3D Target Recognition Using 3-Dimensional SIFT or Curvature Key-points and Local Spin Descriptors

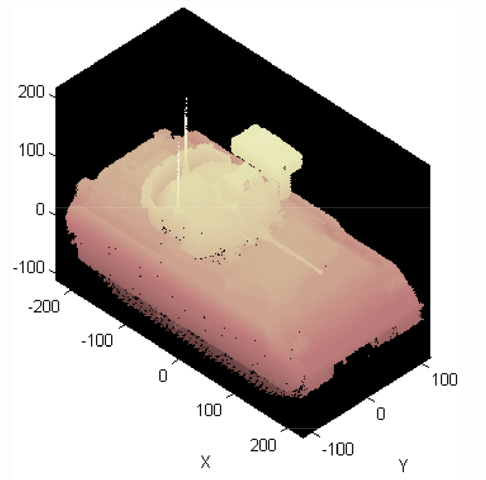
THE UNIVERSITY OF ADELAIDE AUSTRALIA

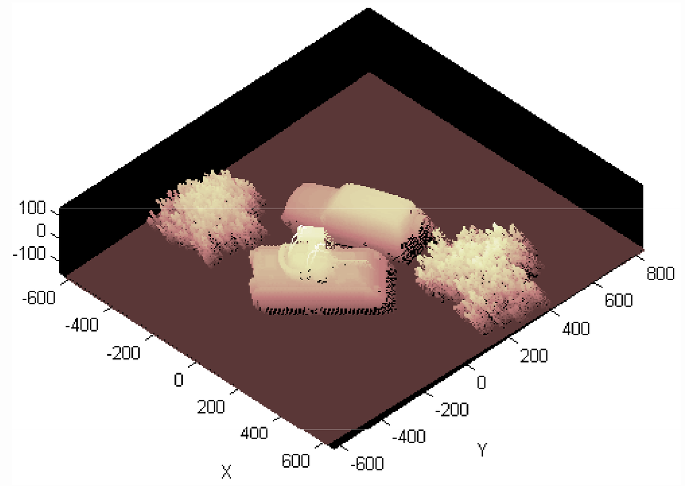
Dr Danny Gibbins,
Dept of Electrical & Electronic Engineering,
The University of Adelaide, Australia

Introduction

The recognition of a target object from 3D range data of a reference object and imaged scene has applications in aerial surveillance and in the controlling of intelligent autonomous UAV's and UGV's. Numerous approaches to recognition from 3D have been investigated such as appearance, exhaustive search and local descriptor based approaches [1].

This work describes a 3D target recognition scheme based on modified versions of 3D-SIFT or local curvature maxima techniques of Flint et.al[2] and Ho et.al.[3] to identify key-points. These are then combined with spin image descriptors[4] which are used to perform key-point to key-point matching. Clustering of these matches using Delaunay triangulation is employed to construct estimates of the target's pose which are then tested.





Above: An example 3D target model and scene generated from simulated range sensor data.

3D SIFT Key-Points

3D-SIFT is a variant of Lowe's popular 2D scale invariant feature transform method which can be applied to voxel representations of 3D data. In THRIFT[2], the 3D data is represented by a density function *D* defined as:

$$D(i,j,k) = \frac{n\left(B(i,j,k)\right)}{\operatorname{argmax}_{(i,j,k)\in I}\left\{n\left(B(i,j,k)\right)\right\}}$$

Where *n(B(I,j,k))* is the number of 3D sample points in voxel location *(i,j,k)*. Stable key-points in this representation are identified across scale-space as local maxima in the Hessian estimates:

$$H(\hat{\mathbf{x}},\sigma) = \begin{pmatrix} S_{xx}(\hat{\mathbf{x}},\sigma) & S_{xy}(\hat{\mathbf{x}},\sigma) & S_{xz}(\hat{\mathbf{x}},\sigma) \\ S_{yx}(\hat{\mathbf{x}},\sigma) & S_{yy}(\hat{\mathbf{x}},\sigma) & S_{yz}(\hat{\mathbf{x}},\sigma) \\ S_{zx}(\hat{\mathbf{x}},\sigma) & S_{zy}(\hat{\mathbf{x}},\sigma) & S_{zz}(\hat{\mathbf{x}},\sigma) \end{pmatrix}$$

where $S_{xx} = \mathbf{D} \otimes \frac{\partial^2}{\partial \mathbf{x}^2}\mathrm{g}(\sigma)$ and g is a Gaussian. The selected key-points of interest on the object are defined as:

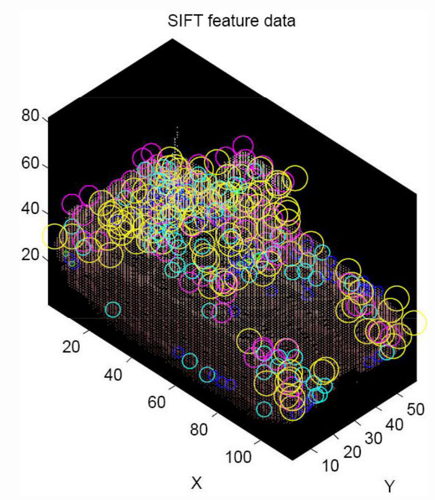
$$\mathrm{interest}(X) = \operatorname{arglocalmax}_{\hat{\mathbf{x}},\sigma} \left|\det\left(H(\hat{\mathbf{x}},\sigma)\right)\right|$$

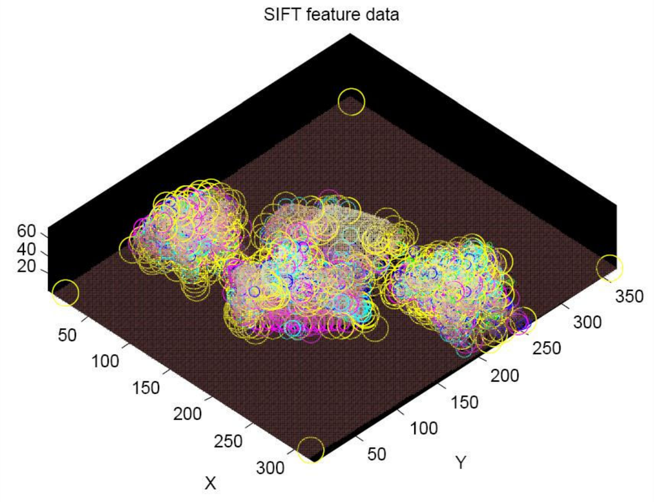
The resulting key-points are invariant to translation, scale and rotation. An examination of the density function *D* reveals that it is sensitive to variations in sampling density which can occur with scanning range sensors. As a result, the work used a modified density function *D'* defined as:

$$D'(i,j,k) = \begin{cases} 1 & \text{if } n(B(i,j,k)) > t \\ n(B(i,j,k))/t & \text{otherwise} \end{cases}$$

Curvature Key-Points

An alternative source of multi-scale key-points proposed in [3] is to consider measures of the local surface shape in terms of its principal curvatures k1 and k2, such as the local Gaussian curvature or shape-index. As with the Hessian used in [2] these estimates are invariant to translation and rotation.





Above: The modified 3D-SIFT Key-point estimation applied to the 3D model examples using a voxel resolution of 4cm.

In terms of *k1* and *k2,* the Gaussian curvature or shape-index is defined as:

$$C(\hat{\mathbf{x}},n) = k_1(\hat{\mathbf{x}},n)k_2(\hat{\mathbf{x}},n) \qquad s(\hat{\mathbf{x}},n) = \frac{\pi}{2}\arctan\left(\frac{k_1(\hat{\mathbf{x}},n)+k_2(\hat{\mathbf{x}},n)}{k_1(\hat{\mathbf{x}},n)-k_2(\hat{\mathbf{x}},n)}\right)$$

From this the curvature based key-point estimator is defined as:

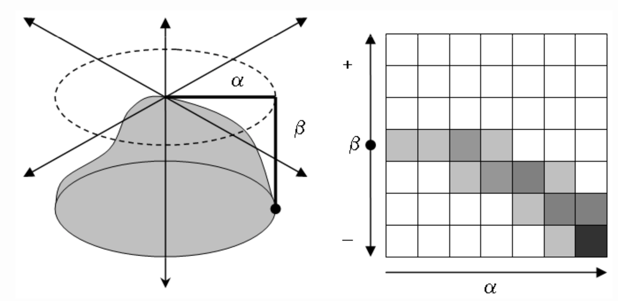
$$\mathrm{interest}(X) = \operatorname{arglocalmax}_{\hat{\mathbf{x}},n} C(\hat{\mathbf{x}},n)$$

where *n* denotes the scale at which curvature is computed for point *x*.

SPIN Image Descriptors

Spin images [4], are a well known shape descriptor used in 3D target recognition. A spin image is a discrete approximation of the surface in terms of a density function of the surface parameterised by a radial distance (α) and a height displacement (β) relative to a given surface point and its local surface normal (see illustration below):

$$\begin{aligned} \alpha(x,y,z) &= \sqrt{x^2+y^2} \\ \beta(x,y,z) &= z \end{aligned}$$



Above: an illustration of spin image representations.

For the point to point matching employed here each key-point is represented by a local 11x11 spin image descriptor representing a local surface patch of a size proportional to the key-points scale. Matching spin images can be readily identified using correlation techniques.

Matching

In the approach described here, matched key-points in the model and scene data are clustered into local groups using Delaunay triangulation. Each triangular region in the model is then compared to each triangular region in the scene data. Combinations which are spatially consistent thus constitute possible registrations of the model and scene which are then testing using a likelihood test. This method can be summarised as:

Algorithm:

- *Find high correlation matches between the model and scene spin feature measures*
- *Form a Delaunay triangulation of all points in the model with good matches to the scene data*
- *Repeat this process for the scene points with good matches with the model data*
- *Identify pairs of triangles in model data and scene data that have consistent size and shape*
- *For each identified pair compute the transformation between them and use this to project the model data into the scene and from this determine a match score*
- *Identify the match as the combination with the highest score exceeding a given match tolerance*

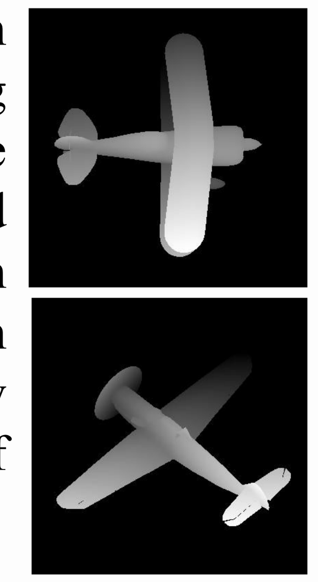
The likelihood test function for each point takes the form of:

$$m(d,c) = e^{-\frac{d^2}{\sigma_d^2}} e^{-\frac{c^2}{\sigma_c^2}}$$

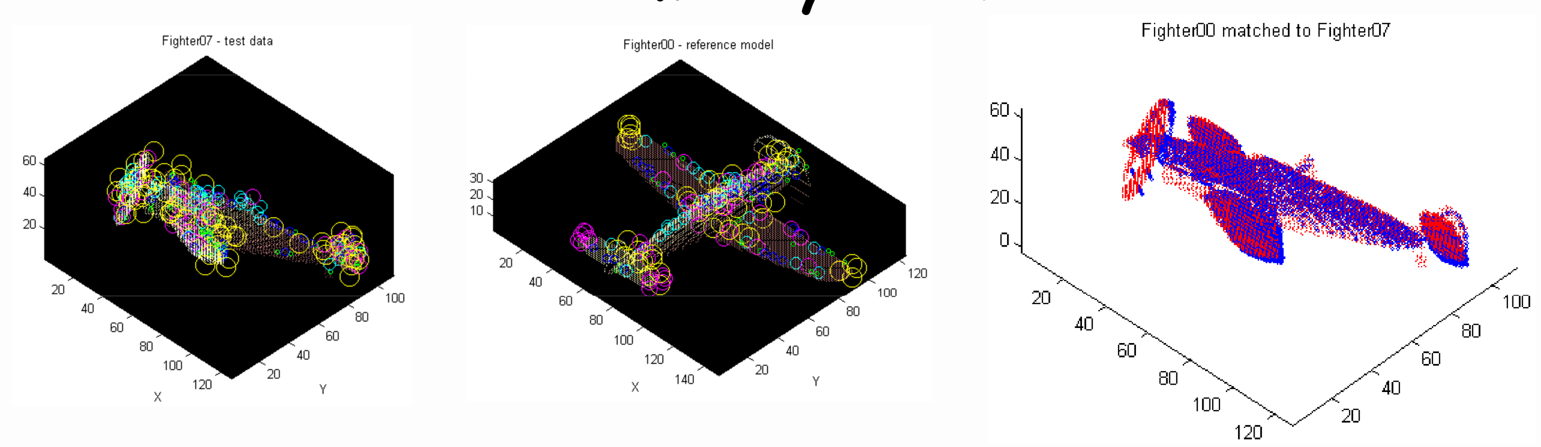
where *d* and *c* are the distance to the nearest match and its associated score.

Experiments

To test the reliability of the proposed approach to changes in object pose, range images of objects observed from differing viewpoints from the Stuttgart range image database were matched using either SIFT or curvature key-points. A second set of experiments then examined the methods ability to match a more realistic scenes containing other targets and clutter. In both cases the percentage of correct matches and the raw registration errors were used to assess the success or failure of the match.. Some results from this are given below.



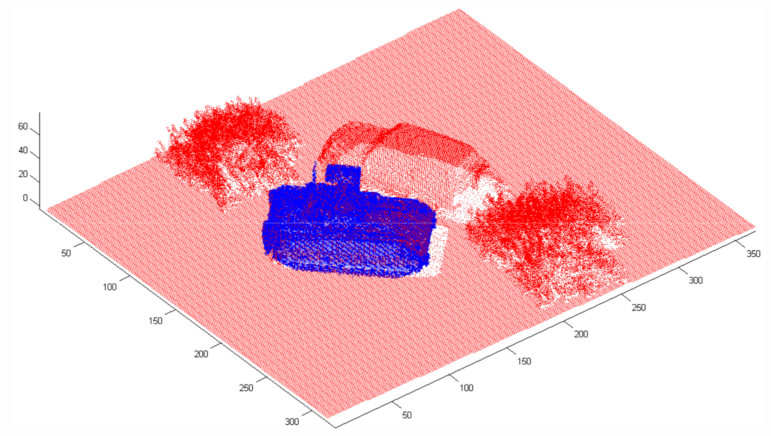
Preliminary Results



Key Point	Match	Mean Abs Error			
		d	h°	p°	r°
Hessian	100%	3.4	1.3	0.9	1.3
Curvature	93%	8.6	3.6	4.8	2.8

Key Point	Match	Mean Abs Error			
		d	h°	p°	r°
Hessian	100%	4.3	1.7	1.2	0.9
Curvature	100%	5.1	1.6	1.1	1.1

Above: Results for the aircraft test models from the Stuttgart range image database.



Key Point	Match	Mean Abs Error			
		d	h°	p°	r°
Hessian	100%	2.9	0.8	0.5	1.1
Curvature	100%	4.8	1.9	1.0	1.4
C variance	88%	21.2	4.1	9.4	17.1
SI variance	100%	6.9	3.5	1.5	3.2

Above: Match results for the M2A2 test model and scene.

Conclusions

These preliminary results demonstrate the capability of this approach to object recognition and registration for 3D targets for the simple test scenes considered here. Alignment errors for the Hessian based feature points were typically less than 2 degrees. Based on the comparative results of the Hessian and Curvature key-points, it would appear that the Hessian key-points are marginally better for target matching.

References

[1] R. Campbell and P. Flynn, "A survey of freeform object representation and recognition techniques," Computer Vision and Image Understanding, vol. 81, pp. 166–210, Feb 2001

[2] A. Flint, A. Dick, and A. van den Hengel, "Thrift: Local 3d structure recognition," in Proc. Digital Image Computing Techniques and Applications, pp. 182–188, December 2007.

[3] H. Tho and D. Gibbins, "Multi-scale feature extraction from 3d models using local surface curvature," in Proc. Digital Image Computing: Techniques and Applications (DICTA'08), (Canberra, Australia), December 2008.

[4] A. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3d scenes," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, pp. 674–686, May 1999.

Acknowledgement

The author would like to acknowledge the sponsorship of Dr Leszek Swierkowski and Dr Anthony Finn (DSTO Edinburgh, Australia) which originally motivated this research work.