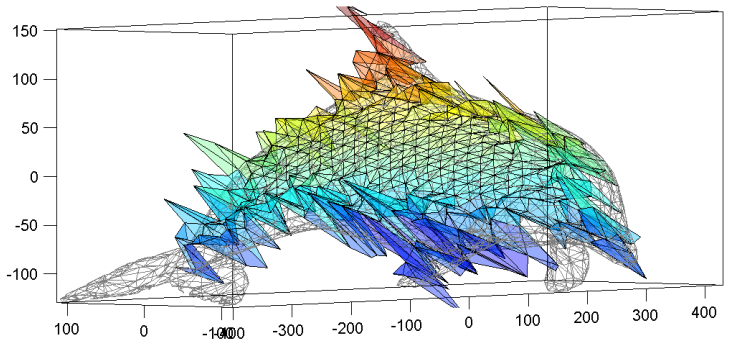


Figure 1. A simulated range image overlaid on top of the 3D model, where at the boundaries of the object, the range readings are more inaccurate than the central ones.

We use matrix $\mathbf{M}_{l\times 3}, \mathbf{M}_{(i,\cdot)} = (v_1, v_2, v_3) \in \mathbb{R}^3$, whose each row denotes the vertex of the triangle meshes and l the number of vertices, to represent the full 3D surface. In order for the vertices $\mathbf{M}$ to be representative enough of the object, the meshes of the surface has to be uniform and dense enough. Then we would like to approximate $\mathbf{M}$ by a weighted sum of different partial views, and wish that as the number of views k increases, the approximation be closer to $\mathbf{M}$. Motivated by Hoppe[12] and Whitaker,[13] our model is

$$\mathbf{M} \approx E\left(\sum_{i=1}^{k} a_i(\mathbf{X}) V_i(\mathbf{X})\right), \tag{1}$$

where matrix $\mathbf{X}_{P\times 3}, \mathbf{X}_{(i,\cdot)} = \mathbf{x}_i = (x, y, z) \in \mathbb{R}^3$, denotes grid points in a 3D bounding box, P the size of the bounding box, and the rest of a_i, V_i, and the edge operator $E(\cdot)$ will be defined as follows. V_i is the signed distance field of a partial view as defined in Ref. 13, which we reveal it again here. For each grid point $\mathbf{x}_i$, we define $R(\mathbf{x}_i) \in \mathbb{R}$ as the signed distance between the spatial point $\mathbf{x}_i$ and the range reading $\vec{r}(\mathbf{x}_i)$ (hereafter we drop the subscript i for brevity)

$$R(\mathbf{x}) = (\mathbf{x} - \vec{r}(\mathbf{x})) \cdot \vec{n}(\mathbf{x}), \tag{2}$$

where $\vec{r}(\mathbf{x}) \in \mathbb{R}^3$ is the range reading acquired along the line of sight from the laser scanner to $\mathbf{x}$, and $\vec{n}(\mathbf{x})$ is the unit vector along the line of sight from the scanner (Fig. 2). Then the signed distance function $V(\mathbf{x})$ can be defined as

$$V(\mathbf{x}) = R(\mathbf{x})W(R(\mathbf{x})), \tag{3}$$

where $W(t)$ is a window function being 1 if $t \leq \varepsilon$ and 0 otherwise, which prevents the distance function from affecting the other side of the surface.

A similar definition for the 3D model is

$$D(\mathbf{x}) = S(\mathbf{x})W(S(\mathbf{x})), \tag{4}$$

where $S(\mathbf{x}) = (\mathbf{x} - \vec{s}(\mathbf{x})) \cdot \vec{n}(\mathbf{x})$ and $\vec{s}(\mathbf{x}) \in \mathbb{R}^3$ is the exact signed distance from the scanner to the 3D model surface along the line of sight passing through $\mathbf{x}$ (i.e. the range reading $\vec{r}(\mathbf{x})$ in Eq. (2) may be inaccurate due to any artifact).

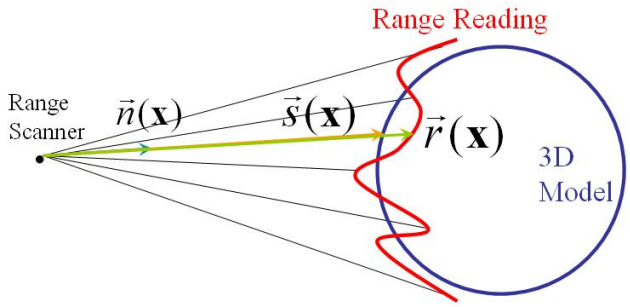


Figure 2. Figure reproduced from Ref. 13. Illustration of range reading $\vec{r}(\mathbf{x})$, exactly surface range $\vec{s}(\mathbf{x})$, and unit vector $\vec{n}(\mathbf{x})$ along one line of sight.

The weight function $a(\mathbf{x})$ can therefore be defined as

$$a(\mathbf{x}) = e^{-|V(\mathbf{x})-D(\mathbf{x})|} = e^{-|(\vec{s}(\mathbf{x})-\vec{r}(\mathbf{x}))\cdot\vec{n}(\mathbf{x}))|}. \tag{5}$$

In plain words, we assign a heavy weight to the signed distance V when the range reading is faithful to the 3D model and a less weight when the reading deviates from the ground truth. Notice that when summing all the weighted signed distances that can be acquired, the zero-crossing (zero *level set*) corresponds to the true object surface, if the weighted distances contributed by overlapped views whose erroneous readings at the boundaries cancel out, which is one important assumption for this model. Therefore the approximated 3D surface is the set denoted by $\mathbf{Z}$

$$\mathbf{Z} = \left\{ \mathbf{x} \in \mathbb{R}^3 \middle| \sum_i a_i(\mathbf{x})V_i(\mathbf{x}) = 0 \right\}. \tag{6}$$

Notice that if the cover of sum of the weighted distant field does not sweep the whole 3D space within the inside of the object, looking for the zero-crossing would not make sense. Therefore asking the union of viewing angles to cover the whole space is one requirement of our model.

The edge operator $E(\cdot)$ has then to be defined as some points on the approximated surface corresponding to the vertices of 3D model. Mathematically, we defined it as the intersection between the outward normals of the 3D model vertices $\mathbf{M}$ and the zero level set $\mathbf{Z}$,

$$E\left(\sum_{i=1}^{k} a_i(\mathbf{X})V_i(\mathbf{X})\right) = \left\{\mathbf{X}'_{l\times 3}\middle|\left(\mathbf{X}'_{(i,\cdot)} - \mathbf{X}_{(i,\cdot)}\right)\backslash\backslash\vec{N}_i, \mathbf{X}'_{(i,\cdot)} \in \mathbf{Z}, \forall i = 1,2,\ldots,n\right\}, \tag{7}$$

where $\vec{N}_i \in \mathbb{R}^3, i = 1,2,\ldots,l$, is the outward normal to each vertex of 3D model $\mathbf{M}$, and l the number of vertices of 3D model.

Curless *et al.*[9] has a very similar model as ours in the sense that they reconstructed the object as the weighted sum of signed distances *without* knowing the 3D model. However, our *weight* is defined differently as theirs and the purpose of the our model, instead of doing reconstruction, is to find out the canonical views. Our model would not make sense if we do reconstruction since the full 3D model is already known as a priori, but it is solely to assist in making the decision, of the canonical views. In the following two sections we present two methods to determine the canonical views.

3. DETERMINE THE CANONICAL VIEWS BY MINIMUM DESCRIPTION LENGTH CRITERION

3.1 Theory

MDL (Minimum Description Length) criterion, proposed by Rissanen,[10] was at first used in coding theory. The criterion essentially looks for the code length achieving the balance between model accuracy and overfitting. One does not want to use too long a code to describe the data because this code may be overfitted, nor too short the description length to sacrifice the accuracy. In other words, one can design a code perfectly fit to a given set of data with certain probability distribution, however, whenever there is new incoming data, this code could be very wrong. The quest for the compromise between parsimony and goodness-of-fit has long been studied and applied on many fields.[14,15] For our case, we are basically looking for a proper number (code length) of views that can best describe the object (3D model), and find out which views they are.

Rissannen's MDL criterion essentially takes a form of penalized likelihood, and the penalty is the cost to encode data with longer length. The associated minimum length criterion for our case would be to first identify that out of N total partial views, there are $2k$ free parameters corresponding to k canonical views together with which viewing angles they are. The criterion to be minimized is therefore

$$L(\mathbf{k}) = -\log H(\mathbf{M},\mathbf{k}) + \frac{1}{2}(2k)\log(N), \tag{8}$$

where $H(\mathbf{M},\mathbf{k})$ denotes the likelihood, representing how close the sum of partial views approximates the 3D model, and $\mathbf{k} \subset \{1,\ldots,N\}$, k the cardinality of $\mathbf{k}$, is the set of views to be searched for. The likelihood $H(\mathbf{M},\mathbf{k})$ is defined as

$$H(\mathbf{M},\mathbf{k}) = \exp\left(\frac{-\left\|\mathbf{M} - E\left(\sum_{i\in\mathbf{k}} a_i(\mathbf{X})V_i(\mathbf{X})\right)\right\|^2}{2\sigma^2}\right), \tag{9}$$

where, as most practical cases, we assume our approximation model, the weighted sum of partial views, obeys a Gaussian distribution; σ^2 denotes the variance, and the norm is a Euclidean (l_2) norm. Rewriting the final minimum length criterion, we have

$$\tilde{\mathbf{k}} = arg\min_{\mathbf{k}\subset\{1,\ldots,N\}} L(\mathbf{k}) = arg\min_{\mathbf{k}\subset\{1,\ldots,N\}} \left\{\frac{\left\|\mathbf{M} - E\left(\sum_{i\in\mathbf{k}} a_i(\mathbf{X})V_i(\mathbf{X})\right)\right\|^2}{2\sigma^2} + \frac{1}{2}(2k)\log(N)\right\}. \tag{10}$$

By minimizing $L(\mathbf{k})$ with respect to $\mathbf{k}$ we should be able to find the best k views, namely our goal of the canonical views, that parsimoniously describe the object. Notice that if we only minimize the first term $H(\mathbf{M},\mathbf{k})$, k would tend to all partial views N, which is the overfitting situation as we described earlier. Therefore adding the penalizing term would give us a proper subset of canonical views from N total views.

3.2 Experimental Results

We conduct the following simulation: Fixing the object at the origin of a spherical coordinate system and take range measurements at $\{(\theta,\phi)|\theta \in (0,30,\ldots,330),\ \phi \in (15,30,\ldots,180)$ degrees $\}$, where θ and ϕ represent azimuth and (90-elevation) in degree. Moreover we have the assumption that the scanner sits at infinitely far so the scanning rays are all of parallel lines. The reason for the interval of θ and ϕ is a compromised result that: 1.) it guarantees the viewing angles are many enough such that the union of them cover the whole 3D space, and 2.) the overlap of some views provides enough reliable readings to compensate for the erroneous measurements of others, while 3.) the number of views is not too many such that the minimization problem is feasible in computation. We assert that this set of viewing angles is overcomplete for reconstruction and out of which we would like to use MDL criterion to decide a subset of views, i.e. the canonical views.

We used a 3D dolphin model for simulation, which is triangle meshed uniformly with 1049 vertices. Laser range measurement simulation is carried out, which is later polluted by Gaussian noise under our assumption that the closer to the boundaries, the readings being more inaccurate. To achieve this, two steps are carried out: 1.) the distance function from the boundary of object is calculated, and 2.) the noise is distributed accordingly. Specifically, first let $\omega \in \mathbb{R}^2$ denote the domain inside the boundaries of the object within one range image, and $\partial\omega$ the boundaries of the object, and the distance function D is defined as

$$D(\mathbf{x}) = \begin{cases} |d(\mathbf{x},\partial\omega) - \max_{\mathbf{x}\in\omega} d(\mathbf{x},\partial\omega)|, & \mathbf{x} = (x,y) \in \omega \\ 0, & \text{otherwise,} \end{cases} \tag{11}$$

where $d(\mathbf{x},\partial\omega)$ is the Euclidean distance between $\mathbf{x}$ and $\partial\omega$.

Secondly Gaussian noise is added with variance σ^2_{noise} according to the distance function

$$\sigma^2_{noise}(D(\mathbf{x})) = D(\mathbf{x})^{\epsilon}, \tag{12}$$

where $\epsilon \in \mathbb{R}$ is chosen to adapt to different models because the scale difference, but in simple words, the closer to the boundaries the more heavily polluted by noise. For our case of dolphin, ϵ is set as 3.5.

With the above settings, we solve the minimization problem of Eq. (10) with respect to k. Basically it is computational expensive because an exhausted search is carried out for both the number of canonical views k and their corresponding angles, which is of $\mathcal{O}(\sum_k^N \mathcal{C}_k^N)$ complexity. The result of this experiment shows that 11 canonical views are needed to minimize MDL criterion, and these views are taken at angles (θ,ϕ) $=\{(180,75)$, $(330,135)$, $(60,15)$, $(300,135)$, $(330,165)$, $(0,165)$, $(30,45)$, $(90,75)$, $(180,135)$, $(240,105)$, $(270,105)\}$ in degree. The corresponding reconstructed 3D surface is shown in Fig. 3. The reason of this coarse resolution is due to computation concern. If finer resolution is adopted, namely larger bounding box size of the signed distance field V_k, more CPU is used. Therefore a smaller size/lower resolution is used as a compromise.

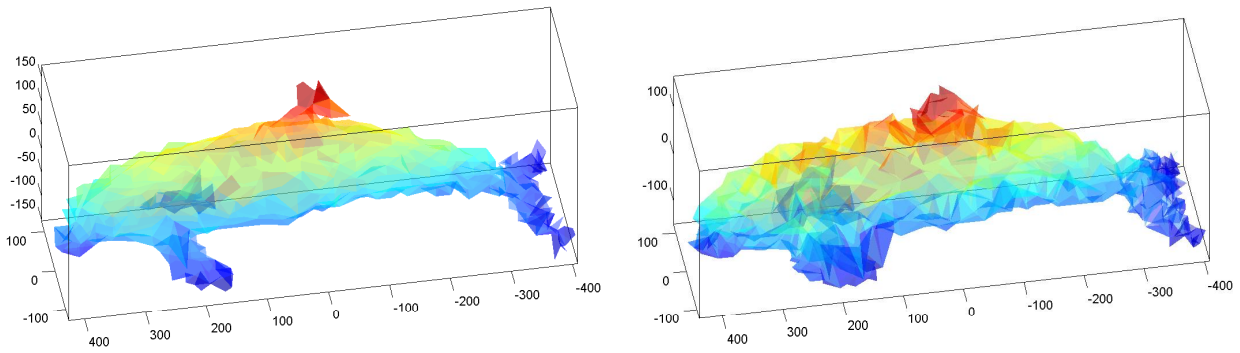
The viewing angles of the canonical views obtained pretty much spread out uniformly in the 3D space, and they have some overlap to each other so that the reliable reading in one view compensates for another's erroneous readings, which is in accordance with our intuitive.

4. CANONICAL VIEW DETERMINATION AND 3D MODEL RECONSTRUCTION BY COMPRESSIVE SENSING METHOD

4.1 Theory

Even though in previous section, minimum description length criterion provides an elegant method to solve the canonical view problem, the computation complexity is very expensive because an exhausted search is carried out in the minimization process. There actually exists a method around this heavy cost process, the compressive sensing method. In other words, compressive sensing method need not acquire all the possible partial views at first and then out of which select a subset of canonical views, but it compressively samples data in the first stage.

There has been a growth of study in compressive sensing recently,[1,2,11,16] and the basic idea says that given a data with some sort of sparsity property, by sampling an incomplete measurements in a random fashion, there is a overwhelming probability the original data can be fully reconstructed. Donoho[1] and Candès[2] are the most



(a) (b)

Figure 3. (a.) 3D model, (b.) Reconstructed surface from the 11 canonical views obtained from minimum description length criterion (the poor resolution is due to computational complexity concern).

renowned two who published their works on this topic separately. Before we apply this method to determining the canonical views and reconstructing a 3D model, let us review the basic ideas.

For illustration and comparison purpose, let's look at one simple example given in Ref. 11. Suppose f is the given data (ex: an image), and (Φ, Ψ) is a pair of orthobases of $\mathbb{R}^n$ (the restriction of their being orthobases is not essential but for the convenience of illustration), where Φ denotes the sensing basis and Ψ the representing basis, with their corresponding elements φ_i's and ψ_i's. The sensing basis can be sinusoids if we sample the data by Fourier transform, or wavelets if we use wavelet transform. For most natural data, the projection of the data on the sensing basis is often sparse, in the sense that the significant coefficients of the projection (ex: Fourier coefficients) is relatively few. Traditional compressing techniques acquire all the measurements, and then do things such as simple thresholding to store those significant ones and throw away other non-informative samples. However they may be impractical when the sensing mechanism is very expensive or the sampling time is long. Practically, we can imagine that taking one airborne laser scan on the battle field is both time and resources consuming, and it is not possible to have as many scans as we wish. Therefore it comes to compressive sensing that skips the collecting data step, but compress things in the sensing stage.

The theory says that when the sensing and representing basis are of *low coherence* pairs,[11] i.e. small $\mu(\Phi, \Psi)(= \sqrt{n} \cdot \max_{1 \le k,j \le n} |\langle \varphi_k, \psi_j \rangle|)$, fully reconstruction of the original data is possible for incomplete sampling. As shown in Ref. 11, random matrices are mostly incoherent with any fixed basis Ψ, and therefore the whole compressive sensing method can be cast in the following way. Suppose the sensing basis Φ has n elements and let the coefficients y_k be the projections of the signal on the basis. By uniformly sampling k coefficients at random

$$y_k = \langle f, \varphi_k \rangle, k \in M, \tag{13}$$

where M is a subset of $\{1, \ldots, n\}$ with cardinality $m < n$, the signal f can be reconstructed with an overwhelming probability by solving the following minimization problem

$$\min_{\tilde{x} \in \mathbb{R}^n} \| \tilde{x} \|_{l_1} \quad \text{subject to} \quad y_k = \langle \psi_k, \Phi \tilde{x} \rangle, \forall k \in M, \tag{14}$$

and the reconstruction of f will be $f^\star = \Psi x^\star$, where $x^\star$ is the solution to the above minimization problem. In words, the problem finds the minimal l_1 norm coefficients $x^\star$ that conform to those few random observations y_k, and then the data f can be reconstructed from the coefficients. Compressive sensing provides us an useful tool to reveal the truth of some signal by merely randomly, and scarcely sensing it, without searching the whole space in which the signal resides.

To make an analogy and apply this method to our case, we make the following observations: First, each partial view scan can be seen as a sample from certain viewing angle (θ, φ) (suppose the object is sitting at the origin of a spherical coordinate and the scanner is at angle (θ, φ) at infinity), i.e. each range image is a projection of the 3D model, $V_i = \langle f, (\theta, \varphi) \rangle$, where f denotes the 3D model. The difference is that, instead of interacting

with a basis, the projection is only a subset (part of the surface, i.e. partial view) of the original data, with the properties that the value being more inaccurate at the boundaries of the object. Secondly, remember one requirement of compressive sensing is that the signal has to possess some sparsity property. Suppose we make the most stringent case that only the central line of the scanning rays captures the correct reading, in other words, the scanner behaves as a Dirac delta function, then infinite scans are needed to reconstruct the object. However this is not the case since each partial view has a fair amount of reliable readings. Therefore the sparsity property comes in the sense that we do not need to sample infinitely many scans, even though at the boundaries of each partial view the reading may be inaccurate, each scan has a fair amount of overlap with others and out of which most readings are reliable. Therefore, it is *the number of canonical views* that is sparse. Following a similar fashion, one can thus construct the compressive sensing method for scanning 3D object by first uniformly sampling the object in (θ, φ) domain at random

$$V_k = \langle f, (\theta_k, \varphi_k) \rangle, k \in M, \tag{15}$$

where $\{(\theta_i, \varphi_i), i = 1, \ldots, n\}$ is an overcomplete set of viewing angles, $M \subset \{1, \ldots, n\}$ and $|M| = m < n$; V_k need not be the distance function as we defined in section 2, but any feature, say vertices of triangle meshes. However for convenience of illustration we admit it to be the signed distance function as defined in Eq. (3).

Then secondly solve the minimization problem

$$\min_{C \in \mathbb{Z}} C \quad \text{subject to} \quad V_k = \left\langle (\theta_k, \varphi_k), E\left(\sum_{i=I_1}^{I_C} a_i V_i' \right) \right\rangle \forall k \in M, \tag{16}$$

where $V_i' = \langle f, (\theta_i, \varphi_i) \rangle, i \in \{1, \ldots, n\}$, $\{I_1, I_2, \ldots, I_C\} \subset \{1, \ldots, n\}$, and the sum of weighted partial view $\sum_{i=1}^{c} a_i V_i'$ and the edge operator $E(\cdot)$ are as defined in section 2, and the signal f can be reconstructed with an overwhelming probability.

However there are differences and problems for this setting compared with the illustrative example we just gave. The difference is that we observe that the number of partial views to be summed in Eq. (16), instead of n element of the basis in Fourier transform, is only C of partial views (which is the number of canonical views to be solved). If we use n partial views in the summation, the projection at angle (θ_k, φ_k) will no matter what conform to V_k. Since n views is an overcomplete set of views, and the weighted sum after the edge operator $E\left(\sum_{i=1}^{n} a_i V_i'\right)$ will be the original signal f. Furthermore, if doing so, the number of canonical view C then plays no role in the minimization problem. Setting this number to C is reasonable in the sense that we would like to reconstruct the signal from as few number of partial views as possible, and for its projection at the sampling angles conforming with those views.

However there are still two problems to be noticed. First, we do not know out of n views, which C partial views should we choose in Eq. (16). This minimization problem is with respect to the number C but there are $\mathcal{C}_C^n$ choices and we do not know which one to pick. Thus we make an important assumption, that is: the canonical views, for reconstruction purpose, are not a fixed set of views, but rather their relative positions are fixed. What this assumption helps us is that we do not need to locate the specific C views, but only need to find out C views with a fixed relative positions. Actually this assumption can be relaxed more by saying that the canonical views have fixed relative positions with some tolerance ε. Mathematically speaking, the *canonical viewing angles* are defined as the set $\mathbf{G}$

$$\mathbf{G} = \left\{ (\theta_k, \varphi_k), 1 \leq k \leq C \middle| (\theta_i - \theta_j, \varphi_i - \varphi_j) = (\Delta\theta_{i,j} \pm \varepsilon, \Delta\varphi_{i,j} \pm \varepsilon), 1 \leq i, j \leq C \right\}, \tag{17}$$

where $\Delta\theta_{i,j}, 1 \leq i, j \leq C$ and $\Delta\varphi_{i,j}, 1 \leq i, j \leq C$, are sets of fixed distances between canonical viewing angles and $\varepsilon \in \mathbb{R}$ is the tolerance.

With this assumption, the possibility of hitting the canonical views greatly increases. However, we still do not know which views to choose. Therefore we need to twist the setting a little more to come up with

$$\min_{C \in \mathbb{Z}} C \quad \text{subject to} \quad V_k = \left\langle (\theta_k, \varphi_k), E\left(\sum_{i=1}^{C} a_i V_i' \middle| V_i' = \langle f, (\theta_j, \varphi_j) \rangle, j \in U \right) \right\rangle \forall k \in M, \tag{18}$$

where $U \subset \{1, \ldots, n\}$, $|U| = C$, is a subset of uniformly random samples from the overcomplete views. The restriction of the uniformly random sampling is where the twist is. It again comes from the compressive sensing idea that by uniformly sampling the object at random, there is a great chance the whole object can be reconstructed. The floating canonical views assumption increase the chance of hitting the right views. Notice though that the solution $C^\star$ to this minimization problem is not the exact number of canonical views, but lager, while we assert in the order of the true canonical number by the result of compressive sensing theory.

The second problem is that it is not reasonable for the projection to be equal to the whole partial view V_k, since remember at the boundaries, the readings are more inaccurate due to the grazing incidence angles. Therefore we should instead require the projection to be equal to the reliable part, say out of some distance of the object boundaries, of the partial view. The final setting of the minimization problem is then as follows

$$\min_{C \in \mathbb{Z}} C \quad \text{subject to} \quad \tilde{V_k} = \left\langle (\theta_k, \varphi_k), E\left(\sum_{i=1}^{C} a_i V_i' \middle| V_i' = \langle f, (\theta_j, \varphi_j) \rangle, j \in U \right) \right\rangle \forall k \in M, \tag{19}$$

where $U \subset \{1, \ldots, n\}$, $|U| = C$, is the set of uniformly random samples from the overcomplete views, and $\tilde{V_k} = \{V_k(\mathbf{x}) | d(\mathbf{x}, \partial\omega) > \xi\}$, where $d(\mathbf{x}, \partial\omega)$ is the distance function as defined in section 3.2, and $\xi \in \mathbb{R}$ is a predefined threshold.

The whole compressive sensing scheme is as follows. Uniformly sampling the signal f (3D object of interst) in (θ, φ) domain m times at random as shown in Eq. (15), and then the signal f can be reconstructed using $C^\star$ uniformly random partial views with an overwhelming probability, where $C^\star$ is the solution of Eq. (19).

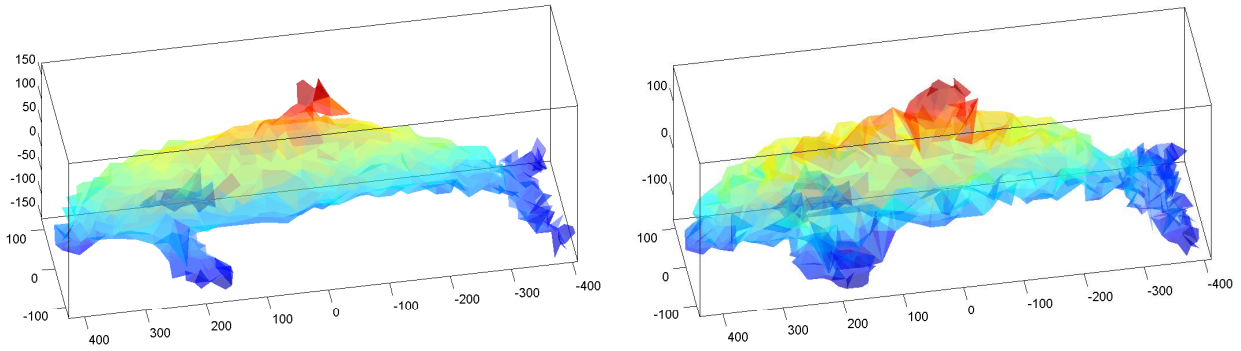
The minimization part is not easy to solve, and the even more difficult problem lies on the number of times m we have to do random sampling. In theory, as long as we randomly sample in the order of $S \log(n)$ times and solve the minimization problem, the signal can be reconstructed,[11] where S is the sparsity number, namely the number of significant coefficients. However the number corresponds to S in our case is $C^\star$ (actually should be $\tilde{C}^\star$; remember we claimed that the solution $C^\star$ is greater than the real number of canonical views $\tilde{C}^\star$ but in the same order), which is the number to be solved in the minimization problem. Theoretically this is an extremely complex problem that we have to go back and forth to exhaustively try different m and solve for $C^\star$ to see if m conforms to the order of $\tilde{C}^\star \log(n)$. In the following section we illustrate a way to go around and show the reconstructed object using $C^\star$ canonical views.

4.2 Experimental Result

The minimization problem of Eq. (19) is more expensive than the one of MDL criterion in section 3.2 because of the calculation of weighted sum of signed distances, the cost to locate the zero crossing of the fusion of views, the projection from angle (θ_k, φ_k), and an exhausted search of the number of canonical views. What is causing more headache is that we do not know the number of times m we should randomly sample the signal f in Eq. (15). Therefore we use a trick to go around. Instead of minimizing Eq. (19) directly, we use the knowledge of the result obtained from previous section to assist us. The number of canonical views has already been found to be 11 in section 3.2 for our particular case of the dolphin model, therefore we say that $C^\star$ is assumed being solved in Eq. (19). The number of times we have sample the signal f is then in the order of $C^\star \log(n)$ (actually $\tilde{C}^\star \log(n)$, but presumably in the same order) as we described at the end of previous subsection. Therefore we let $m = C^\star \log(n)$ and use this number to randomly sample the signal, and then we go back to solve the minimization problem Eq. (19).

Therefore the whole procedure for our dolphin model is: 1.) uniformly sample the object f in the (θ, φ) domain at random $11 \cdot \log(72)$ times ($72 = n =$ the number of elements in the overcomplete set $\{(\theta_i, \varphi_i), i = 1, \ldots, n\}$ as illustrated in section 3.2), 2.) solve for the number of canonical views $C^\star$ in Eq. (19), and 3.) the object can be reconstructed by the fusion of $C^\star$ uniformly random samples.

The result of $C^\star$ is in the range $30 - 40$ in our simulation, and the reconstructed object is the fusion using 35 random partial views as shown in Fig. 4. As can be seen the reconstructed model is faithful to the model in this resolution. Again the poor resolution is due to the concern of computational complexity.



(a) (b)

Figure 4. (a.) 3D model, (b.) Reconstructed surface from 35 uniform random projections obtained using compressive sensing method (the poor resolution is due to computational complexity concern).

5. CONCLUSION

In this paper we proposed two methods to determine the canonical views of a 3D object. The MDL criterion provides an elegant formulation through minimizing which we obtain the canonical views; compressive sensing provides us an useful scheme to parsimoniously scan the object in a random manor, with sampling times of a number in the order of $C\log(n)$, where C is the number of canonical views and n is the total number of elements of the overcomplete representation function. In both analysis, a full 3D model is assumed known and each partial view is assumed to be already in the corresponding position. Furthermore, through the development of these methods, a novel model to represent a 3D object by the weighted sum of range images was proposed as well. These methods can be particularly useful when determining which partial views to sample for reconstruction in the situations that a single sample cost much resource and time.

REFERENCES

1. D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on* **52**(4), pp. 1289–1306, April 2006.
2. E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.* , 2006.
3. P. Hall and M. Owen, "Simple canonical views," *The British Machine Vision Conf* , pp. 835–848, 2005.
4. V. Blanz, M. J. Tarr, and H. Bülthoff, "What object attributes determine canonical views," *Perception* **28**, pp. 575–599, 1999.
5. T. Denton, M. Demirci, J. Abrahamson, A. Shokoufandeh, and S. Dickinson, "Selecting canonical views for view-based 3-d object recognition," *Proc. International Conference on Pattern Recognition* **2**, pp. 273–276, 2004.
6. F. Cutzu and S. Edelman, "Canonical views in object representation and recognition," *Vision Research* , 1994.
7. Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," *Image and Vision Computing* **10**(3), pp. 145–155, 1992.
8. C. Dorai, G. Wang, A. Jain, and C. Mercer, "Registration and integration of multiple object views for 3d model construction," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**, pp. 83–89, 1998.
9. B. Curless and M. Levoy, "A volumetric method for building complex models from range images," *Proc. SIGGRAPH* , 1996.
10. J. Rissanen, "Modeling by shortest data description," *Automatica* **14**, pp. 465–471, 1978.
11. E. J. Candès and M. Wakin, "People hearing without listening: an introduction to compressive sampling," *To appear in IEEE Signal Processing Magazine* , 2007.
12. H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, "Surface reconstruction from unorganized points," *Computer Graphics (SIGGRAPH '92 Proceedings)* **26**, pp. 71–78, 1992.

13. R. Whitaker, "A level-set approach to 3d reconstruction from range data," *Int'l J. Computer Vision* **29**, pp. 203–231, 1998.
14. H. Krim and J. Pesquet, *On the Statistics of Best Bases Criteria*, pp. 193–207. Springer Verlag, New York, 2004.
15. I. Krim, H.; Schick, "Minimax description length for signal denoising and optimized representation," *Information Theory, IEEE Transactions on* **45**(3), pp. 898–908, Apr 1999.
16. E. Candes and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," 2004.