

# Object Recognition Using Range Maps for Pose Estimation

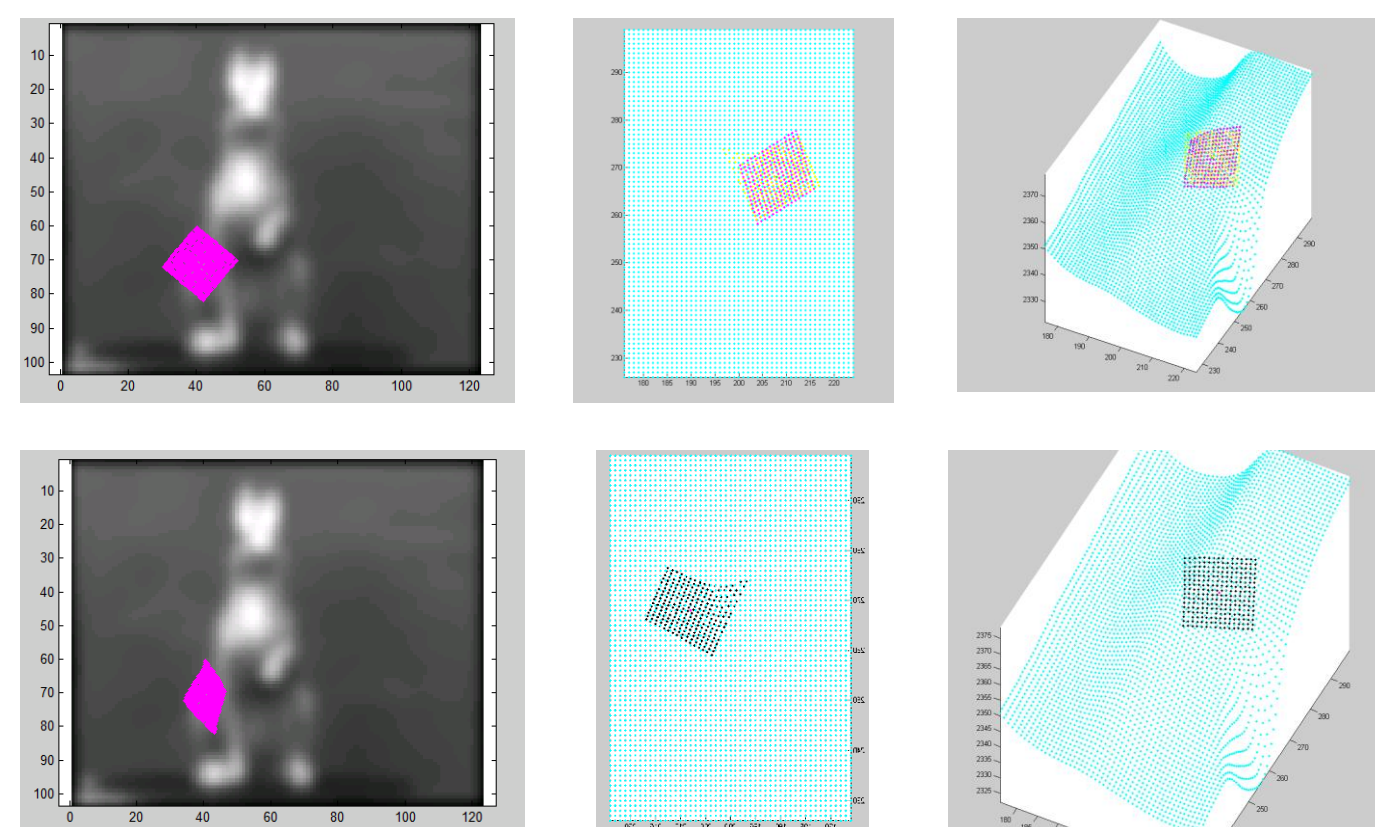
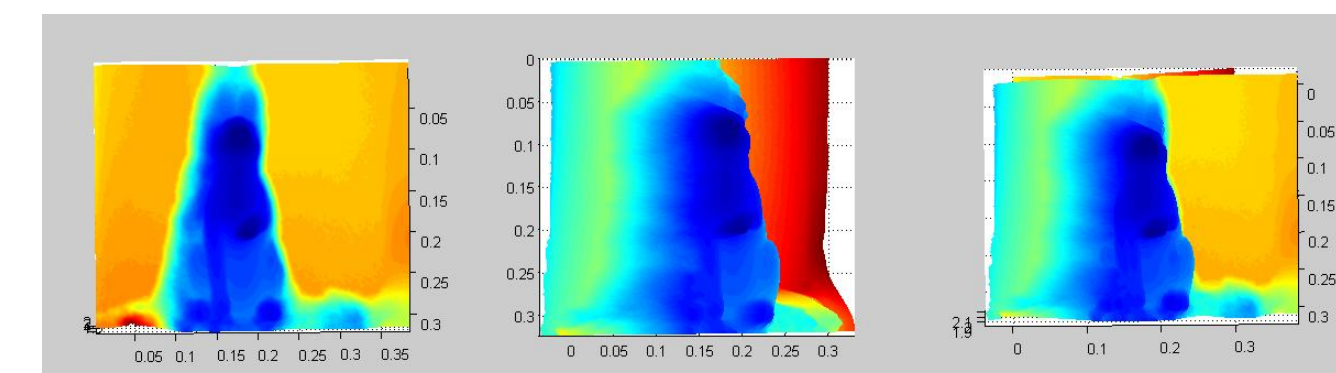
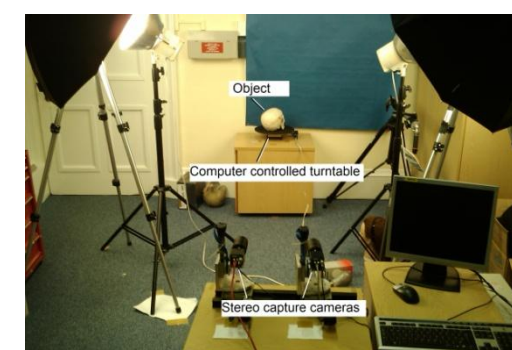
Euan Strachan, University of Glasgow; J.Paul Siebert, University of Glasgow

**Abstract** – In this work SIFT features were extracted and the performance of the point to point matches between images was compared for 2 proposed pose correction methods. Keypoints were localized in image space, the range data is then used to find the surface normal at the keypoint location and correct the SIFT descriptor sample patch in image space such that it has rectilinear sampling a normal to the range surface. Counter to expected results this form of pose correction for 3D free form objects decreased the matching performance, in our conclusions we highlight possible reasons for this.

## Pose Corrected Features

Intensity features were extracted from the 2D texture image using: SIFT, SIFT with an affine pose corrected sample patch, and SIFT with a projective corrected sample patch. The affine corrected patch was extracted by fitting a plane to the range surface underlying the support region of the local intensity feature, and then sampling the intensities using evenly spaced coordinates on the range-plane to form the 16x16 sample patch used to create a SIFT intensity feature descriptor.

To create the projective corrected sample patch the range map was treated as a 3D point cloud which could be rotated so that the surface normal at a local intensity feature could be set to zero. Hidden points were then removed and the intensity surface was resampled using evenly spaced range coordinates. Each local intensity feature sample patch was transformed in this way, so that the sample patch 3D view direction was normalised for all SIFT keypoints, descriptors were then extracted from the texture image.



## Experimental Setup

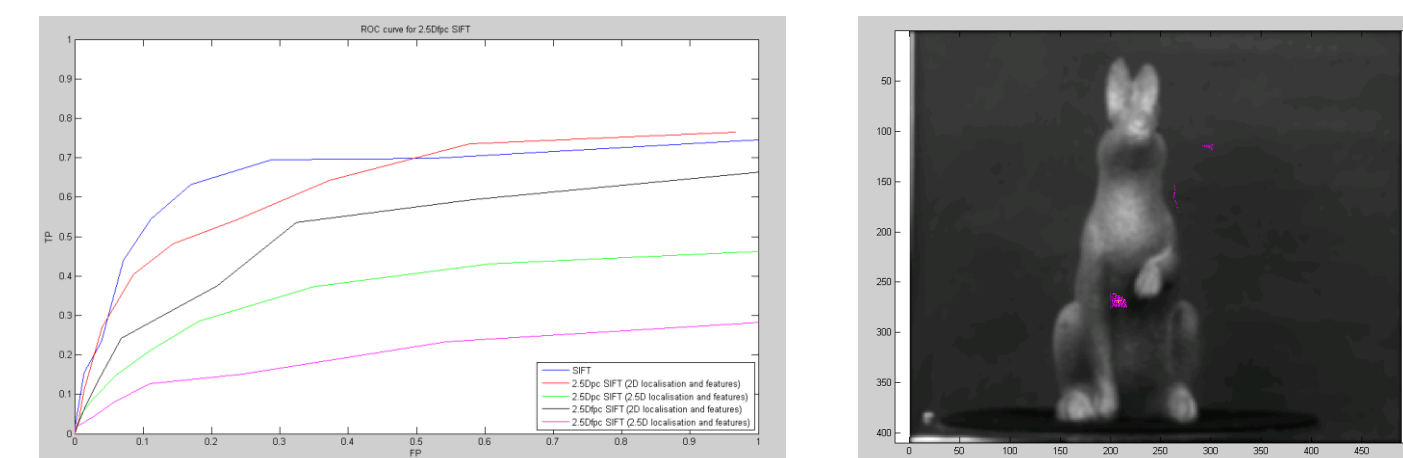
The experimental setup uses a stereo capture rig and a computer controlled turntable to capture models with a known transformation for every rotation of the turntable. This transformation together with the range map are then used as ground truth for the keypoint locations between different views of the object. Range maps for the models were produced using the software package C3D. Captures were taken in 5 degree interval to cover the full 360 degrees of models.

Model views were stored every 30 degrees, all the descriptors from these views were stored in a database for query keypoints to match to. The model views were not included in the query views.

Keypoint descriptors from a query view were matched to all model keypoints. The keypoint matches were deemed to be correct matches if the scale was within  $\frac{1}{2}$  sigma, and the keypoint footprint overlapped by 75%. Each query view was matched only to the nearest model view to avoid including multiple instances of the same keypoint descriptor.

## Results

It was found that standard SIFT outperformed both the affine corrected SIFT and projective corrected SIFT. Affine corrected SIFT performed better than projective corrected SIFT. The results are presented in the ROC curve below.



## Conclusions and Future Work

It was found when examining the cases where the matches failed that there were 3 causes. These are

- Instability in the estimate of keypoint location. This causes an instability in the surface normal, which results in the Affine and projective corrected features changing their sample patch and resulting descriptor significantly. However, as standard SIFT suffers only a translational change the resulting feature descriptor is left mostly unaffected.
- For the projective corrected features localized at an occlusion boundary, the resampling process causes some sample points to project and sample the areas in the background with large variation. This can dominate the resulting feature descriptor at the normalization stage producing an unstable feature descriptor.
- Sampling a range map off the optical axis causes the descriptor sample patch to attempt to sample 3D data which is unavailable, and create an invalid feature descriptor.

Clearly range maps do not contain sufficient information to construct a pose invariant intensity feature descriptor as described here. Our future work will be directed toward an alternative approach that learns the feature descriptor space under varying viewing transformations, using the range and image data for creating multimodal features.