# 2.5D Local Feature Matching System

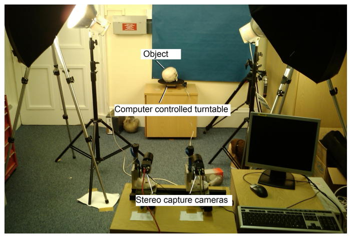
Euan Strachan, Dr J.Paul Siebert

March 2011

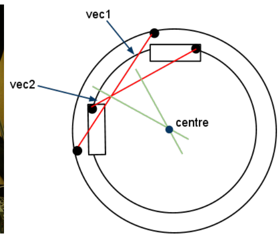
## 1 Introduction

A 3D visual tracking system was developed to investigate pose corrected surface sample patches for creating pose invariant SIFT features. The system comprises of a binocular stereo vision system and computer controlled turntable, Figure 1a. Rangemaps and images were captured of the object situated on the turntable, these were then aligned using the ground truth of the turntable motion. The system can capture both 3D surface structure and 2D surface texture simultaneously, without suffering from texture alignment issues affecting Lidar systems, such as seen in FRGC 2.0 [1].

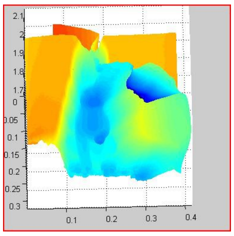
The system was used to evaluate projective corrected salient features for 3D invariant sample patches. This work was intended to extend affine invariant SIFT features [3] by using the partial 3D information from rangemaps to correct the viewing angle to the normal of the surface at the position of the interest point.



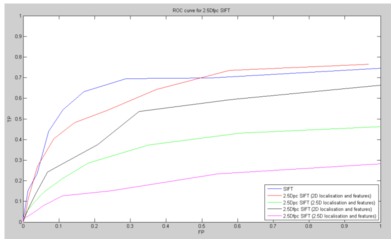

(a) Setup




(b) Turntable Center



(c) Aligned Rangemaps



(d) ROC curve

## 2 Data Capture Setup

The data capture system comprises of a stereo camera setup and calibrated turntable, rangemaps were built using the software C3D [4]. To calibrate the turntable, points on a calibration target placed on the turntable were tracked. Subsequent positions of the points were compared to create tangents to the rotation of the turntable, these were then used to find a least means square fit for the center of rotation of the turntable, Figure 1b.

The axis of rotation was found by taking the quaternion of the rotation matrix between 2 turntable rotations, and finding the axis of rotation from the quaternion. This gives the vector for the axis of rotation for the comparison of two rotations, to include all measurements the median of this vector from all observations was used.

## 3 2.5D Local features

Features were extracted from the 2D texture image using SIFT[2], SIFT with an affine pose corrected sample patch, and SIFT with a projective corrected sample patch. The affine corrected patch was formed by fitting a plane to the range surface of the local feature, and sampling evenly on the plane to form the 16x16 sample patch used to create the SIFT feature descriptor.

To create the projective corrected sample patch the rangemap was treated as a 3D point cloud which could be rotated so that the surface normal at a local feature could be set to zero. Hidden points were then removed and the surface was evenly resampled. Each local feature was extracted in this way, so that the view point was normalised for all SIFT keypoints.

## 4 Results and Conclusions

The local features were tested on their ability to match the same location on the object in successive rotations. The calibrated turntable was used as ground truth for the position changes in 3D. The results of this experiment are shown in Figure 1d. Keypoint localisation was also investigated, ie localising keypoints in texture and in range.

It was found that standard SIFT outperformed both the affine corrected SIFT and projective corrected SIFT. Affine corrected SIFT preformed better than projective corrected SIFT. Closer examination of the cases where the two proposed systems failed showed that there were 3 causes for this.

In some cases errors in the estimate of the surface normal caused the patch to warp to fit an incorrect surface, this caused no change in the original SIFT sample patch, small changes in affine corrected SIFT and large changes in projective corrected SIFT. For projective corrected SIFT the assumption that the rangemap can be treated as 3D and that rotating the points and resampling, should produce pose invariant features does not hold. This caused the resampled surface to be fit to invalid data, and the sample patch to differ further from between differing object views. Finally, for projective corrected SIFT when a keypoint is taken at a depth discontinuity, part of the sample patch will project to sample some point in the background, as this part of the sample patch varies the feature varies and loses its descriptability.

Future work will be based on learning the feature descriptor space for varying views

## References

[1] C. Boehnen and P. Flynn. Impact of involuntary subject movement on 3d face scans. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 1 –6, 2009.

[2] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[3] Krystian Mikolajczyk and Cordelia Schmid. Scale & affine invariant interest point detectors. *Int. J. Comput. Vision*, 60(1):63–86, 2004.

[4] J.P Siebert X. Ju, T. Boyling. A high resolution stereo imaging system. In *Proceedings of 3D Mod03*, 2003.