Simultaneous registration of multi-view range images with adaptive kernel density estimation

By Steven McDonagh and Robert B. Fisher

The University of Edinburgh, UK
s.g.mcdonagh@sms.ed.ac.uk rbf@inf.ed.ac.uk

Abstract

3D surface registration can be considered one of the crucial stages of reconstructing 3D objects from depth sensor data. Aligning *pairs* of surfaces is a well studied problem that has resulted in fast and usually reliable algorithms addressing the task. The generalised problem of globally aligning *multiple* surfaces is a more complex task that has received less attention yet remains a fundamental part of extracting a model from multiple 3D surface measurements for most useful applications.

In this paper, we propose a novel approach for the global registration of depth sensor data, represented by multiple dense point clouds. Point correspondences between scans and view order are unknown. Given many partial views, we estimate a kernel-based density function of the point data to determine an accurate approximation of the sampled surface. We define an energy function which implicitly considers all viewpoints simultaneously. We use this density to guide an energy minimisation in the transform space, aligning all partial views robustly. We evaluate this strategy quantitatively on synthetic and range sensor data where we find that we have competitive registration accuracy through comprehensive experiments that compare our approach with existing frameworks for this task.

1. Introduction

Surface modelling from range data includes the important step of registering many partial range scans into a common coordinate system. The task is required in several areas of computer vision, computer graphics and reverse engineering. Point data sets are routinely generated by optical and photometric range finders and have become increasingly popular in applications such as autonomous navigation (Whitty et al. (2010)), e-Heritage (Ikeuchi et al. (2007)) and medical imaging (Clements et al. (2008)). Typical sensors can only measure the visible surface of the target, and therefore only provide a partial view of the object due to (self)-occlusions, blind areas or otherwise missing data. Generating high quality geometric representations from real-world objects requires the fusion of such partial views into a common coordinate frame by estimating the transforms between the data sets. This is the multi-view registration problem.

An initial coarse registration can often be found directly from the sensor scanning system or interactively from the user. The registration can then be refined by accurately aligning the overlapping parts of the scans. This refined registration task is typically subdivided into the correspondence and alignment sub-problems. The correspondence problem is defined as: given a point in one scan, determine the points in the other scans

that are near to the same physical point on the object surface. An exact correspondence may not actually be sampled due to sample quantization. The alignment problem involves estimating the motion parameters that bring one scan into the best possible alignment with the others. Fixing either of these objectives renders the other trivial to solve.

A naive method for accurately registering many views is by sequential registration. Unregistered views are registered, one at a time, to a base scan or another previously registered viewpoint. The main problem with this approach is that pairwise registration errors can begin to accumulate and propagate and thus sequential approaches are not optimal. Moreover the view sequence order must be known or manually specified in advance. We consider the problem of simultaneous global registration, where point correspondences and view order are unknown and the aim is to align all views simultaneously by distributing the registration errors evenly between overlapping viewpoints. Previous approaches for tackling this problem are surveyed in Section 2. Given many partial views, we estimate a kernel-based density function of the point data to determine an accurate dense approximation of the sampled surface. We use this density to guide an energy minimization in the transform space, aligning all partial views robustly. The details of our method are found in Sections 3 and 4. We evaluate this strategy quantitatively on synthetic and range sensor data where we find that we have competitive registration accuracy while improving convergence behaviour over existing frameworks for this task.

Our working hypothesis is: A *simultaneous* registration method will improve registration accuracy over *sequential* approaches by distributing errors evenly between overlapping viewpoints. We can robustly approximate an object surface $\mathcal{S}$, represented by coarsely misaligned partial views, with kernel density estimation and use this to iteratively guide simultaneous registration, improving accuracy and robustness over existing techniques.

2. Related Work

When only pairwise registration between two scans is required, the problem can be considered well studied in the literature. The Iterative Closest Point (ICP) algorithm proposed by Besl and McKay (Besl & McKay (1992)) and subsequent modifications make up the predominant share of proposed methods. ICP-based techniques start from a coarsely aligned pose and iteratively revise a transformation (typically composed of rotation and translation) to minimize a distance function between pairs of neighbouring points in the two point clouds. Point correspondences are typically found by considering the point in the other view currently of minimum distance. Since this family of techniques is based on local iterative decisions they are generally susceptible to local minima (for example, when poor initial coarse alignment is provided).

Variants of ICP tend to make improvements on either registration speed or robustness and accuracy. The latter is typically achieved by modifying the point pair matching strategy. These works include rejecting conflicting point pairs or weighting correspondences with similarity measures, for example, making use of probabilistic tools such as Expectation Maximisation to assign probabilities to each candidate point pair (Granger, Pennec & Roche (2001)). Additional work involves altering the measure of alignment error, typically using a weighted sum of the squared Euclidean distances between corresponding points. Employing data structures such as a KD-tree facilitate fast point pair search and therefore fast registration speed. A comprehensive review of existing ICP variants is found in (Rusinkiewicz & Levoy (2001)). Alternative pairwise registration techniques include that of Chen and Medioni (Chen & Medioni (1992)) who minimise

point to (tangent)-plane distances rather than point pairs and also propose the strategy of merging additional views into a single meta-view. This is similar to our approach in that information from all previously merged views is made use of when registering a new scan however their technique still registers views incrementally using simple averaging while we make use of information from all views simultaneously.

Multiple view registration is a more challenging problem. Typically ten or more views are required to reconstruct a complete object with each view overlapping with several neighbouring scans. As discussed the two main approaches are sequential (local) registration and simultaneous (global) registration. Simple sequential registration techniques such as that proposed by Chen and Medioni align two overlapping views in turn. The points of each aligned view are then merged into the metaview until each view has been aligned. This approach, although common, is usually suboptimal as pairwise registration errors are able to accumulate and propagate. Global registration attempts to mitigate this by aligning all scans at the same time and mediate registration error evenly between all overlapping views.

Extensions of the ICP algorithm have been proposed for simultaneous registration of multiple range images. However handling multiple range images simultaneously dramatically increases the computational time. As shown by Eggert et al. (1998) it takes $\mathcal{O}(r^2 N log(N))$ operations to find all point correspondences across pairs of r point sets with N points each. Such methods become impractical as the number of range images become large.

To solve the registration error accumulation and propagation problems Bergevin et al. (Bergevin et al. (1996)) match points in each view with all overlapping views and compute a rigid transform that registers the active scan using the matching points from all overlapping views. By making use of all overlapping views this approach attempts to diffuse errors among all viewpoints as the process is iterated to convergence. Converging to a steady-state using this approach may be slow and computationally expensive. A similar iterative approach that computes the "mean rigid shape" of multiple point sets was proposed by Pennec (1996) but point correspondences had to be specified manually as a point matching algorithm was not included. An early numerical solution is proposed by Stoddart & Hilton (1996) yet slow convergence may occur (particularly with cases of near degenerate point sets). A review of comparable early multi-view registration methods was carried out by Cunnington and Stoddart (Cunnington & Stoddart (1999)).

Further early multi-view work by Eggert et al. (Eggert et al. (1996)) constrain the point pairings such that points of each scan match with exactly one other point and then minimise the total distance between paired points. The transformation update is then solved for by simulating a spring model. Our work is similar in that we minimise an energy system representing the scan positions but we do not constrain points to an individual match.

One of the early true simultaneous registration works was proposed by Pulli (Pulli (1999)) where after pairwise scan alignment, each pair of registered scans are used as constraints in a multi-view step aimed at diffusing the pairwise errors. By treating the role of multi-view registration as projecting the point transformation onto a common frame of reference, accumulated pairwise registration errors can be reduced. This is achieved by limiting the difference between the position of point sets as positioned in two frames, transformed by the pairwise registration of the two frames. This effectively moves each scan, relative to its neighbours, as little as possible. In practice a greedy approach is used to limit the difference between the position of point sets as positioned in two frames (and transformed by the pairwise registration of those two frames). Formally Pulli attempts to keep the distortion $\mathcal{D}(U)$ of the points from a set U within a given tolerance ϵ where

we define $\mathcal{D}(U)$ as:

$$\mathcal{D}(U) = \sum_{u \in U} \sum_{(i,j) \in \mathcal{V}} ||P_i(u) - T_{i,j}(P_j(u))||^2$$

In this formulation $P_i(u)$ is a transformation that transforms a point u into the coordinate system of view i while $T_{i,j}$ is the transform that maps the coordinate frame j into the coordinate frame i (as found by the pairwise registration between the two frames) and $\mathcal{V}$ is the set of neighbouring view pairs for which pairwise registration is carried out. The set of points U on which to perform this greedy approach must be specified and Pulli suggests that these points can be sampled uniformly from the overlapping areas of the scan views. Since only the space of transformations is explored in this approach, memory usage is small as there is no need to retain all of the points from all views in memory at once. This allows for global registration on data sets that are too large to keep directly in memory. There is however no guarantee that optimal solutions are found. Williams and Bennamoun (Williams & Bennamoun (2001)) took a similar view attempting to minimise a similar distortion on a set of sampled points, computing the minimisation using an iterative approach and optimising individual transforms using singular value decomposition.

More recently Torsello et al. (2011) introduce a method that extends Pulli (1999), by representing the transforms as dual quaternions, and by framing the multi-view registration problem as the diffusion of rigid transformations over the graph of adjacent views. Like the approach introduced here correspondences are allowed to vary in alternation with the optimisation over the rigid transformations (but they do not discuss the convergence of this procedure). Thomas & Matsushita (2012) introduce a simultaneous registration method for dense sets of depth images that employs a convex optimisation technique for obtaining a solution via rank minimisation. They work directly with depth images rather than point clouds and extend previous work on simultaneous alignment of multiple 2D images. Fantoni & Castellani (2012) perform initial coarse alignment by proposing a voting scheme to discover view overlap relationships and then extend LM-ICP to multiple views in order to minimise a global registration error as part of their completely automated registration pipeline.

Toldo et al. (Toldo et al. (2010)) proposed a global registration framework based on embedding the Generalized Procrustes Analysis (GPA) method in an ICP framework. A variant of the method, where the correspondences are non-uniformly weighted using curvature similarity measures was also presented. Similar to our work, Toldo et al. iteratively minimise a cost function considering all views simultaneously but rely on *mutual* correspondences; matches are defined between points that are *mutually nearest neighbour* and appropriate view transforms are found by employing GPA to minimise the distance between mutual neighbours.

In this paper we propose a novel multi-view registration algorithm where view poses are estimated through an optimisation process using an energy measure defined over every point of the input data simultaneously. With every point of the input data we associate a local measure capturing the likelihood that a 3D point is located on the sampled surface. Using kernel density estimation, a fundamental data smoothing technique, we make inferences about where surfaces exist based on the data samples available and use these inferences to align scan views by gradient ascent optimisation over the rigid pose space parameters. Following pose space optimisation we then refine our kernel density model estimate iteratively. The rationale is that since only a limited number of points are sampled from the true surface, the position of every surface point is partly uncertain. By

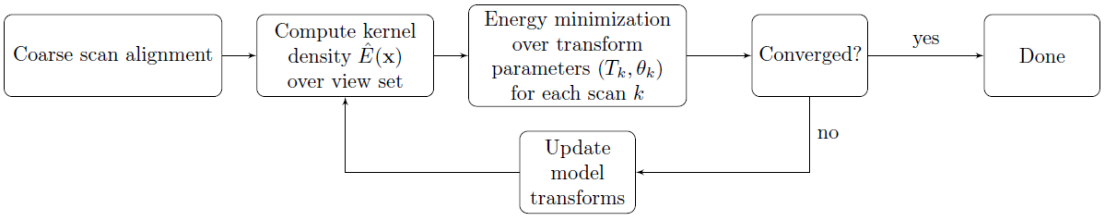


Figure 1: Our kernel density based registration approach.

capturing this fact in our density estimation approach we are able to exhibit robustness in the presence of sensor noise and scan misalignment. By optimising parameters in the transform space with respect to our energy function we reduce the amount of registration work required since no pairwise point correspondences, view pair scan alignment or view order information is required, as is commonly the case.

3. Method

3.1. *Kernel Density Estimation*

Our registration approach includes three main steps as illustrated in Figure 1: coarsely aligned scan viewpoints are provided as input, kernel density estimation is performed to determine where surfaces are most likely to exist. View poses of all point clouds are then optimised simultaneously via gradient ascent on an energy measure that relates scan position to local point density in an iterative fashion. Once convergence is reached we have the option of applying a surfacing method to the registered point set.

Point cloud data is often used as input for modelling and rendering applications and is typically acquired by sensors capable of capturing physical point geometry. The robust and accurate alignment of scattered point data is a subject of continued research (Toldo et al. (2010), Torsello et al. (2011), Tam et al. (2012), Fantoni & Castellani (2012), Thomas & Matsushita (2012)). Here we define robustness in terms of an alignment or registration technique that works well on noisy data that may also have a small number of gross errors or outliers. In this work we develop a method for the alignment of point sets acquired from varying views of physical object surfaces. Point data sets such as this are routinely generated by optical and photometric depth sensors and more recently, inexpensive consumer depth cameras such as the Microsoft Kinect (Microsoft Kinect (2010)). Sensor quality continues to improve but real-world acquisition inherently contains measurement noise.

We propose a statistical method to perform multi-view registration on potentially noisy point cloud surface data. Our method uses a non-parametric kernel density estimation scheme. Kernel density estimation is a fundamental data smoothing technique where inferences about a population are made based on finite data samples. We define a density function that reflects the likelihood a point $\mathbf{x} \in \mathbb{R}^3$ is a point on the unknown true surface $\mathcal{S}$ which is observed by point samples $\mathcal{P}$. This surface estimate is then used to guide view registration in the sensor transform space as we alternatively refine view pose positions and our model surface estimate.

3.2. *Kernel density functions for point cloud data*

Given point data $\mathcal{P} = \{p_1, p_2, \ldots, p_N\}$ one can estimate the unknown density of the data $f(\mathbf{x})$ using a simple non-parametric kernel estimator $\hat{f}(\mathbf{x})$ with kernel function $\mathcal{K}$ which

is often chosen to be a Gaussian function (although there are various other common options) with bandwidth parameter h which is a smoothing parameter. A simple kernel based estimation $\hat{f}(\mathbf{x})$ of the true density $f(\mathbf{x})$ is given as:

$$\hat{f}(\mathbf{x}) = \frac{1}{Nh} \sum_{i=1}^{N} \mathcal{K}\Big(\frac{\mathbf{x} - p_i}{h}\Big) \tag{3.1}$$

Figure 2 gives an illustration of the kernel density estimation technique for scattered data points in one dimension. Local maxima of the density estimation naturally define clusters in the scattered point data $\mathcal{P}$. Our surface density estimate (see Equation 3.2) is an adaptation of this generic approach. We use the local maxima to guide our approximation of where the sampled surface is most likely to exist and in turn adjust scan pose position in relation to this inferred surface approximation via parameter optimisation in the transform space.

3.3. *View registration using density functions*

To register multi-view point cloud data from the set of views $\{V_1, V_2, ..., V_M\}$ we first infer likely true surface positions from potentially noisy, coarsely aligned sets of scan views using our adaptation of the kernel density technique outlined above. We then use these inferences to optimise the pose of each viewpoint V_k simultaneously by rigidly moving the scan in pose space and evaluating the new point positions of the moving scan on a density function defined by the positions of all other views $\{V_j | j \neq k\}$. Using these evaluations we are able to improve the alignment of each scan by following the gradient of an energy function defined by the position of the other scan views. By moving all scans simultaneously to positions of high energy we effectively move each view to a best fit position that is most likely a location on the sampled surface.

The core of this approach involves defining a smooth energy function $\hat{E}$ that reflects the likelihood that a point $\mathbf{x} \in \mathbb{R}^3$ is a point on the surface S estimated by the current alignment of partial views. This allows us to produce an accurate approximation of the sampled surface from which we are able to guide view alignment by way of optimisation in the transform space. Once view positions have been simultaneously and independently optimised we can iteratively re-estimate $\hat{E}$ and therefore S. Moving scans by optimising in the transform space to find poses with high values of our kernel density energy lets us perform registration without requiring explicit point pair correspondences.

Based on our kernel density approach we define the energy function $\hat{E}(\mathbf{x})$ to evaluate the positions of points $\mathbf{x}$ belonging to view V_k by accumulating local contributions $\hat{E}_i(\mathbf{x})$ at $\mathbf{x}$ for each sample point $\{p_i \in \mathcal{P}\}$ where p_i are local neighbouring points contained in the stationary views $\{V_j | j \neq k\}$. We measure the local energy $\hat{E}_i(\mathbf{x})$ at a point $\mathbf{x}$ by considering the local plane fitted to a spatial neighbourhood of p_i. This plane is fitted using all points located within a spatial distance h (the bandwidth) of p_i. In practice we fit a least-squares plane (normal n_i, centroid μ_i) to the points in the spatial neighbourhood of p_i. The centroid μ_i is the weighted mean of the spatial neighbours and the plane normal n_i is found by applying singular value decomposition to a weighted covariance matrix Σ_i such that neighbour points nearer to p_i are given higher weighting:

$$\Sigma_i = \sum_{j \in \text{Neighb}(p_i)} (p_j - \mu_i)(p_j - \mu_i)^T \chi(p_j, p_i, h)$$

$$\chi(p_j, p_i, h) = \frac{1}{h \cdot \sqrt{(p_j^x - p_i^x)^2 + (p_j^y - p_i^y)^2 + (p_j^z - p_i^z)^2}}$$

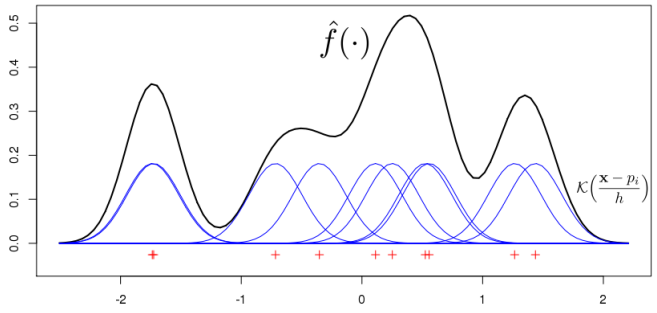


Figure 2: Example of the kernel density estimation technique for 1D data points. Data points are marked with the '+' symbol. Note that local maxima of the kernel estimation define clusters of the original data. The density approximation $\hat{f}(\cdot)$ (Equation 3.1) at a given location $\mathbf{x}$ is the summation of values contributed by local Gaussian kernels $\mathcal{K}$, located at each data point.

As noted h is our bandwidth, μ_i is the weighted average position of all samples inside the kernel and χ is a monotonically decreasing weight function, we choose a simple normalised inverse Euclidean distance weighting. Since Σ_i is symmetric and positive semi-definite, the corresponding eigenvectors form an orthonormal basis such that the largest two eigenvectors v_i^1,v_i^2 span a local plane and the third v_i^3 provides the normal direction n_i that we require. A schematic example of this is depicted in Figure 3. If normals are provided by the scanning device we can use them instead of the estimated normals.

The distance from $\mathbf{x}$ to this fitted plane provides the first component of the local energy contribution $\hat{E}_i(\mathbf{x})$. We orthogonally project $\mathbf{x}$ onto the plane and using the squared distance, $((\mathbf{x}-\mu_i)\cdot n_i)^2$, we measure the first term of the local contribution $\hat{E}_i(\mathbf{x})$ as:

$$[h^2 - [(\mathbf{x}-\mu_i)\cdot n_i]^2]$$

The bandwidth h provides the maximal distance that points may lie from $\mathbf{x}$ and still contribute to the locally estimated plane that we project the query point $\mathbf{x}$ to. The value of this first $\hat{E}_i(\mathbf{x})$ term is therefore greater than or equal to zero and by definition, positions $\mathbf{x}$ closer to locally fitted surface structure are assigned higher energy than positions that are more distant. Orthogonal plane projection provides us with a good measure of error as our 3D surface structure is locally planar and query points in well registered positions will lie on or near this local plane.

An additional assumption made is that the influence of point p_i on position $\mathbf{x}$ diminishes with increasing distance. To account for this fact we make use of monotonically decreasing weight functions ϕ_i to reduce influence as distance increases. For this we follow Schall et al. (2005) and choose a trivariate anisotropic Gaussian function ϕ_i that additionally lets us adapt to the local shape of the point distribution in the spatial neighbourhood of p_i. In practice we estimate ϕ_i parameters μ_i, Σ_i by reusing the same neighbouring points of p_i according to the distance h. From these points we again form a weighted mean vector μ_i and covariance matrix Σ_i providing the second contribution to the local $\hat{E}_i(\mathbf{x})$ as:

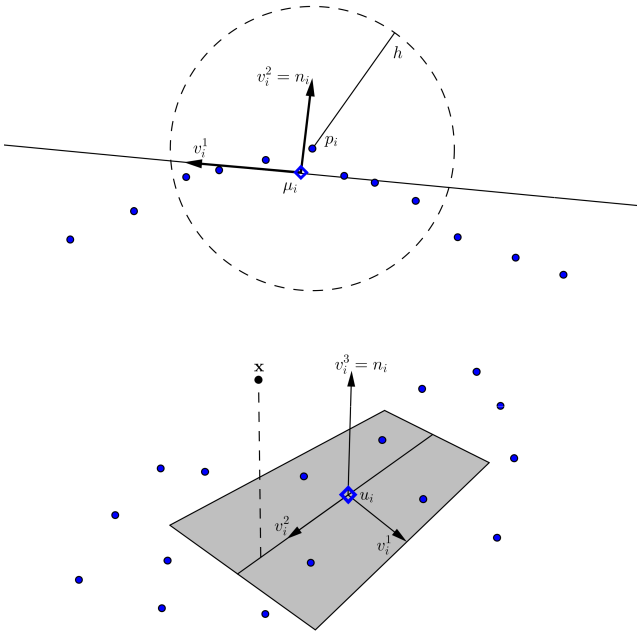


Figure 3: Upper: 2D example of a least-squares line fit found using the neighbours of p_i within bandwidth distance h. Lower: In three dimensions we orthogonally project $\mathbf{x}$ to the locally fitted plane and find the energy contribution of $\hat{E}_i(\mathbf{x})$ using n_i to provide our estimated plane normal.

$$\phi_i(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}|\Sigma_i|^{1/2}} \exp(-\frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1}(\mathbf{x} - \mu_i))$$

The product of the local plane distance term and this trivariate Gaussian term provides the local energy contribution $\hat{E}_i(\mathbf{x})$:

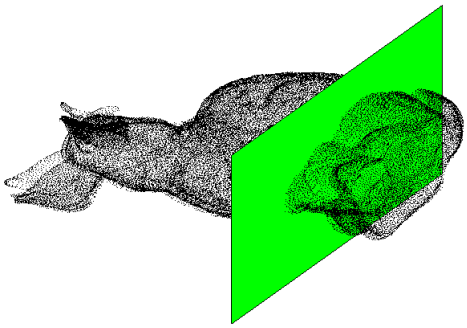
$$\hat{E}_i(\mathbf{x}) = \phi_i(\mathbf{x} - \mu_i)[h^2 - [(\mathbf{x} - \mu_i) \cdot n_i]^2]$$

This leaves us to define the full energy function $\hat{E}(\cdot)$ modelling the likelihood that a given point $\mathbf{x}$ is a point on the unknown true surface S represented by points $\mathcal{P}$. This involves accumulating and summing the local $\hat{E}_i(\mathbf{x})$ contributed by all points p_i in the spatial neighbourhood of $\mathbf{x}$ as defined by the bandwidth size h.

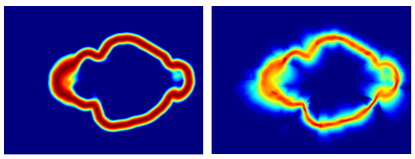
$$\hat{E}(\mathbf{x}) = \sum_{i \in \text{Neighb}(\mathbf{x})} w_i \hat{E}_i(\mathbf{x}) \tag{3.2}$$

We are able to incorporate scanning confidence measures $w_i \in [0, 1]$ associated with measurement point p_i by scaling the amplitudes of our energy functions as Schall et al. (2005) suggest. If no scanning confidences are provided we use $w_i = 1$. Figure 4 shows an example of a slice of our energy function $\hat{E}(\cdot)$.

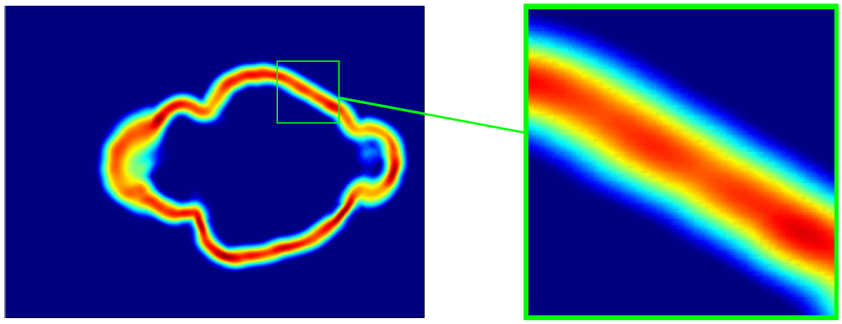
In summary we evaluate the current position of a point cloud pertaining to view V_k by passing each of its member points $\mathbf{x}$ through an energy function $\hat{E}(\cdot)$ that is defined by the current position of the remaining scans. We use the summation of the



(a) A planar slice of our energy function through coarsely aligned partial scans (Bunny data set).



(b) The resulting energy component terms before they are combined. Left: Orthogonal projection to local plane fit term. Right: Gaussian distance term. By using the product of the components we are able to dampen areas of low density yet retain valid surface shape.



(c) Our product energy function $\hat{E}(\mathbf{x})$ approximating the underlying surface defined by the coarsely aligned scans. A zoom of the slice region shows function values that are represented by colours increasing from deep blue to red. The smooth nature of our function aids the pose parameter search.

Figure 4

resulting $\hat{E}(\mathbf{x})$ values to evaluate the current position of view V_k. We then improve the position of V_k by searching for transforms that result in higher energy using quasi-Newton optimisation in the rigid transform parameter space. We guide this search by finding approximate derivatives $\nabla\hat{E}(\mathbf{x})$ with finite differencing (using the MATLAB Toolbox function `fminunc`). We apply this process simultaneously to the position of each of our M views V_k on energy functions defined by the current position of the set of the other $M-1$ views $\{V_j | j \neq k\}$. Once optimal rigid transforms $[R_k, T_k]$ are found for each pointset, we apply these to each view V_k and then update the energy functions using the new scan positions. This process of transform parameter optimization for each view V_k followed by surface re-estimation combine to form one iteration of our technique. This process requires no view order information and we terminate the procedure either after a fixed number of iterations or when energy convergence is reached (data sets experimented with here converge in less than 15 iterations). This provides a simultaneous global alignment strategy for all views. Merging the data points of the M views into a single point cloud post registration provides suitable input for surface reconstruction or other applications.

3.4. *Adaptive kernel size*

One of the difficulties with any kernel density estimation approach is that of bandwidth size selection. In a basic setting, the bandwidth parameter h which governs kernel width is typically a globally fixed radius for all kernels. In regions of high data density, a large value of h may lead to over-smoothing and a washing out of structure that might otherwise be extracted from the surface data and aid fine registration. However, reducing h may lead to noisy estimates elsewhere in the data space where the density is smaller. We note that constant kernel sizes may not be suitable for data sets with varying sampling density, poor initial alignment or considerable sensor noise. To address this problem in addition to selecting neighbours p_i in a given radius h we also follow the approach of Schall et al. (2005) and alternatively we select the k-neighbourhood of each point sample p_i to compute the kernel contribution of E_i using exactly k closest neighbours and define the local h value as the distance to the furthest neighbour in this set. This provides a method to ensure that we always have a suitable number of neighbouring points as we guarantee to include the k closest neighbours in each kernel contribution. Experimentally we find 150 - 600 neighbours per local kernel suitable in practice. This results in h bandwidth values that are on the order of one to ten times the mean inter-point distance of the data sets that we experiment with. We include the exact point neighbourhoods used for our experiments in Table 1. Using this approach we are able to adapt the local kernel to the point sample distribution in a neighbourhood of each p_i and also adjust for local spatial sampling density. As scan registration improves, the local point sampling density typically becomes tighter and the kernel width h is able to reduce adaptively to account for this.

Some justification for this adaptive bandwidth choice can be seen in Figures 5 and 6. If a small fixed radius h is used local maxima of our energy function E can be created far from the likely object surface in regions where point clouds exhibit large amplitude noise. "View cliques" may also be formed due to scan misalignment, creating regions of unwanted multi-modal density. Cliques are found when sets of scan views form groups such that views within a clique are in a good registration, but the cliques themselves are not well registered to each other. This is a common problem found in previous multi-view strategies such as (Eggert et al. (1996)). A simple example is a set of scans that forms two cliques such that each scan is well aligned within a clique but not between cliques. For methods that make use of exact point pair matching, if each point is always paired with a member from within its own clique, the inter-clique registration will not be able to improve.

Our use of an adaptive kernel size leads to larger kernel sizes in regions of large amplitude noise due to the lower sampling density and smaller kernels where scans are tightly registered. This decreases the effect of noise by reducing the contribution of noisy local maxima which in turn aids registration. We also give view cliques a high chance of intersecting during registration due to the adaptive bandwidth addressing the problem of view clique point pair matching. This results in improved registration of point sets with large scale noise and sets that are likely to form local cliques during registration.

4. Results and Applications

4.1. *OSU laser database experiments*

We tested our proposed approach using global alignment experiments and compare to other recent work on multi-view registration. We perform experiments with laser range

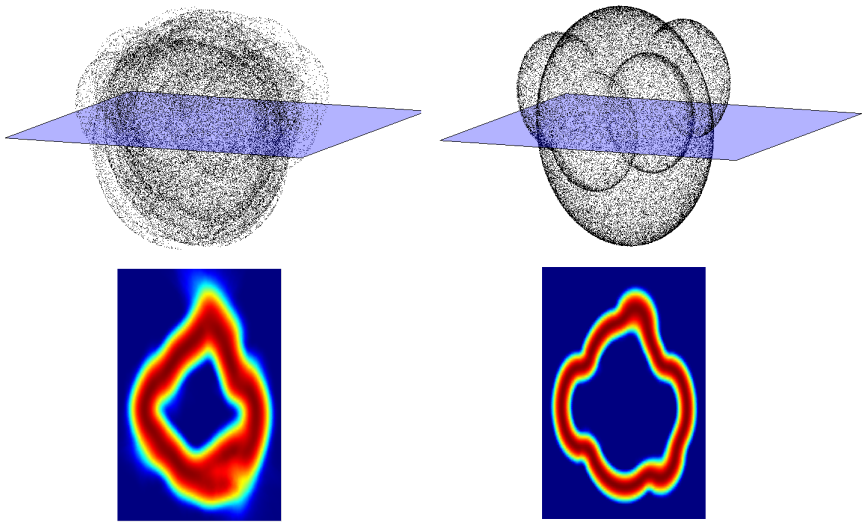


Figure 5: Adaptive kernel bandwidth illustration. Left: A planar slice through our synthetic data set with 20 scan views in their initial perturbed configuration. Right: The same slice after ten iterations of simultaneous scan pose parameter optimisation. Note how the surface estimate becomes tighter due to the improved alignment combined with our adaptive kernel bandwidth.

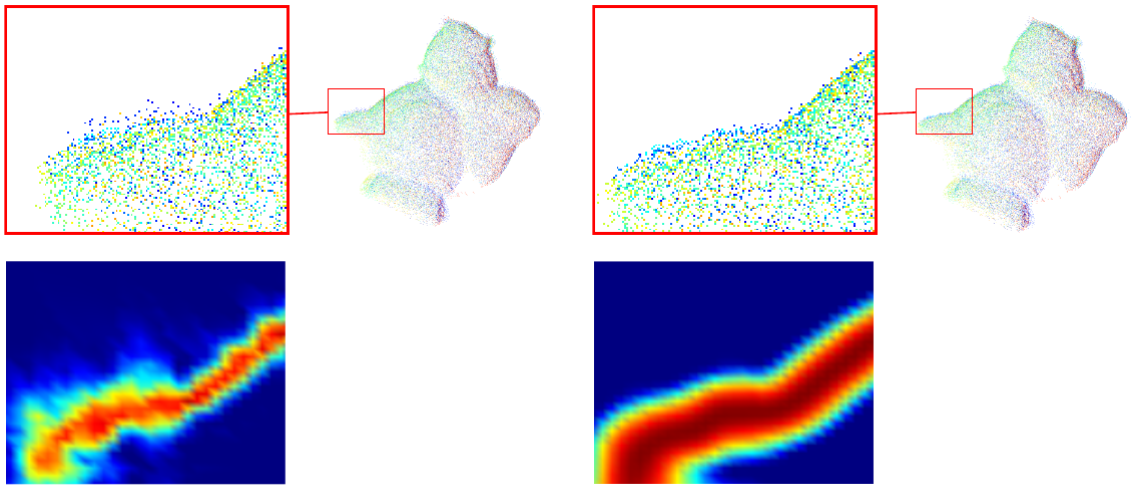


Figure 6: Adaptive kernel bandwidth illustration with OSU data. Left: The bird data set has a kernel density surface approximated using a small fixed spatial kernel size h. Slight view misalignment often cannot be drawn into a better registration configuration if local energy maxima are not in agreement with the global most likely surface position. Right: Registration result of the same data set using an adaptive k-neighbourhood kernel. Outlying maxima due to minor misalignment and sensor noise are diminished and the possibility of view cliques developing is reduced. The alignment of the scans in the highlighted region is visually improved.

Table 1: Data set statistics

Data set	**Number of viewpoints**	**Mean points per view**	**Bandwidth size k-neighbourhood**
Angel	18	6314	$k = 560$
Bird	18	4521	$k = 400$
Bottle	11	2883	$k = 160$
Teletubby	20	2671	$k = 270$
Synthetic spheres	20	4672	$k = 460$

scans from the OSU/WSU Minolta laser database (OSU database) and synthetic data sets. The OSU views come from laser scanning and the subjects used here include: angel, bird and teletubby figurines and a spray bottle. Each scan viewpoint is composed of between 4000 and 14000 points and experiments were performed by sub-sampling, for computational reasons, the views to 50 percent of their original data points. Since the data sets come from real-world acquisition, noise is intrinsically present and object point sampling is not uniform across views.

Registration experiments were also performed on synthetic data sets as they provide a straightforward way to compare registered view output to ground truth alignment. Synthetic data was created by generating sphere-like surface models with Gaussian surface noise added and partial views of these models were defined by simulating a camera position and sampling sets of points from the surface, visible to the camera. The test object sets contain between 11 and 20 scans each and a summary of the data set properties and kernel bandwidths we use for registration are given in Table 1. We choose the k-neighbourhood size for each data set such that k is 0.5% of the total number of points in the set of views. The variance of point cloud size between scans within our data sets is low so, by choosing k in this fashion, we find that our resulting bandwidth h is on the order of one to ten times the mean inter-point distance of the coarsely aligned input data. A different strategy would be required for a data set containing large variance in point cloud sizes.

Prior to registration the OSU data sets were coarsely hand aligned but some misregistration is still evident. Each viewpoint of the aligned synthetic data was perturbed with random (T_{xyz}, θ_{xyz}) transform parameters (c. 10% of sphere size translations and 10 degree rotations) to simulate a similar level of coarse hand alignment. We sampled individual viewpoints using a simulated camera.

We compare with several algorithms including a standard chain pairwise ICP method that makes use of an anchor scan and performs pairwise alignment for each pair of subsequent views. Annealing is used to decide when convergence has been reached. Although fairly straightforward in isolation, a similar approach was used by Zhou et al. (2009) as an initial registration step in their cluster based surface reconstruction work. We also compare to recent work by Toldo et al. (Toldo et al. (2010)) who perform a multi-view alignment by making use of a Generalized Procrustes Analysis based technique. The sequential order of view capture is not needed by our approach or that of Toldo et al. however the pairwise ICP technique does require this information because the algorithm depends on heavily overlapping scans.

The initial scan configurations and final alignments pertaining to the different methods are shown in Figures 11 to 15. We note that our technique is able to converge to an acceptable global minima in all experiments, however the pairwise ICP method in particular exhibits relatively large failure modes in some cases. In particular, the ICP

technique does not find acceptable alignments for the "bird", "bottle" and synthetic data experiments. The Generalized Procrustes Analysis technique in general fares well yet also exhibits some failure with the "bottle" experiment. We confirm the findings of Toldo et al. (2010) that, applied to multi-view registration problems, sequential ICP algorithms require the additional information that view order is known *a-priori* yet exhibit results that are generally worse than more recent simultaneous techniques.

We further analyse the results by examining three statistical error measures. We compute RMS residual point pair and mean inter-point distances of the converged alignment poses. The RMS residuals are computed as the root mean square distances between the points of every view and the single closest neighbouring point from any of the other $M-1$ views. With n points in total in the data set this provides n distance values $\{d_1, d_2, ..., d_n\}$ and the RMS residual is given as:

$$\epsilon_{\text{rms}} = \sqrt{\frac{1}{n}(d_1^2 + d_2^2 + ... + d_n^2)}$$

For the collection of M views our second RMS metric forces each point to identify the closest neighbouring point in *every* other scan view providing $(M-1) \cdot n$ distance measurements $\{d_1, d_2, ..., d_{(M-1)\cdot n}\}$ for a set of M views each containing n points. In a similar fashion

$$\epsilon_{\text{group\_rms}} = \sqrt{\frac{1}{(M-1) \cdot n}(d_1^2 + d_2^2 + ... + d_{(M-1)\cdot n}^2)}$$

defines our second RMS metric. This secondary RMS measure is useful in addition to the first as it penalises the previously discussed "view clique" problem where scans may exhibit good local registration yet poor inter-clique registration.

The mean inter-point distance μ_{ipd} considers the average distance between each point p_i and the nearest point p_j from all other scans combined. This once more provides n distances and we disallow pairs of points that have the same parent viewpoint. The mean inter-point distance is therefore:

$$\mu_{\text{ipd}} = \frac{1}{n}(d_1 + d_2 + ... + d_n)$$

This metric attempts to provide an evaluation measure of how tightly a group of viewpoints has been registered. Well registered sets of scans will typically exhibit a low mean inter-point distance.

We apply these error measures to the resulting alignments generated by the three registration methods. Figure 7 shows the experimental results. The approach introduced in this work performs best in terms of our RMS residual and inter-point distance evaluation metrics in the data sets experimented with. In two cases the pairwise chain ICP technique converges to a solution that increases our residual error metric above the baseline. These cases exhibit some reasonable pairwise scan alignment but global object shape is poor. In contrast, the group RMS value corresponding to applying the ICP method to the "Bottle" data set is found to be the lowest of the three techniques. This is due in part to what we call "over merging" of scans. Views are drawn together as a group but the original object shape is detrimentally affected as scans are only being registered in a pairwise fashion. Visually inspecting the alignment in this case provides evidence that the ICP result is not optimal. With this data set our method provides an improvement in the remaining two statistical measures and a visually improved registration (see Figure 8).

In two of the OSU data sets ("Angel" and "Bird") the final RMS residuals and mean

inter-point distance values our method achieves are very similar to the values resulting from the Procrustes alignment (Toldo et al. (2010)). In these cases the geometrical registration results are also visually similar and both methods exhibit good fine registration for these data sets although some differences are evident during surface reconstruction (see Section 4.4).

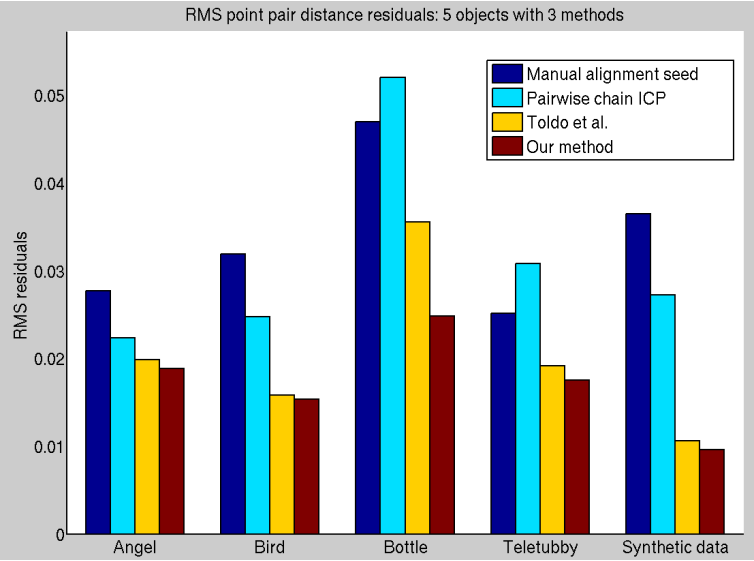
4.2. *Synthetic data experiments*

In addition to our registration accuracy exploration, we perform a further experiment with the synthetic data set to investigate the robustness of our method. A repeat experiment was carried out by seeding the synthetic data with random sets of pose perturbations and assessing alignment algorithm performance on these sets of random seed positions. Seed positions were again obtained by perturbing each scan from a set with random (T_{xyz}, θ_{xyz}) transform parameters such that the seed positions resembled coarse manual alignment. An example perturbed seed position for the synthetic data can be found in Figure 15.

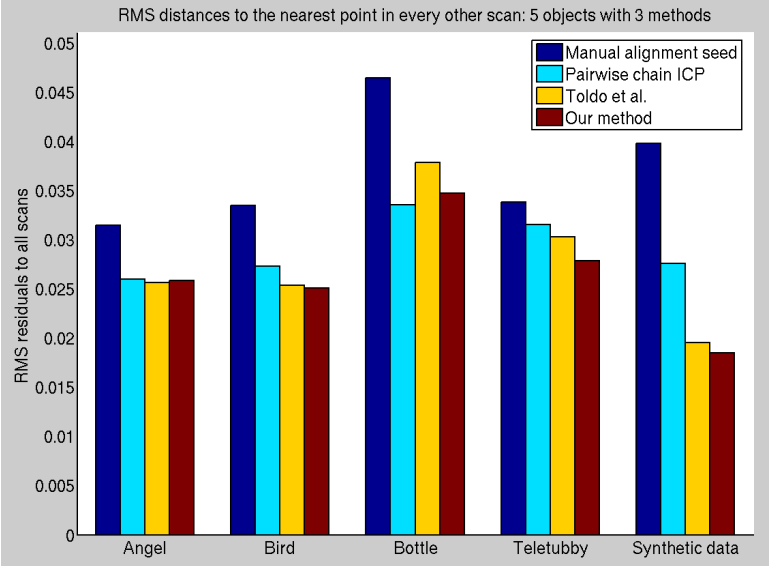
This experiment attempts to provide some insight into which algorithms are able to converge most consistently and how often gross alignment errors or failure to converge to a reasonable solution are likely to occur. We initialise the synthetic data with 20 different seed positions and measure the alignment results produced by three registration algorithms with the error measures outlined previously in Section 4.1. We report the three measures averaged over 20 seed positions for each of the three alignment methods and also report the mean seed position and known ground truth pose to estimate how far from the solution we are. Error bars indicate one standard deviation of the repeated trials. Results are found in Figure 9(a). We note that the values resulting from the Procrustes alignment (Toldo et al. (2010)) method are again similar to our method. We perform a simple paired two-tailed t-test on the post-registration results from the Procrustes alignment samples and those from our method. We find that with all three of our error metrics we obtain statistical significance at the $p \leq 0.001$ level between these techniques. We note that the mean differences between the two methods on these metrics is small $(0.0013, 0.0019$ and $0.0012)$ and our $N = 20$ is relatively low. However we find our method consistently produces lower values in almost every trial, leading to the low p values. A larger number of trials with more complex synthetic data sets would provide more evidence for the robustness of our method.

We further analysed the synthetic data results by computing RMS residuals at each intermediate step of the algorithms (Figure 9(b)). For each method, we measure RMS residuals after each view transform is applied until convergence is reached. Give that the number of transforms applied by each method varies, for display purposes, we rescale the horizontal step axis to $0 - 100$ so timing comparisons are not valid but convergence behaviour is. As with the OSU data sets, the global residual of the Procrustes method is comparable to our approach but converges to a weaker solution on our synthetic data set in 17 of the 20 trials performed. Both the Procrustes and the chain ICP methods exhibit fast initial RMS error convergence by pulling the viewpoints close together yet, particularly in the case of the ICP method, they plateau at suboptimal solutions. For tasks where final registration accuracy is of prime importance one could initialise alignment by performing a single pairwise ICP iteration before switching to our technique. We plan to explore this possibility in future work. The proposed approach converges to the best final solution in terms of closest to the ground truth RMS value in the majority (17/20) of the experimental runs on the synthetic data.

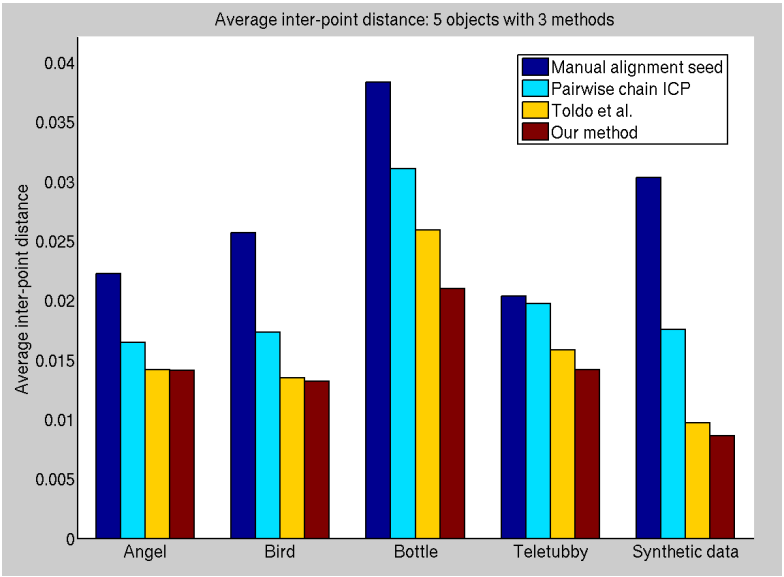
Our synthetic data set is straightforward in construction yet provides simple surface structure and a useful tool in terms of assessing how close to a ground truth position



(a) RMS residuals on converged OSU and synthetic data sets

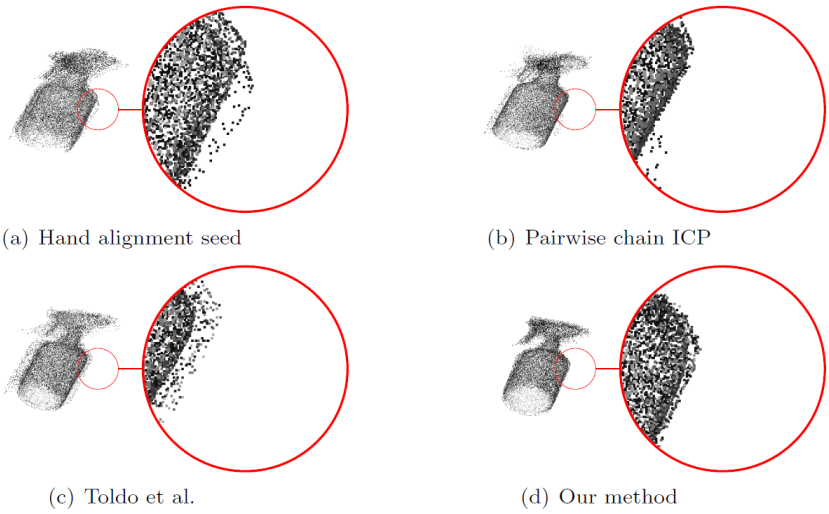


(b) Group RMS residuals on converged OSU and synthetic data sets



(c) Mean inter-point distances on converged OSU and synthetic data sets

Figure 7: Registration metrics



(a) Hand alignment seed (b) Pairwise chain ICP

(c) Toldo et al. (d) Our method

Figure 8: The OSU "Bottle" data set converges to similar poses using all three alignment techniques however we make an improvement to local surface quality. This facilitates an improvement to the global alignment of the object.

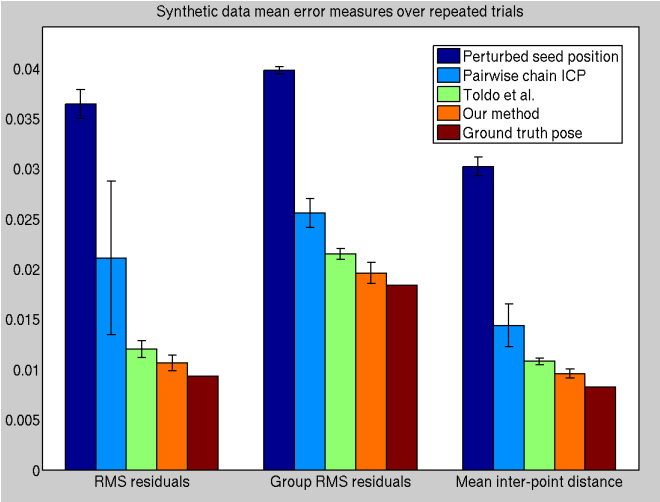
a registration algorithm is able to achieve across many runs. Averaged across 20 runs, our method consistently comes closest to the ground truth values across our three error measures. This provides initial evidence that we are able to exhibit a wide basin of registration convergence. In future we plan to perform additional experiments of this nature and increase the complexity of the synthetic data.

4.3. *Registration experiment summary*

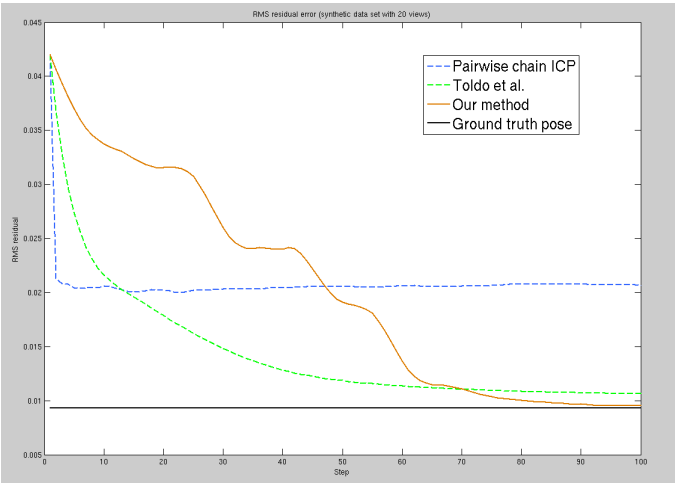
We perform registration experiments across multiple data sets evaluating results visually and with statistical error measures. Given the varied range of points per data set we note that the number of points does not seem to affect the capability of our method, working well across the range of point cloud sizes tested here. Our experimental set up using both synthetic and real data sets has demonstrated the robustness and accuracy of our proposed method for simultaneous view registration.

4.4. *Surface Reconstruction*

Surface reconstruction is an important fundamental problem in geometry processing and often uses point cloud data as input. Most reconstruction methods can be classified in terms of an explicit or implicit surface representation. Implicit methods are an important class of reconstruction algorithm as they tend to offer topological flexibility and robustness to sensor noise. However, these methods often require points supplemented with normal information to be able to reconstruct surfaces. When reconstructing implicit surfaces from data acquired from multiple views, e.g. from laser scans, accurate fine registration is especially important if point normals are not provided by the scanning technology. Alternative normal acquisition typically involves estimation using adjacent nearby points. It is therefore not practical to apply such surfacing methods to multi-view data sets that contain significant view registration errors. The registration process applied prior to constructing implicit surfaces from sets of multi-view data is an area of current research and provides a further assessment for our method.



(a) Mean values for our three error measures across 20 registration trials on a synthetic data set. Mean seed position and ground truth positions are also measured for comparison.



(b) RMS residuals at each step of the algorithms evaluated on a synthetic data run. RMS residuals are calculated as the root mean square distances between the points of every view and the single closest neighbouring point from any of the other views.

Figure 9

We apply our registration technique to sets of multi-view point clouds and then reconstruct a surface from the aligned data. For comparison we also reconstruct surfaces from the coarsely aligned input point clouds and the final registered viewpoint positions provided by the two methods we compare our registration results with. For surface reconstruction we use a well known implicit reconstruction technique, specifically the Poisson surfacing method (Kazhdan et al. (2006)). Poisson surfacing computes a 3D indicator function χ (defined as 1 at points inside the model and 0 at points outside, as dictated by the point surface normals), and then obtains the reconstructed surface by extracting an appropriate isosurface. We estimate point surface normals by fitting a plane to the

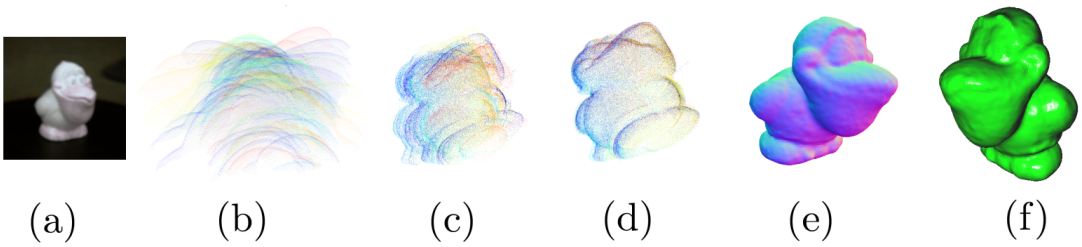


Figure 10: (a) RGB data from Ohio State University (Bird set) (b) Partial depth scans from OSU (c) Pre-energy minimization (coarse alignment by hand) (d) Multi-view registration performed with our method (e) Meshed with normal orientations (f) Phong shaded surface

ten nearest neighbour points in the aligned view sets and propagate coherent normal directions from an arbitrary starting point and use a user defined camera view point to influence the indicator function χ. Surfacing result comparisons are shown in Figures 16-20.

Applying the surfacing method directly to the coarsely hand aligned data often produces gross reconstruction failures as might be expected. The surfacing technique alone is often not able to recover appropriately from the relatively poor registration provided by our hand aligned data sets. This is especially evident in our "Angel", "Bird" and synthetic data experiments where poor alignment causes gross errors and non-smooth surfaces. Visual flaws are also evident in the surfaces that result from point clouds aligned using the simple chain ICP method in the cases of the OSU data sets. In particular results from the "Angel" data set exhibit the failure of the simple ICP method to faithfully reconstruct the wing portion of the model. We argue that this can be attributed to the minor yet evident misregistration during the alignment process. The Procrustes algorithm (Toldo et al. (2010)) generally provides good input for surface reconstruction and the "Angel" and "Bird" data sets provide surface results that are visually very similar to ours. Our method produces slightly better quality limb reconstruction of the "Angel" data set however some small geometrical errors are still present in both results. The resulting model from the "Bird" data set using the Procrustes algorithm and our method are also very similar yet our method provides small visual improvements to areas of intended high smoothness such as the feet. Enlarged versions of these results are found in Figures 21 and 22. The "Bottle" and "Teletubby" data sets exhibit significant surface reconstruction failure from the input provided by the Procrustes method yet fair better when using our technique. In conclusion, the results of applying a surfacing method to the registration results provided by our method tend to show visually improved reconstructions in the data sets experimented with.

5. Discussion and Future Work

Multi-view scan registration is typically cast as an optimisation problem. The error landscape depends on the type of data being registered, outliers, noise and missing data. As noted by Tam et al. (2012) and observed in our experiments, if the surfaces are relatively

clean and there is a good initial estimate of alignment then local optimisation such as using an ICP based method is an efficient choice. However, if there is significant noise, or the initialisation is poor, these methods may not converge.

When the view ordering $V_1, ..., V_M$ is known, registration can be performed pairwise between consecutive views and global registration can be obtained by concatenating the obtained pairwise transformations. As we observe even when all pairs are apparently well registered, misalignments occur at the stage of full model reconstruction due to registration error accumulation and propagation. In this work we propose a novel technique to tackle the task of simultaneous alignment of multiple views. By attempting to solve simultaneously for the global registration by exploiting the interdependence between all views we implicitly introduce additional constraints that reduce the global error. We base our approach on sound kernel density estimation theory. We have shown that our technique is capable of aligning depth scan sets with real-world noise amplitudes and when the seed alignment is only coarsely defined. We demonstrate the capability of our algorithm on synthetic and real-world data sets captured using laser scanners. Further to this we show that our approach can be used in conjunction with a surface reconstruction method (Kazhdan et al. (2006)) and produce surfaces for display purposes. Figure 10 provides an example of how our algorithm fits into an object reconstruction pipeline when starting from unaligned depth data.

Methods such as ours that involve kernel density estimation of point sets contain no special handling of sharp features as which may be common in data from mechanical parts for example. Work addressing sharp features has been introduced by (Haim et al. (2010)). Incorporating this into our registration framework is one avenue of interesting future work. Our method allows every scan view to converge independently to a maxima of the energy function, so parallelism at the depth scan level is fairly straight forward. We have an implementation exhibiting this and we plan in future to apply our method to extremely large data sets.

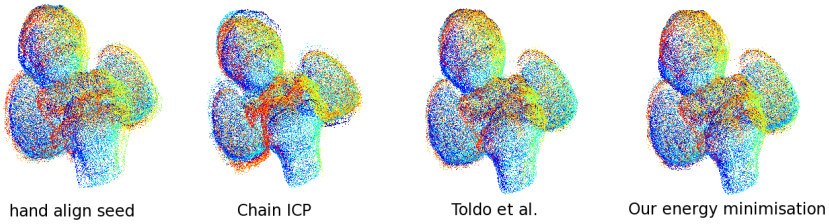


Figure 11: Angel data set final position comparison.

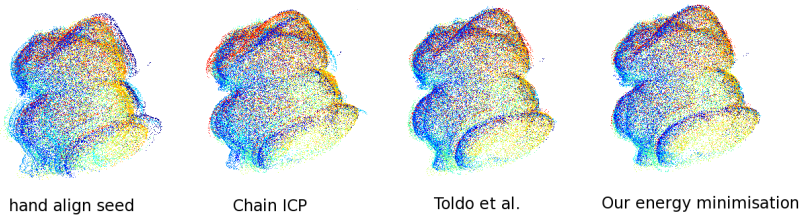


Figure 12: Bird data set final position comparison.

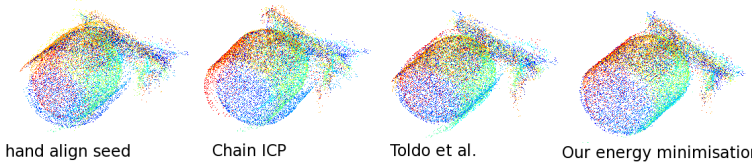


Figure 13: Bottle data set final position comparison.

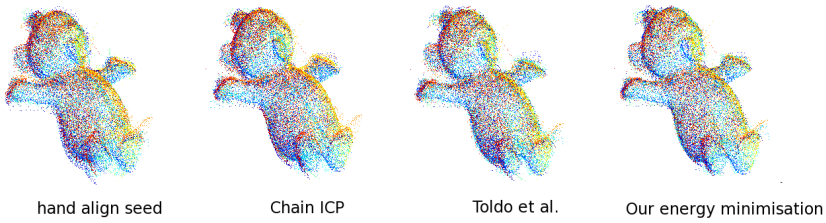


Figure 14: Teletubby data set final position comparison.

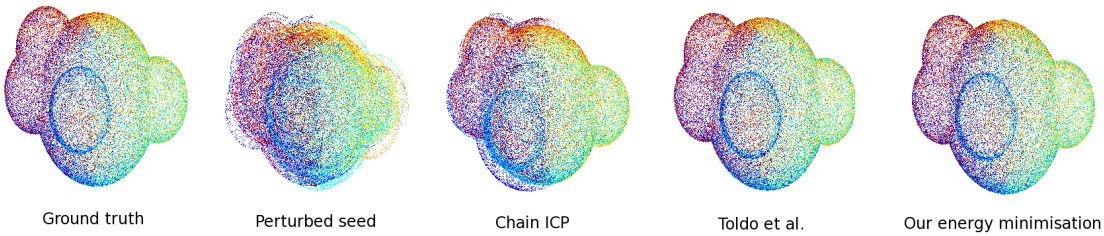


Figure 15: Synthetic data set ground truth and final position comparison.

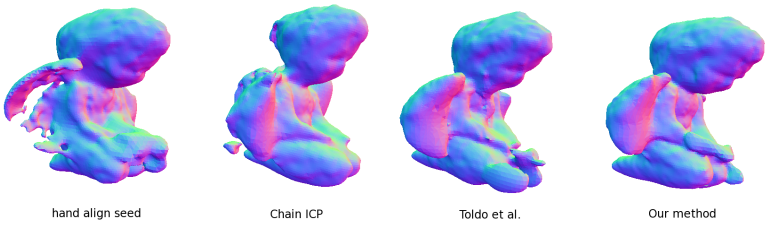


Figure 16: Angel data set. Poisson surfacing applied to final configurations.

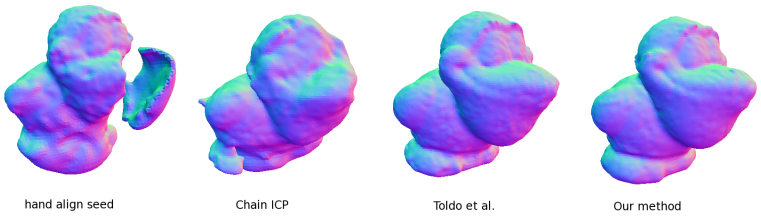


Figure 17: Bird data set. Poisson surfacing applied to final configurations.

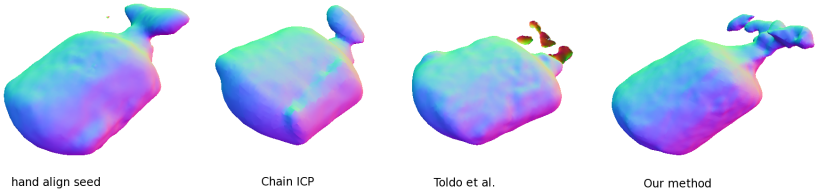


Figure 18: Spray bottle data set. Poisson surfacing applied to final configurations.

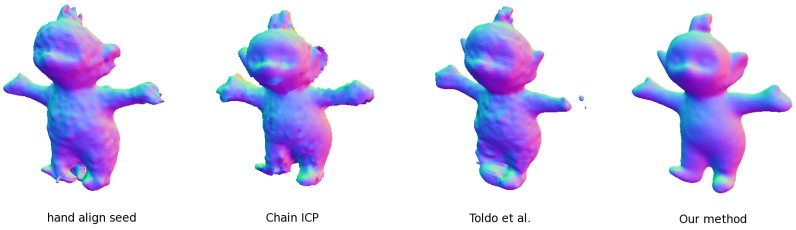


Figure 19: Teletubby toy data set. Poisson surfacing applied to final configurations.

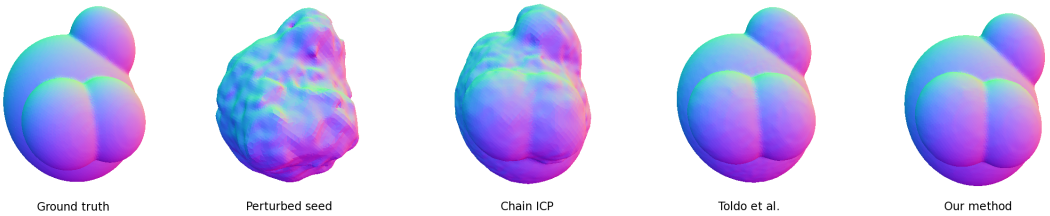


Figure 20: Synthetic data set. Poisson surfacing applied to final configurations.

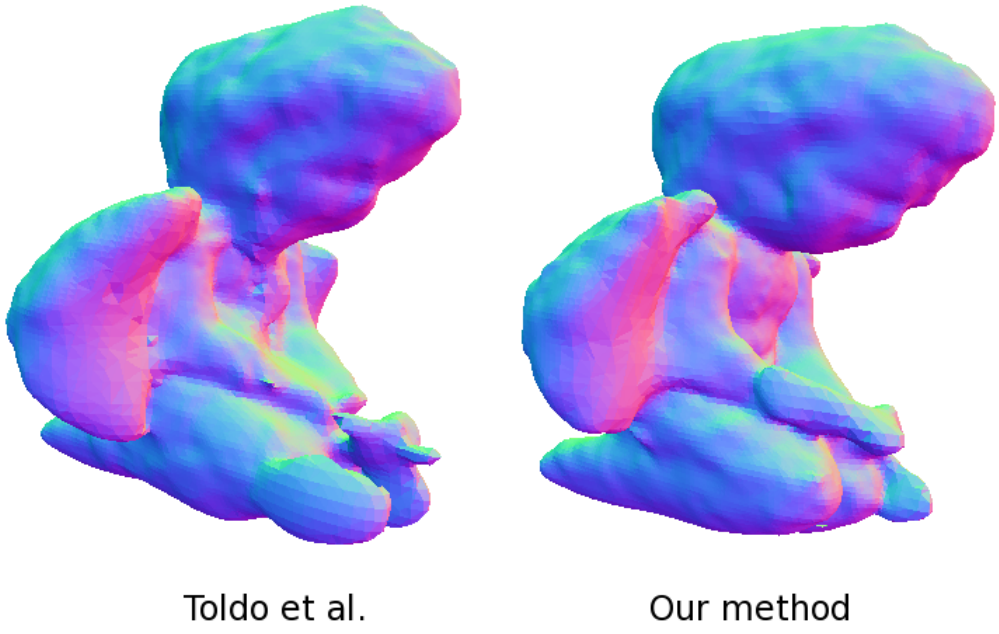


Figure 21: Angel data set registration results with Poisson surfacing applied.

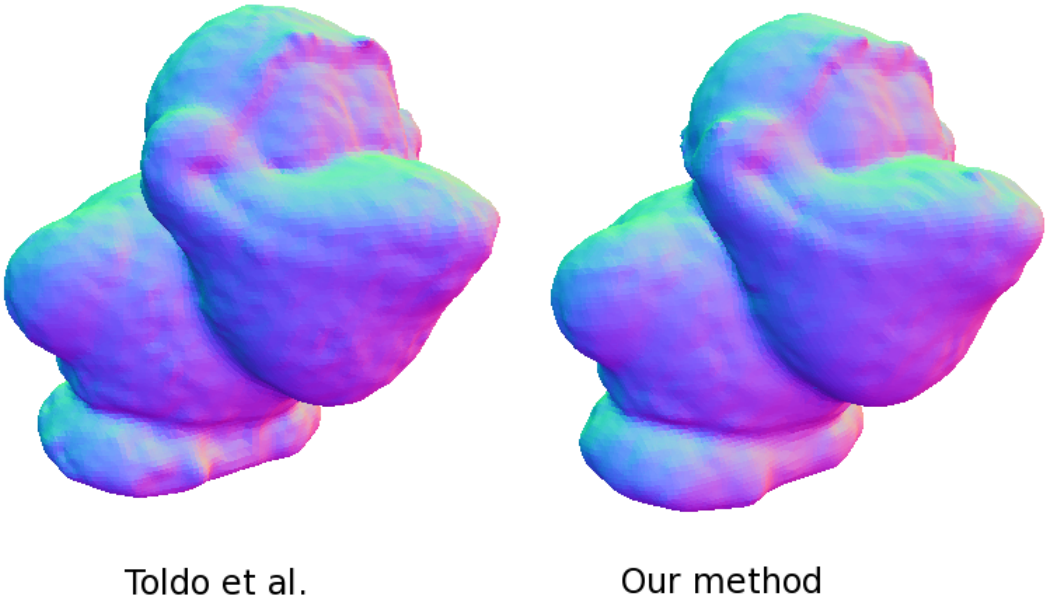


Figure 22: Bird data set registration results with Poisson surfacing applied.

REFERENCES

Audette M., Ferrie F. & Peters T. 2000 An algorithmic overview of surface registration techniques for medical imaging *Med. Image Analy.* **4**, 201–217

Bergevin R., Soucy M., Gagnon H. & Laurendeau D. 1996 Towards a general multi-view registration technique. *IEEE. Trans. Patt. Anal. Machine Intell.* **18(5)**, 540–547.

Besl P. & McKay N. 1992 A method for registration of 3-D shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI).* **14(2)**, 239–256.

Bonarrigo F. & Signoroni A. 2011 An enhanced optimisation-on-a-manifold framework for global registration of 3D range data *3D Im. Modeling Proc. Vis. and Trans 2011 (3DIMPVT).* , 350–357

Chen Y. & Medioni G. 1992 Object modelling by registration of multiple range images. *Int. J. Computer Vision and Image Understanding (IJCVU).* **3(10)**, 145–155.

Clements L., Chapman W. C., Dawant B. M., Galloway R. L. & Miga M. I. 2008 Robust surface registration using salient anatomical features for image-guided liver surgery: algorithm and validation *Medical Physics* **35(6)**, 2528–2540

Cunnington S. & Stoddart A. J. 1999 N-View point set registration: A comparison. *British Machine Vision Conf.*

Eggert D. W., Fitzgibbon A. W. & Fisher R. B. 1996 Simultaneous registration of multiple range views for use in reverse engineering *Proc. of the 13th Int. Conf. on Pattern Recognition.* , 243–247.

Eggert D. W., Fitzgibbon A. W. & Fisher R. B. 1998 Simultaneous registration of multiple range views for use in reverse engineering of CAD models *Computer Vision and Image Understanding* **69(3)**, 253–272.

Fantoni S. & Castellani U. 2012 Accurate and automatic alignment of range surfaces. *3D Im. Modeling Proc. Vis. and Trans 2012 (3DIMPVT).* , 73–80.

Granger S., Pennec X. & Roche A. 2001 Rigid point-surface registration using an EM variant of ICP for computer guided oral implantology. *Int. Conf. on Medical Image Comp. and Comp. Assisted Intervention* **4**, 752–761.

Haim A., Sharf A., Greif C. & Cohen-Or D. 2010 L1-Sparse reconstruction of sharp point set surfaces *ACM Trans. Graph.* **29(5)**, 135–147

Ikeuchi K., Oishi T., Takamatsu J., Sagawa R., Nakazawa A., Kurazume R. , Nishino K. , Kamakura M. & Okamoto Y. 2007 The Great Buddha Project: Digitally archiving, restoring, and analyzing cultural heritage objects *Int. Journal of Computer Vision* **75(1)**, 189–208

Kazhdan M., Bolitho M. & Hoppe H. 2006 Poisson surface reconstruction *SGP Proc. of the fourth Eurographics symposium on Geometry processing.* , 61–70.

Krishnan S., Lee P. Y., Moore J. B. & Venkatasubramanian S. 2007 Global registration of multiple 3D point sets via optimization-on-a-manifold *Int. Journal of Intell. Systems Tech. and App. (IJISTA)* **3(4)**, 319–340

Microsoft Corp. 2010 Microsoft Corp. Redmond WA. Kinect for Xbox 360 and Windows.

OSU Ohio State University Range Image Collection http://sampl.ece.ohio-state.edu/data/ Accessed April 2013

Pennec X. 1996 Multiple registration and mean rigid shapes. *Image Fusion and Shape Variability Techniques - Leeds Annual Statistical Workshop.* , 178–185.

Pulli K. 1999 Multi-view registration for large data sets. *3D Digital Imaging and Modelling* , 160–168.

Rusinkiewicz S. & Levoy M. 2001 Efficient variants of the ICP algorithm. *Proc. of the Third Intl. Conf. on 3D Digital Imaging and Modeling.* **1(3)**, 145–152.

Schall O., Belyaev A. & Seidel H. 2005 Robust filtering of noisy scattered point data *IEEE/Eurographics Symposium on Point-Based Graphics.* , 71–77.

Song R., Liu Y., Martin R. R. & Rosin P. L. 2011 Higher Order CRF for Surface Reconstruction from Multi-View Data Sets *3DIM/3DPVT Conf. on 3D Imaging, Modelling, Proc., Vis. and Trans.* , 156–163.

Stoddart A. J. & Hilton A. 1996 Registration of multiple point sets. *Proc. 13th Int. Conf. on Pattern Recognition* , 40–44.

Tam G. K. L., Cheng Z., Lai Y., Langbein F. C., Liu Y., Marshall D., Martin R. R.,

Sun X. & Rosin P.L. 2012 Registration of 3D Point Clouds and Meshes: A Survey From Rigid to Non-Rigid *IEEE Trans Vis Comput Graph.* ,

Thomas D. & Matsushita Y. 2012 Robust Simultaneous 3D Registration via Rank Minimization. *3D Im. Modeling Proc. Vis. and Trans 2012 (3DIMPVT).* , 73–80.

Toldo R., Beinat A. & Crosilla F. 2010 Global registration of multiple point clouds embedding the Generalized Procrustes Analysis into an ICP framework. *3D Proc. Vis. Transmission.*

Torsello A. , Rodola E. & Albarelli A. 2011 Multi-view registration via graph diffusion of dual quaternions. *IEEE Conf Computer Vision and Pattern Recognition.* , 2441–2448

Williams J. & Bennamoun M. 2001 Simultaneous registration of multiple corresponding point sets. *Comput. Vis. Image Understanding.* **81(1)**, 117–142.

Whitty M., Cossell S., Dang K., Guivant J. & Katupitiya J. 2010 Autonomous navigation using a real-time 3D point cloud *Australasian Conference on Robotics and Automation*

Zhou H., Liu Y., Li L. & Wei B. 2009 A clustering approach to free form surface reconstruction from multi-view range images *Image and Vision Computing,* **27(6)**, 725–747.