

Satellite Pose Acquisition and Tracking with Variable Dimensional Local Shape Descriptors

Babak Taati¹

Michael Greenspan^{1,2}

¹*Department of Electrical and Computer Engineering, ²School of Computing
Queen's University, Kingston, Ontario, Canada*

3bt1@qmlink.queensu.ca

michael.greenspan@queensu.ca

Abstract

A set of local shape descriptors for range data are proposed for reliable point matching in pose acquisition and tracking applications. Several local shape properties are extracted from the principal component space of various neighbourhoods and the similarity between histograms of these properties is used for hypothesizing point matches. The dimensionality of the descriptors may vary, and the effectiveness of several 1D through 9D histograms are experimentally analysed in tracking a range data sequence of a satellite model. The best performing descriptors are used for pose determination between pairs of range images and are shown to outperform two well-known existing descriptors.

keywords: *3D pose acquisition, tracking, local shape descriptors, principal component analysis, range data*

1 Introduction

Establishing feature correspondences reliably is necessary for efficient pose acquisition and tracking. *Local Shape Descriptors (LSDs)* have been utilised by researchers for encapsulating the local geometry information of various neighbourhoods and their similarity is used as an indicator of the similarity between the corresponding geometries. Point matches are then used for registering range image sequences for tracking or for registering a range image with a complete surface model for pose determination.

Ideally, descriptors should be computationally and memory efficient, discriminating, and robust with respect to a variety of inevitable conditions such as noise and self-occlusion. They should also be robust with respect to the viewing angle.

In the *Spin Images* technique [8], 2D histograms are formed based on the distances of the neighbouring points from the normal vector and the tangent plane. The *Point Signatures* method [3] constructs 1D arrays based on the distance profile of intersection of a sphere

with the object from the tangent plane. Other local shape descriptors developed for pose estimation include *Surface Signatures* [12], *Pairwise Geometric Histograms* [1], *Statistical Matrices* [11], and *Point Fingerprints* [9]. A review of some of these techniques is offered in [2].

The abovementioned methods take a minimalist approach, in that they try to construct low dimensional and compact descriptors, presumably to favour efficiency. We propose that using high dimensional descriptors could be the key to more reliable and robust point matching without sacrificing computational and memory efficiency. To this end, we have developed a large variety of variable dimensional shape descriptors, based on local properties derived from principal component analysis. These descriptors can be considered as a generalization of well-known LSDs such as Spin Images and Point Signatures and our experimental results confirm their effectiveness in comparison to the existing methods.

The main motivation for our research is developing a pose acquisition technique for tracking a satellite based on LIDAR data as part of a larger system that could enable automatic rendezvous and robotic capture of malfunctioning satellites by an unmanned spacecraft for service and redeployment [7, 6]. Other applications for a 3D pose estimation system include 3D model building, industrial inspection, augmented reality, and computer aided surgery.

2 PCA-Based Descriptors

We construct our shape descriptors based on properties extracted from the principal component space of the neighbourhood of each point in a point cloud. Eigenvalue decomposition of the covariance matrix of each neighbourhood is used to associate an orthonormal frame $(\vec{i}, \vec{j}, \vec{k})$ and three eigenvalue scalars (e_1, e_2, e_3) , representing vector lengths along each axis of the frame, to each point. Using these vectors and

Position Property	Description	Range
x, y, z	coordinates along main axes	$[-R, R]$
X_a, Y_a, Z_a	axes distances $X_a = \sqrt{y^2 + z^2}$ $Y_a = \sqrt{z^2 + x^2}$ $Z_a = \sqrt{x^2 + y^2}$	$[0, R]$
d_{ist}	reference point distance $d = \sqrt{x^2 + y^2 + z^2}$	$[0, R]$

Table 1: Basic and Extended Position Properties

scalars we generate nine basic and several extended *properties* for each point in the neighbourhood and can form a variety of histograms that carry various levels of information about the local geometry.

The orthonormal frame of each point can be treated as a Cartesian coordinate frame known as the *principal component space*. While constructing a descriptor for a point, we refer to it as the *reference point* (\mathbf{p}') and to its frame as the *reference frame*. For each neighbouring point (\mathbf{p}), the basic properties define its relationship to the reference point and consist of three *position* scalars, three *direction* scalars, and three *dispersion* scalars.

The coordinates of each neighbouring point expressed in the reference frame, x , y , and z along \vec{i} , the *major*, \vec{j} the *semi-major*, and \vec{k} , the *minor* axes respectively, form the three basic position properties. Several extended position properties can be calculated based on these coordinates. Table 1 lists some possible extended position properties, the nomenclature we use for referring to them, and the range of possible values for each property, where R denotes the neighbourhood radius.

The rotation that aligns the orthonormal frame of a neighbouring point with the reference frame can be defined with three parameters. We chose the inner products of the corresponding axes between the two frames as the three basic direction properties. The rotation can be represented in various forms and therefore it is possible to construct several extended properties. Some of these extended direction properties are listed in Table 2, where C_i is the shorthand notation for $\cos(i)$.

Eigenvalues of the neighbourhood covariance matrix form the three basic dispersion scalars. Three scale-independent extended dispersion properties are generated by normalizing the basic values by their corresponding dispersion property of the reference point. Table 3 lists the dispersion properties, where \hat{e}_i refers to a basic dispersion property of the reference point.

Figure 1 shows a satellite range image containing nearly 50,000 3D points. The neighbourhoods of two

Direction Property	Description	Range
C_θ, C_ϕ, C_ψ	inner products of axes $C_\theta = \vec{i} \cdot \vec{i}'$ $C_\phi = \vec{j} \cdot \vec{j}'$ $C_\psi = \vec{k} \cdot \vec{k}'$	$[-1, 1]$
$r_{oll}, Pitch, yaw$	ZYX Euler angles	$[-\pi, \pi]$
α, β, γ	ZYZ Euler angles	$[-\pi, \pi]$

Table 2: Basic and Extended Direction Properties

Dispersion Property	Description	Range
e_1, e_2, e_3	eigenvalues $ e_1 > e_2 > e_3 $	$[-R^2, R^2]$
$\hat{e}_1, \hat{e}_2, \hat{e}_3$	$\hat{e}_1 = e_1/\acute{e}_1$ $\hat{e}_2 = e_2/\acute{e}_2$ $\hat{e}_3 = e_3/\acute{e}_3$	$[-\infty, \infty]$

Table 3: Basic and Extended Dispersion Properties

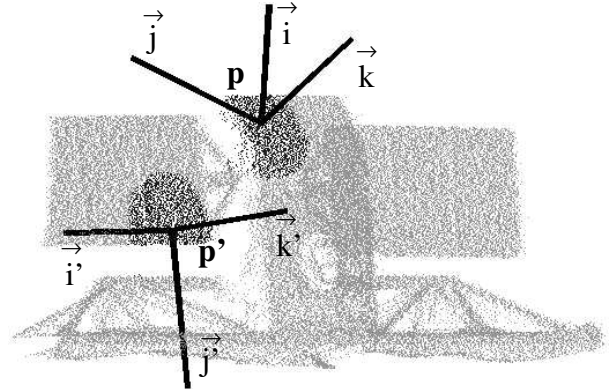


Figure 1: Two points on a satellite point cloud, their neighbourhoods, and their reference frames

selected points (\mathbf{p} and \mathbf{p}') are shown in a darker shade. The local frames of the two points are illustrated and are labeled with $(\vec{i}, \vec{j}, \vec{k})$ and $(\vec{i}', \vec{j}', \vec{k}')$. Values of the nine basic properties for point \mathbf{p} are listed in Table 4 with respect to point \mathbf{p}' as the reference point.

For each reference point, a histogram could be constructed based on one or any combination of the properties of neighbouring points. We note that only nine independent properties exist for each neighbour point and therefore we limit our histograms to have one through nine dimensions, with each dimension being one of the 22 properties listed in Tables 1, 2, and 3.

Histograms are normalized to the number of neighbouring points that contribute to their construction so

position (mm)			direction			dispersion (mm^2)		
x	y	z	C_θ	C_ϕ	C_ψ	e_1	e_2	e_3
-356	-265	16	-0.53	-0.46	0.90	4652	3429	118

Table 4: **Basic properties of point p in Figure 1 with respect to point p' as the reference point**

the descriptors are robust to sampling resolution. The neighbourhood radii for performing the PCA and for selecting points that contribute to each histogram do not necessarily need to be the same, but were set to the same value in the experiments. Histogram intersection [10] is used for measuring the disparity between histograms and provides a continuous similarity measure in the $[0, 1]$ interval, where 1 denotes identical histograms.

Since the range for the scale-independent dispersion properties is unlimited, we clip the values beyond a pre-determined maximum before constructing a histogram based on them. We have also chosen to clip the values of the basic dispersion properties beyond a certain range since the great majority of their values lie within a range much smaller than their absolute possible maximum. In the experiments, the clipping ranges were empirically set to $[0, R^2/2]$, $[0, R^2/4]$, and $[0, R^2/8]$ for e_1 , e_2 , and e_3 properties respectively, and to $[\frac{1}{2}, 2]$ for the extended dispersion properties.

3 Experiments

A set of experiments were performed to analyse the point matching capabilities of the PCA-based descriptors and compare their performance with that of Spin Images and Point Signatures. A one-fifth scale model of the RADARSAT satellite and an Optech ILRIS 3D LIDAR sensor were mounted on two manipulator robots so that the LIDAR could scan the satellite from various orientations and distances while the robotic arms moved to simulate the motion of a spacecraft approaching the satellite. Movement of the arms was slow relative to the data acquisition rate so the motion skew was negligible. Since the configuration of both manipulators was known during the scanning period, a ground truth rigid transformation existed between various scans. A VRML model of the RADARSAT satellite is shown in Figure 2(a).

Two range images from the sequence were selected for the first point matching experiments in order to select a small subset of PCA-based descriptors. The disparity in the viewing angle for the two range images was 10° in the roll angle and 10° in the yaw angle. One hundred points were randomly selected on the

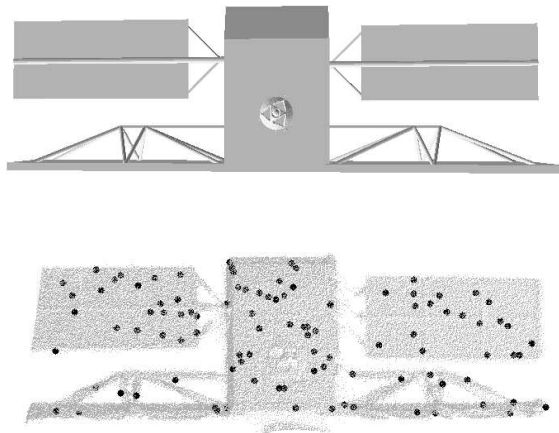


Figure 2: (a) RADARSAT VRML model (b) Corresponding points from the second image for the 100 randomly selected points on the first image

first point cloud such that they were also visible on the second point cloud. Based on the ground truth transformation, the corresponding point for each of these random points on the second scan were identified. The second point cloud and the selected points on it are illustrated in Figure 2(b).

Figure 3 illustrates a sample confusion matrix generated for the hundred points based on 6D $[e_2, e_3, X_a, e_1, Z_a, \gamma]$ histograms of size 4^6 . This confusion matrix is a 100×100 matrix in which the rows and columns represent the selected points on the first and second point clouds respectively. In the illustration, higher similarities are shown in darker shades and it can be seen that the main diagonal elements of the matrix are mostly darker than the off diagonal elements. Each diagonal element that is larger than all the other element in its row (or column), represents a point that is correctly matched to its true match on the other range image. In this case, there were 73 points that were correctly identified.

Since we use RANSAC [4] for generating statistically stable poses from candidate correspondences, it is not strictly necessary for the points to be matched exactly with their true match as long as their true match gains a high rank based on the similarity between their descriptors. We define $Rank_i$ as the number of true matches that were ranked within the top- i highest similarities. $Rank_3$ was used as a performance measure in the experiments. The results showed that the relationship between $Rank_3$ and $Rank_1$ (i.e. number of correct matches) was roughly linear.

The bounding box for the satellite model is approximately $2.1m \times 0.7m \times 0.5m$. The sampling resolution is such that range images of the tracking sequence each

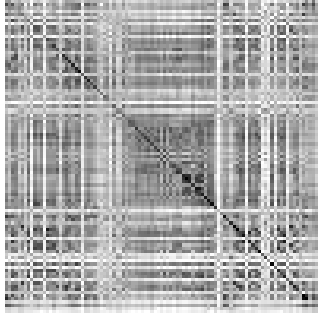


Figure 3: **A sample confusion matrix based on 6D $[e_2, e_3, X_a, e_1, Z_a, \gamma]$ histograms**

contain about 40k-70k 3D points (47,884 and 56,743 for the two selected images). The neighbourhood radius was empirically set to 15cm. The number of bins for histograms along each dimension were set to 100, 35, 11, 6, 5, 4, 3, 3, and 3 for 1D through 9D histograms respectively.

Based on the 22 properties listed in Tables 1 to 3, over a million possible histograms could be constructed for each point¹. A forward selection scheme was utilised for obtaining a suitable subset of descriptors from the vast pool of possible PCA-based histograms. First 22 1D histograms based on properties listed in Tables 1, 2, and 3 were tested and the top 8 descriptors were selected based on the *Rank₃* measure. The top 8 1D descriptors were then combined with all the 22 properties to form several 2D histograms and the 8 best performing 2D histograms were identified. Similarly, the top 8 histograms of higher dimensions were selected and each in turn were used to construct 8×22 higher dimensional histograms. In the end, we identified 8 top-performing descriptors for each dimensionality based on their point matching capabilities.

Table 5 lists the 2 top performing histograms for each dimensionality. The basic dispersion properties seem to provide the highest discrimination power among the one dimensional histograms. Figure 4(a) illustrates the *Rank₃* and *Rank₁* measures for the top performing descriptor and the average of these measures for the top 8 descriptors at each dimensionality. The *Rank₁* and *Rank₃* measure are larger than 50% for 70% respectively for all 2D and higher dimensional descriptors. The measures are larger than 60% and 80% respectively for 4D and higher dimensional descriptors. It can also be observed that higher dimensional descriptors ($\geq 4D$) provided better point matching than the lower dimensional ones despite the fact that they were more coarsely sampled.

¹ $C_1^{22} + C_2^{22} + \dots + C_9^{22} = 1,097,789$

Dim.	Selected Histograms
1D	$[e_1], [e_2]$
2D	$[\hat{e}_2, e_3], [e_2, e_3]$
3D	$[\hat{e}_2, e_3, y_{aw}], [\hat{e}_2, e_3, \hat{e}_1]$
4D	$[e_2, e_3, X_a, e_1], [e_2, e_3, Z_a, e_1]$
5D	$[e_2, e_3, X_a, e_1, Z_a], [e_2, e_3, X_a, e_1, d_{ist}]$
6D	$[e_2, e_3, X_a, e_1, Z_a, \gamma]$ $[e_2, e_3, X_a, e_1, d_{ist}, \gamma]$
7D	$[e_2, e_3, X_a, e_1, Z_a, Y_a, z]$ $[e_2, e_3, X_a, e_1, Z_a, Y_a, d_{ist}]$
8D	$[e_2, e_3, X_a, e_1, Z_a, Y_a, z, d_{ist}]$ $[e_2, e_3, X_a, e_1, Z_a, \hat{e}_3, \hat{e}_1, pitch]$
9D	$[e_2, e_3, X_a, e_1, Z_a, \hat{e}_3, \hat{e}_1, \alpha, \hat{e}_2]$ $[e_2, e_3, X_a, e_1, Z_a, \hat{e}_3, \hat{e}_1, pitch, \alpha]$

Table 5: **Top-2 selected histograms at each dimension**

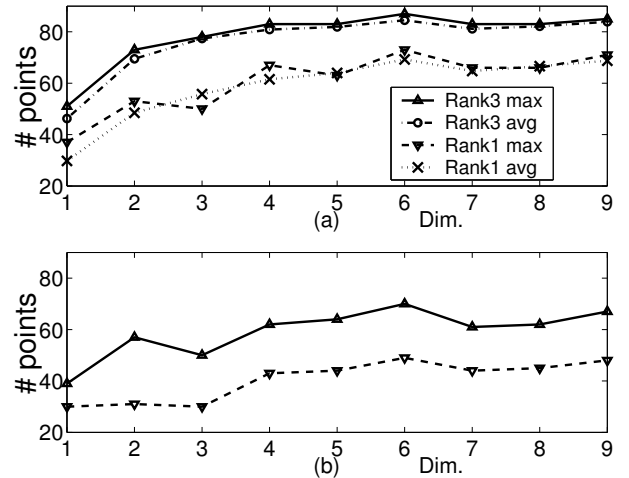


Figure 4: **Rank₃ and Rank₁ measures for the first (a) and the second (b) pair of satellite range images**

A similar experiment was performed on a second pair of range images to confirm the discrimination power of the selected descriptors. One of the range images from the first pair was replaced with another image from the scanning sequence such that the new pair had a larger view point disparity with 20° and 10° differences in roll and yaw angles respectively. Figure 4(b) plots the *Rank₃* and *Rank₁* measures for the previously selected top performing descriptors applied for point matching across the second pair. As expected, the numbers are smaller due to the larger disparity between the view points but we notice that *Rank₃* is still over 50% for all 2D and higher dimensional descriptors. We also observe that the 6D histograms provide the best point matching capability.

Our final experiment was point matching between

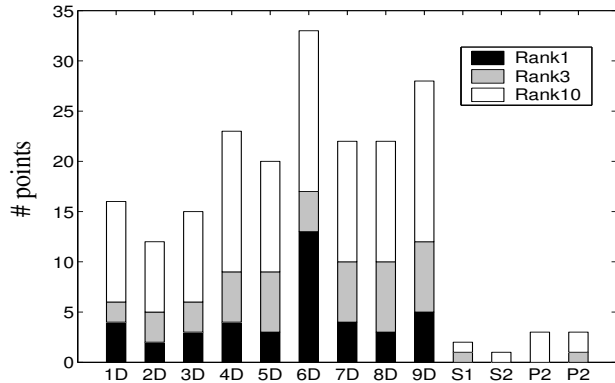


Figure 5: **Rank₃, Rank₁, and Rank₁₀** for matching sparse scene points to dense model points

the second pair of satellite range images for pose estimation. Descriptors were constructed for 2% of the points on one image and for 0.1% of the points on the second image. Due to memory limitations, the number of bins along each direction for 9D histograms was set to 2 in this experiment.

Since the LIDAR images are dense, the 2% subsampling provides a rather large set of points and this experiment is equivalent with constructing descriptors for a dense subset of model points offline and for a sparse set of scene points online. Therefore, in this experiment we refer to the densely subsampled image as the *model* and to the sparsely sampled image as the *scene*. The subsampling rates lead to almost 1,000 descriptors on the model and 57 descriptors on the scene.

Scene descriptors were matched to model descriptors based on the selected PCA-based descriptors, as well as Spin Images and Point Signatures. Figure 5 shows the *Rank₁*, *Rank₃*, and *Rank₁₀* measures for the PCA-based descriptor at each dimensionality, for Spin Images of size 35×35 (S1) and size 50×50 (S2), and for Point Signatures of length 24 (P1) and length 100 (P2). The parameters for Spin Images and Point Signatures were set according to the suggested values in [8] and [3]. The sphere radius for Point Signatures was set to 10cm. The neighbourhood radius for constructing Spin Images was set to 15cm and the neighbourhood radius of 4cm was used for estimating the surface normal directions. We observe that even the low dimensional PCA-based descriptors outperform both Spin Images and Point Signatures.

To further illustrate the point matching qualities of our PCA-based histograms against Spin Images and Point Signatures, Figure 6 shows rank histograms of the top PCA-based 1D and 6D histograms, 35×35 Spin Images, and Point Signatures of length 24. The rank histograms contain 20 bins, each representing 50

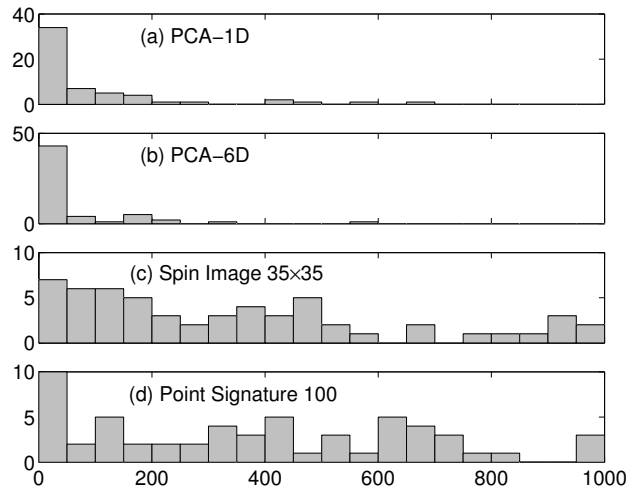


Figure 6: **Rank histograms**

points. The first column therefore, represents *Rank₅₀* or the number of scene points (out of 57) for which their true match on the model was ranked between 1 and 50 based on their histogram similarity. The second column represents *Rank₅₁₋₁₀₀* and so on. We observe that the first column contains a significantly larger number of points for the PCA-based descriptors than it does for Spin Images and Point Signatures.

Our implementations of Spin Images and Point Signatures were slightly simplified version of the original methods. The simplifications in Spin Images include not compressing images and using histogram intersection rather than the image matching technique presented in [8]. Neither simplification should affect the point matching performance of Spin Images since image compression is performed solely for the purpose of gaining memory efficiency, and the suggested image matching in [8] is devised for dealing with outliers and no outliers exist in our experimental data. We use simple histogram intersection for comparing Point Signatures instead of using the shift and compare strategy devised in [3]. This affects the matching quality for signatures with multiple global maxima.

4 Conclusion

We have developed a large set of variable dimensional local shape descriptors for 3D point matching across range images. The descriptors are based on nine basic and several extended properties extracted from the principal component analysis of each point's neighbourhood. The extended properties are different expressions of the same information as the basic properties. Nevertheless, different representations of the same

information could lead to histograms that differ in their point matching capacity and experimenting with various extended properties is therefore a useful exercise.

Unlike the minimalistic approach of the previously developed descriptors, our descriptors include variable dimensional histograms of up to 9 dimensions. Previous LSD approaches such as Spin Images obtain view independence by exploiting properties such as distance from the tangent plane or distance from the normal vector. They form a low dimensional array that carries some information about the distribution of these properties for some neighbourhood around each point. View independent properties used in these methods are a subset of our PCA-based properties. For instance, when a small radius is used for neighbourhood generation, the xy plane approximates the tangent plane (possibly with an offset along the z axis) and the distance from the normal vector and the distance from the tangent plane are equivalent to the Z_a and z position properties respectively. Therefore, our PCA-based descriptors encompass well-known existing descriptors such as Spin Images as well as a large variety of novel descriptors.

We have conjectured that since the existing descriptors were selected in an ad hoc manner, they did not necessarily contain the optimal subset of information from the vast pool of available properties and we could empirically determine a more optimal subset. Our experimental results confirmed that our descriptors indeed provided more reliable point correspondences than Point Signatures and Spin Images.

While our current implementation is in the Matlab environment and does not provide meaningful results regarding the processing times, we note that the amount of computation required for constructing the PCA-based descriptors is about the same level as that of Spin Images or Point Signatures. Furthermore, the increased reliability of the point matches provided by the PCA-based descriptors could drastically reduce the RANSAC phase of a pose estimation routine and could lead to more efficient and reliable pose estimation for vision based tracking.

Our future work involves performing further experiments in registering various models with their range images. We also intend to utilize nonparametric statistical methods [5] for efficiently comparing the high dimensional descriptors in order to deal with the curse of dimensionality.

Acknowledgments

The authors would like to thank MDA Space Missions and NSERC for their support for this work.

References

- [1] A. Ashbrook, R. B. Fisher, C. Robertson, and N. Werghi. Finding surface correspondence for object recognition and registration using pairwise geometric histograms. *Proceedings of the 5th European Conference on Computer Vision*, pages 674–680, 1998.
- [2] R. Campbell and P. Flynn. A survey of free-form object representation and recognition techniques. *Computer Vision and Image Understanding*, pages 166–210, 2001.
- [3] C. S. Chua and R. Jarvis. Point signatures: a new representation for 3d object recognition. *International Journal of Computer Vision*, Vol. 25, No. 1, pages 63–85, 1997.
- [4] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, Vol 24, pages 381–395, 1981.
- [5] I. Fraser and M. Greenspan. Color indexing by non-parametric statistics. *International Conference on Image Analysis and Recognition*, 2005.
- [6] M. Greenspan, L. Shang, and P. Jasiobedzki. Efficient tracking with the bounded hough transform. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pages 520–527, 2004.
- [7] P. Jasiobedzki., M. Greenspan, and G. Roth. Pose determination and tracking for autonomous satellite capture. *The 6th International Symposium on Artificial Intelligence, Robotics and Automation in Space*, June 2001.
- [8] A. E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 3, pages 433–449, 1999.
- [9] Y. Sun, J. A. Koschan, D. L. Page, and M. A. Abidi. Point fingerprints: A new 3-d object representation scheme. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, Vol. 33, No. 4, pages 712–717, 2003.
- [10] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, pages 11–13, 1991.
- [11] G. Xiao, S. H. Ong, and K. W. C. Foong. Three-dimensional registration and recognition method for tooth crown. *International Congress on Biological and Medical Engineering - The Bio-Era: New Frontiers, New Challenges*, 2002.
- [12] S. M. Yamany and A. A. Farag. Surface signatures: An orientation independent free-form surface representation scheme for the purpose of objects registration and matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 8, pages 1105–1120, 2002.