Evaluation of 3D Feature Descriptors for Multi-modal Data Registration

Hansung Kim
Centre for Vision, Speech and Signal Processing
University of Surrey
Guildford, Surrey, GU2 7NT, UK
h.kim@surrey.ac.uk

Adrian Hilton
Centre for Vision, Speech and Signal Processing
University of Surrey
Guildford, Surrey, GU2 7NT, UK
a.hilton@surrey.ac.uk

***Abstract*—We propose a framework for 2D/3D multi-modal data registration and evaluate 3D feature descriptors for registration of 3D datasets from different sources. 3D datasets of outdoor environments can be acquired using a variety of active and passive sensor technologies. Registration of these datasets into a common coordinate frame is required for subsequent modelling and visualisation. 2D images are converted into 3D structure by stereo or multi-view reconstruction techniques and registered to a unified 3D domain with other datasets in a 3D world. Multi-modal datasets have different density, noise, and types of errors in geometry. This paper provides a performance benchmark for existing 3D feature descriptors across multi-modal datasets. This analysis highlights the limitations of existing 3D feature detectors and descriptors which need to be addressed for robust multi-modal data registration. We analyse and discuss the performance of existing methods in registering various types of datasets then identify future directions required to achieve robust multi-modal data registration.**

***Keywords*-3D feature descriptor; Multi-modal data registration; 2D/3D registration; Evaluation;**

I. Introduction

The trend in digital media production is increasingly towards full 3D representation of the actors, objects and their environment. This simplifies the integration, artistic manipulation and photo-realistic rendering but requires capture of real-world scenes. A variety of techniques are commonly used to capture scenes from Light Detection and Ranging (LIDAR) which captures highly detailed geometry but lacks photometric information and is relatively slow to digital photographs which provide full photo-realism but result in lower geometric precision. The 3D approach requires a large amount of data and meta data such as camera tracks and calibration. The result is an ocean of unstructured footage which is hard to search, arrange and manage efficiently. Moreover, datasets exist in different domains with different types of format, characteristic and sources of error. Scene modelling requires the registration of multi-modal 3D datasets acquired from different sensor techniques into a common coordinate frame. Multi-modal data normally includes 3D point clouds from active sensors, manually generated 3D CG models, and 3D reconstruction from 2D video and photos as illustrated in Fig. 1.

Registration of multi-modal 2D and 3D datasets acquired

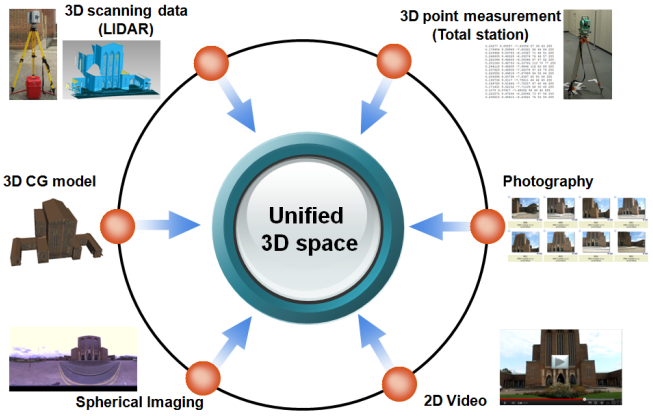


Figure 1. Multi-modal data registration

using different sensor is challenging due to the difference in information available. There have been a few researches for 2D-to-3D data matching and registration. Sattler et al. proposed a method to register a 2D query image to a reconstruction of large scale scenes using Structure-from-Motion (SfM) [12]. They implemented a two way matching scheme for 2D-to-3D and 3D-to-2D using the SIFT descriptor and RANSAC-based matching. Stamos et al. integrated 2D-to-3D and 3D-to-3D registration technique [14]. They registered 2D images to dense point cloud from range scanners by reconstructing sparse point cloud using a SfM method. 3D lines and circular feature matching were used for registration. Restrepo and Mundy evaluated performance of various local 3D descriptors for models reconstructed from multiple images [9]. They reconstructed urban scenes with a probabilistic volumetric modelling method and applied different descriptors for object classification to find the best descriptor. However previous research has focused only on registration for a single data modality.

In this paper, we propose a framework for registering multi-modal data into a unified 3D space and evaluate existing 3D feature descriptors for the framework. Direct matching and registration of a 2D image to 3D structure is a difficult problem due to differences in the information content between a 2D projection and the 3D scene structure.

Therefore we assume that 2D images are at least a stereo pair, video sequence or multiple images so that we can extract 3D geometric information from the images. Another problem of multi-modal data registration is the fact that they have different structure, resolution and sources of error. Active sensors such as LIDAR or Total Station produce a set of 3D points or point cloud without color information. Reconstruction from stereo images normally have mesh structure with colour information. SfM or Multi-view stereo (MVS) methods for a video sequence or multi-view images produce 3D point clouds with colour information.

We use 3D point clouds without colour as a basis for feature extraction and registration in this research because it is the most common element for all types of input sources. Point normals are calculated with neighbouring points and 3D keypoints are extracted using a 3D extension of Kanade-Tomasi detector based on the surface normals. We test four 3D descriptors to evaluate their robustness for registering various types of dataset.

II. Input Sources

Sensor technologies for acquisition of real-world scene structure can be classified into two categories: active methods using laser or depth sensors and passive methods using 2D photographic devices. In this section we briefly review capture devices and processing methods used in the evaluation.

Range sensors:: LIDAR is one of the most popular depth ranging techniques, which measures the range by the time delay between transmission of a pulse and detection of the reflected signal. A LIDAR scanner produces accurate depth information, but each scan only yields points visible from the scanner position. Combination of multiple scans from different locations is commonly required for full scene coverage. Architectural surveying equipment such as a total station can be used to measure a sparse set of measured points in the scene. The distance to the point as well as its azimuth and elevation are recorded. The resulting data is used to provide outline geometry and scale of the scene and to position multiple LIDAR scans with a larger context.

Video and multi-view images:: Video and still cameras are the most common sources for scene capture due to the wide availability and ease of use. Reconstructing 3D information from 2D image is one of the most active research fields in computer vision. Video frames from a moving camera can be considered as multi-view images. We use Bundler, a bundle adjustment algorithm for initial point cloud reconstruction and camera calibration for unordered image collections [13]. Then more accurate and dense point cloud is reconstructed by the PMVS algorithm [6] from the results of the Bundler. Once we register this 3D point cloud to other 3D model, we can say the original 2D images have been registered to the 3D model because 3D positions of the images are bound to the point cloud by the calibration parameters.

Spherical images:: Omnidirectional spherical imaging is commonly used to get an environment map or lighting source detection. The most common way to capture the full 3D space instantaneously is to use a catadioptric omnidirectional camera using an ellipsoidal mirror combined with a CCD. However, the catadioptric camera is difficult to calibrate and has limited resolution. Therefore, we use a commercial off-the-shelf line-scan camera, Spheron[1], with a fisheye lens in order to capture the full environment as a high resolution spherical image. In order to recover depth information from a spherical image pair, we assume that the scene is captured with the camera at two different heights as a vertical stereo pair. We use a PDE-based method for spherical line-scan image pairs proposed in [8] for 3D scene reconstruction.

Proxy model:: Simplified scene models are useful in understanding and representing rough geometry of the scene with small amount of data. Google SketchUp[2] provides a simple 3D reconstruction tool by mapping multiple photos to 3D primitives. It uses manual vanishing point alignment for photo registration to a 3D coordinate. It is useful to build simple scenes but has limitations in building complex scenes because it requires manual matchings for each primitive. We use a simple method to reconstruct a axis-aligned plane-based model from point cloud in the experiments. We assume that the world is piecewise planar and aligned to orthogonal axes (Manhattan world) [5]. Independent 3D rectangular planes are built by the plane fitting algorithm [3] and they are aligned into x,y,z planes. We refine planes by expanding planes, detecting intersections and cropping planes based on visibility.

III. 3D Keypoint Detection

Keypoint detection is an essential step prior to matching and registration. Keypoints are referred to as interest points, salient points or feature points which are distinctive in their geometry and locality. Dutagaci et al. [2] and Tombari et al. [18] survey and evaluate 3D keypoint detectors mainly for 3D objects. However all evaluations were carried out for accurate 3D models generated by computer graphics or uni-modal sensors. The evaluations were mainly performed in terms of distinctiveness and repeatability. The distinctiveness is performance to describe the characteristics of the point and find correct point matches while the repeatability means the stability to detect the same keypoints in various environments. Some 3D keypoint detectors show both high distinctiveness and repeatability according to the evaluations [18]. However, previous evaluations have focused on registration of 3D data of the same modality and are commonly limited

[1]Spheron, http://www.spheron.com/en/spheron-cgi/products/spherocam-hdr.html

[2]http://www.sketchup.com/product/newin7.html

to high-accuracy range data. A problem in applying those detectors to our datasets is the fact that the detectors do not show such high repeatability and distinctiveness for multi-modal data sets because they have potentially different source of errors especially in reconstruction from images as we pointed out in the Introduction. For example, Heat Kernel Signature (HKS) detector [15] shows good repeatability and distinctiveness in those experiments, but it is too selective to yield enough number of repeatable keypoints between active sensor model and image-based reconstruction models due to geometrical errors induced from incomplete 3D reconstruction methods. Evaluation of 3D keypoint detectors is an important topic, but we focus on the evaluation of 3D descriptors for one fixed keypoint detector in this research.

The repeatability of keypoints between models reconstructed from different sources can be extremely low. Therefore we use the Kanade-Tomasi detector [16] to extract a large number of evenly distributed keypoints for all 3D data modalities. Input of the 2D detector is replaced by surface normal vectors of a point cloud which were calculated with neighbouring points within the radius R_n, and all calculations of the detector are extended to a 3D domain. The Kanade-Tomasi detector uses an eigenvalue decomposition of the covariance matrix of input 3D normal vectors. Eigenvalues represent the principal surface directions and the ratios of eigenvalues are used to detect plane, edge and corner features. We use the Kanade-Tomasi detector to extract 3D corner features in regular grid regions by thresholding the smallest eigenvalue. We set the threshold F_t as 0.1 for all experiments in this paper.

IV. Multi-modal data Registration

A. 3D Feature Descriptors

There has been extensive research on 2D feature descriptors, but relatively little on 3D descriptors. Most 3D feature descriptors descriptors operate on a 2D domain by local projection onto a 2D tangent plane or extend existing 2D descriptors to the 3D domain. We consider four 3D descriptors which directly operate on 3D point clouds rather than meshes or 2D projection. They are used for uni-modal data classification in Restrepo and Mundy's work [9], here we evaluate their performance for multi-modal data registration.

Spin Images (SI) [7]:: Spin Images are a classic 3D shape descriptor encoding surface properties in a local object-oriented system. Using a single point basis constructed from an oriented point, the position of other points in the support radius R_s is described by two parameters in a cylindrical system. The SI computes a 2D histogram of points falling within a cylindrical volume by means of a 2D plane spinning around the cylinder axis. We set the number of bins along one dimension as 8, minimum support cosine angle between surface normals as 0.5, and the minimum number of points in the support as 16. The SI descriptor is represented as 153 dimensional vectors.

3D Shape Context (SC) [4]:: This is a 3D extension of the 2D shape context descriptor. The support region for a 3D shape context is a sphere centred on the basis point and its north pole oriented with the surface normal. A region in the support radius R_s is divided into bins by equally spaced boundaries in the azimuth and elevation dimensions and logarithmically spaced boundaries along the radial dimension. Each bin accumulates a weighted count by local point density for each point. We set the number of bins as 12 for azimuth, 11 for elevation and 15 for the radial dimension. As a result, the SC descriptor is represented as 1980 dimensional vectors.

Signature of Histograms of Orientaions (SHOT) [17]:: SHOT descriptor relies on the definition of a repeatable local Reference Frame (RF) based on the eigenvalue decomposition of the scatter matrix of surface points in the support radius R_s. Given the local RF, an isotropic spherical grid is used to define a signature structure. For each sector of the grid a histogram of normals is defined and the overall descriptor results from the juxtaposition of these histograms. The set the number of spatial bins as 32 as suggested in [17] and the angle between normal vectors as 10. The SHOT descriptor also requires a 9 dimensional vector for RF. Therefore the descriptor is represented as 329 dimensional vectors.

Fast Point Feature Histograms (FPFH) [10]:: Point Feature Histograms (PFH) are based on the combination of certain geometrical relations between neighbours in the support radius R_s. FPH extracts 4 features [α, φ, θ, d], where α is angle to the second axis, φ is an angle to the first axis, θ is a rotation on the UW plane and d is a distance between 2 points which is used for weighting parameter. The FPFH is a simpler and faster version of PFH by caching previously computed values in feature histogram computation. The number of bins is set as 11 for each α, φ, θ. Therefore the FPFH descriptor can be represented with 33 dimensional vectors.

B. Feature matching and registration

The Kanade-Tomasi detector generate a relatively large number of keypoints over the whole model. The resulting set of key-points will have a sub-set of points which can be matched between modalities but may also have many outliers which cannot be matched. In order to eliminate such outliers and accelerate matching speed, we use a sample consensus method, SAC-IA [10]. Instead of greedy search for feature matching, SAC-IA iteratively selects samples whose pairwise distances are distant enough and compute a rigid transform matrix with 6 DOF to find the best transform matrix which minimise the error metric. We do not consider scale because all datasets are reconstructed for the real world scale.

This feature-based registration provides a good initial alignment for the further refinement over the whole point cloud using the Iterative Closest Point (ICP) algorithm [1]. The ICP algorithm finds the optimal transformation between two point sets but it requires an initial rough alignment to avoid local minima. Therefore the feature matching and registration can be used as a prior step for automatic ICP registration. We use this ICP result as a ground-truth registration and evaluate initial registration performances of descriptors in Section VI.

V. Multi-modal Datasets

To the best of our knowledge, there is no public multi-modal datasets for 3D acquisition of a common scene with different sensor technologies. Therefore we captured two scenes with the devices in Section 2 and tested the proposed multi-modal data registration framework on the datasets.

A. Gate scene

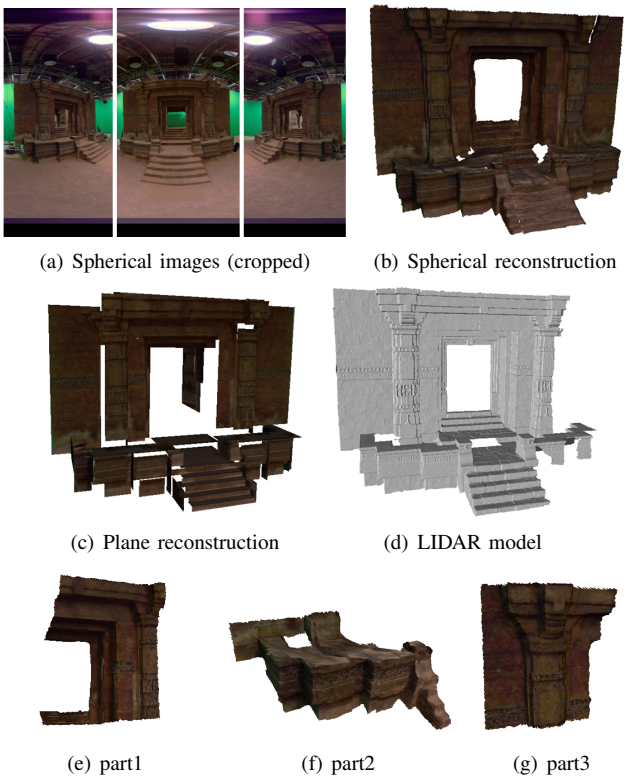
The Gate scene was captured in a indoor film set as multiple spherical stereo pairs by the Spheron and also as point clouds with a LIDAR scanner. The Gate model has a width of 9m and height of 6m. Three pairs of spherical stereo pairs were captured with a baseline of 60cm and resolution of 12574 $\times$5658. Dense 3D geometry of the scene was reconstructed using the PDE-based reconstruction method [8]. For testing proxy model registration, a plane-based model was reconstructed by plane fitting method from the dense reconstruction. Figure 2 (a) shows original spherical images of the scene and Fig. 2 (b)-(d) show their 3D reconstructions from images and 3D model from the LIDAR scans. We also took parts from the spherical reconstruction in order to verify the performance of partial model registration. Fig. 2 (e)-(g) show parts of the spherical reconstruction.

As can be see in Fig. 2 (d), the LIDAR model is incomplete due to occlusion. For evaluation against the LIDAR data we only consider the overlapping parts of the image-based reconstruction. We set the support radius R_n and R_s as 10cm for both surface normal calculation and 3D descriptor computation. Figure 3 shows the detected keypoints by the 3D Kanade-Tomasi detector [16]. The keypoints are relatively evenly distributed over the model and located on geometrically distinctive points in local regions. For the plane-based reconstruction, we sub-sampled the planes to get enough points to calculate normal vectors. However, the number of detected keypoints is still small because many planes are not connected or close enough to each other.

B. Cathedral scene

The "Cathedral" scene was captured in an outdoor environment. The scene is composed of one main building and surrounding open areas. The facade of the main building has a width of 30m and height of 20m, and has a complex structure with many self-occlusions and complicated details such as sculptures.

The main building was scanned as multiple spherical stereo image pairs with the spherical line-scan camera at three points. The resolution of each spherical image was 6284$\times$2794 and we used the same dense reconstruction and plane proxy model reconstruction techniques as the Gate model. We also took 92 still photos of the main building with a normal digital camera. The resolution of each photo was 2272$\times$1704 and we used the Bundler [13] for camera pose estimation and the PMVS algorithm [6] to reconstruct a dense point cloud. Finally we scanned the main building with a LIDAR scanner at 7 points and also measured around 50 points with a Total station to get a ground-truth model. We manually registered all LIDAR scans to the reference points from the total station and generated an integrated

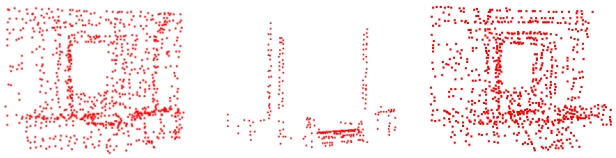


(a) Spherical images (cropped) (b) Spherical reconstruction

(c) Plane reconstruction (d) LIDAR model

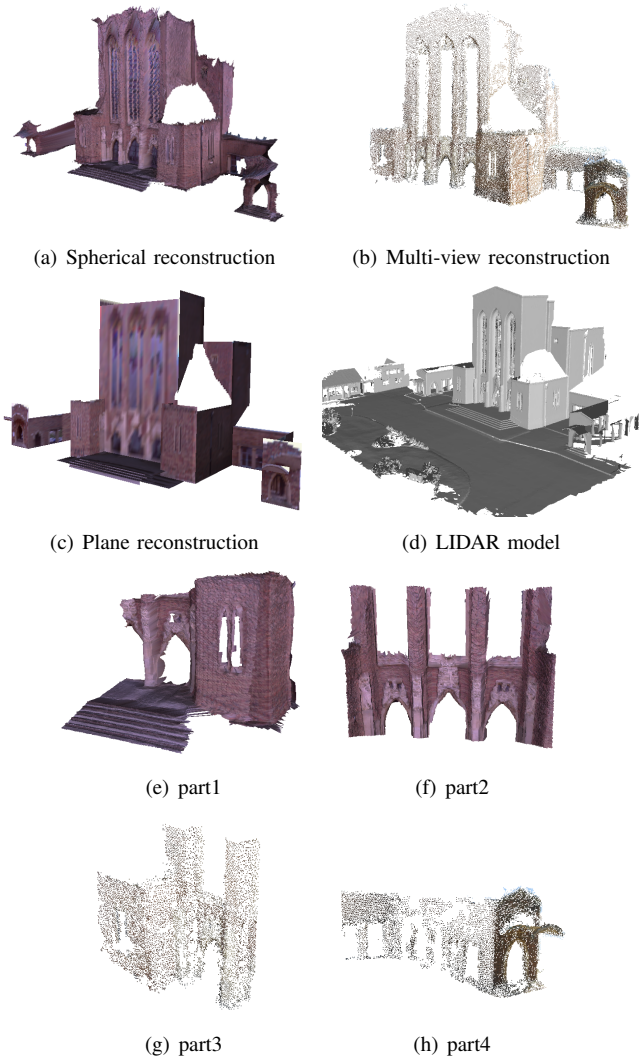
(e) part1 (f) part2 (g) part3

Figure 2. Gate dataset



(a) Spherical (885) (b) Plane (263) (c) LIDAR (844)

Figure 3. Point clouds and detected keypoints for Gate set (Number of detected keypoints)



(a) Spherical reconstruction (b) Multi-view reconstruction

(c) Plane reconstruction (d) LIDAR model

(e) part1 (f) part2

(g) part3 (h) part4

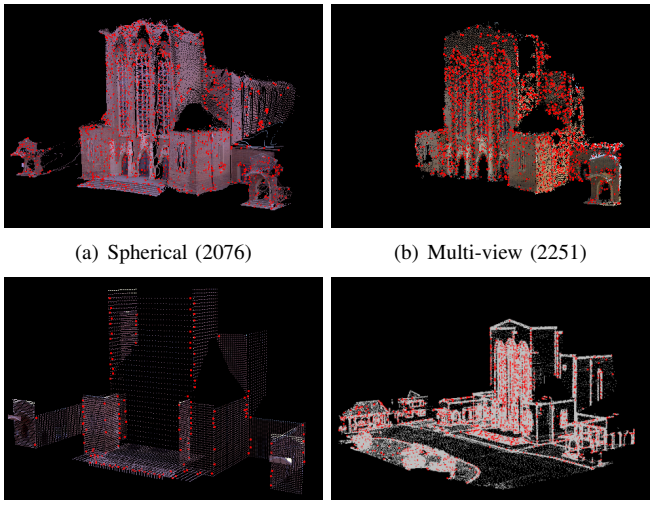
Figure 4. Cathedral dataset

mesh model. Examples of the original inputs are briefly shown in Fig.1, and Fig. 4 (a)-(d) show the 3D models from various input sources. Figure 4 (e)-(h) show partial reconstructions. Figure 4 (e) and (f) are from spherical reconstruction and we removed window and door regions because both spherical reconstruction and LIDAR model have errors in those regions. Figure 4 (g) and (h) are reconstruction from multiple photos.

The scale of this Cathedral scene is bigger than the Gate model. Therefore we set the support radius as 20cm. Figure 5 shows detected keypoints for the common region.

VI. Evaluation of Descriptors for Registration

In this section, we evaluate registration performances of the 3D feature descriptors introduced in Section IV.A on the datasets in Section V. The proposed framework was implemented based on the open source Point Cloud Library [11].



(a) Spherical (2076) (b) Multi-view (2251)

(c) Plane (153) (d) LIDAR (2562)

Figure 5. Point clouds and detected keypoints for Cathedral set (Number of detected keypoints)

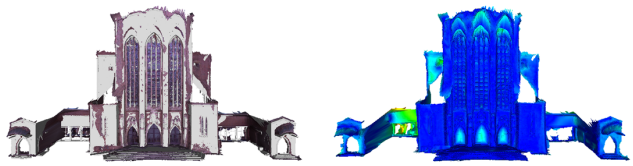
A. Registration to LIDAR sets

In this experiment, we assume the LIDAR model as the reference model and tried to register all other reconstructed models to the LIDAR model. For objective performance evaluation, we generated a ground-truth registration by manual registration followed by the ICP refinement [1]. Figure 6 shows the ground-truth registration result and the error map. We measured Hausdorff distance and mapped the distance in the range of 0-5m to a Blue-Red colour range. It is important to note that even the ground-truth registration has errors between models because those models are from different sources and have geometrical errors from capture and reconstruction process.

For the registration evaluation, we measured RMS error to the ground-truth registration instead of the Hausdorff distance. Figure 7 shows RMS error to the ground-truth in registration for different sets and descriptors.

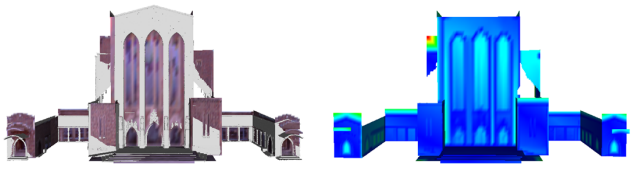
In the experiment with the Gate sets, all descriptors could register all test sets with errors in the range of 1m to the ground truth. The SHOT descriptor shows particularly good performance for the dense reconstruction model but poor for the proxy model registration. The FPFH descriptor generally shows more stable performances than others. Figure 8 shows registration results of the part 1-3 to the LIDAR model. However, this level of registration errors can be refined by the ICP algorithm. In terms of computational load, FPFH is normally 2-3 times faster in computing and matching descriptors than other descriptors.

The Cathedral sets are from a large scale outdoor scene which can induce more errors in capture and reconstruction. In the experiment with the Cathedral sets, we can see that some results show very high RMS errors in registration. Most cases under 3m RMS error range could be refined by



(a) Spherical

(b) Multi-view

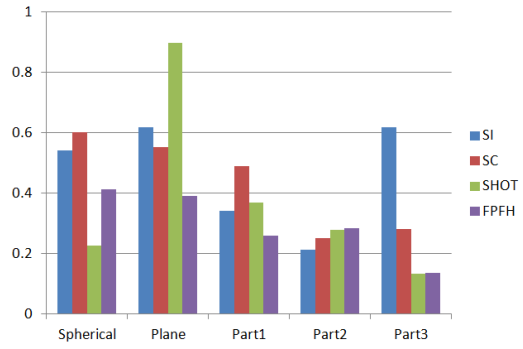


(c) Plane

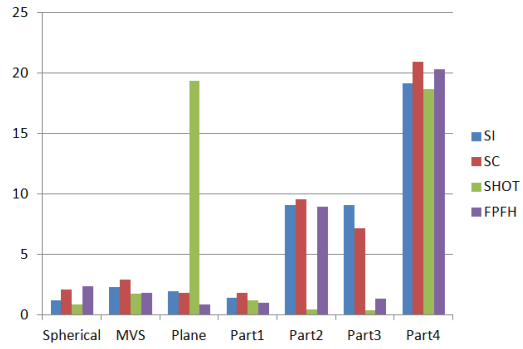
Figure 6. Ground-truth registration (Left: Registration result. Right: Error map)

the ICP in Fig. 7 (b). Therefore we refer the cases as successful and cases over 3m RMS error as failure in registration. Figure 9 shows examples of successful registration as their error maps and Fig. 10 shows the failed cases which cannot be refined by the ICP algorithm. Generally SHOT and FPFH show better performances in registration. In the registration of the plane model, only the SHOT descriptor failed in registration. The plane model has a relatively small number of keypoints and most of keypoints on the step of the scene were matched to windows in the LIDAR model. The SHOT descriptor also showed the worst performance in the Gate plane model registration. However, SHOT could successfully register the Part2 model while all other descriptors failed. Actually other descriptors registered the Part2 model to similar locations, but it is placed upside down as shown in Fig. 10 (b) because the Part2 model is pseudo-symmetric horizontally and vertically. The Part3 model is from multi-view images which is much coarser and noisier. SI and SC failed in registration.

Unfortunately, none of the descriptors could register the Part4 model. Most keypoints in the Part4 were extracted from the side wall, but relatively few keypoints were detected from the side wall in the LIDAR model. Keypoints on the side wall in Part4 were matched to the keypoints on the frontal windows in the LIDAR model. From the failed cases in Fig. 10, we can see that the errors are affected by the features on the frontal windows of the LIDAR model which were induced by noise from scanning transparent or reflective regions. Those erroneous features dominated matchings in the failure cases.

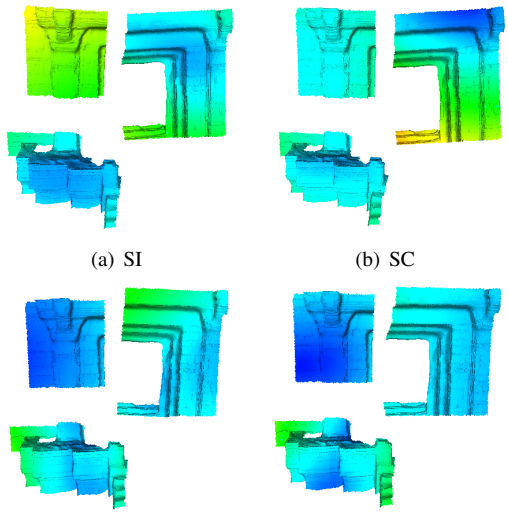


(a) Gate



(b) Cathedral

Figure 7. RMS error in registration



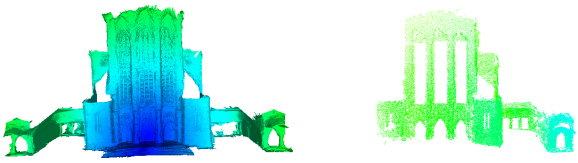
(a) SI (b) SC

(c) SHOT (d) FPFH

Figure 8. Registration error maps of partial Gate models (Distance in the range 0-1m are mapped to a Blue-Red colour range)

B. *Cross-modal registration*

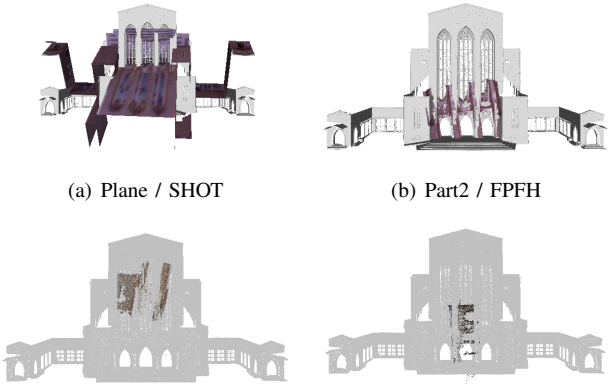
In this experiment, cross-evaluation is performed for data from all modalities against all others. We selected the same part of the Cathedral model from all reconstructions as shown in Fig. 11, and registered them to the other full



(a) Spherical / SC (b) Multi-view / SI

(c) Plane / FPFH (d) Part1-3 / SHOT

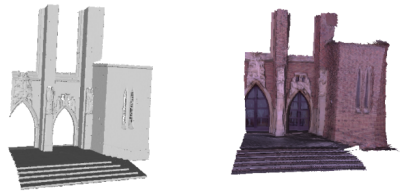
Figure 9. Registration error maps of Cathedral model (Successful cases. Distance in the range 0-5m are mapped to a Blue-Red colour range)



(a) Plane / SHOT (b) Part2 / FPFH

(c) Part3 / SI (d) Part4 / SC

Figure 10. Failure cases in registration (Cathedral models)



(a) LIDAR-part (b) Spherical-part

(c) MVS-part (d) Plane-part

Figure 11. Partial models for cross-modal registration

Table I
RMS ERROR IN CROSS-MODAL REGISTRATION (FPFH)

FPFH	LIDAR	Spherical	MVS	Plane
LIDAR-part		1.080	0.897	8.476
Spherical-part	1.531		0.853	10.015
MVS-part	7.801	0.440		13.608
Plane-part	6.924	4.484	16.846	

Table II
RMS ERROR IN CROSS-MODAL REGISTRATION (SHOT)

SHOT	LIDAR	Spherical	MVS	Plane
LIDAR-part		0.449	0.651	8.583
Spherical-part	0.986		8.879	16.025
MVS-part	10.346	1.015		12.670
Plane-part	9.058	12.840	17.086	

datasets. Registration error was measured as RMS error against the ground-truth registration generated by manual alignment followed by the ICP.

Table I and II show the RMS errors in registration using the FPFH and SHOT descriptors, respectively. In both tables, all part registrations related to the plane models failed in registration due to lack of geometrical features. Considering that the full plane model could be registered in the previous evaluation, we can conclude that simple proxy models require more feature points over a wide area to be registered. We can also observe that both descriptors failed in registration of the MVS-part to the LIDAR set. The MVS-part model does not have the lower part as shown in Fig. 11 (b) and the corresponding part is rotationally pseudo symmetric in the full dataset. They were 180° rotated and mapped to the left part of the scene as it happened for the Part2 model in the previous section. Geometrical symmetry is one of main causes of failure in registration.

The FPFH descriptor shows relatively stable performances in the cross-model registration. The SHOT descriptor could register datasets more accurate than the FPFH descriptor, but it sometimes failed in certain cases. This coincides with the results in the previous experiment.

VII. Discussion and Conclusions

In this work, we proposed a framework for 2D and 3D multi-modal data registration and evaluated registration performances of various 3D feature descriptors. 3D LIDAR scan data, 3D proxy models, spherical images and multiple photographs were considered in the framework. For 2D data registration, 3D structures are reconstructed from the 2D images using stereo or multi-view methods, then they are registered in a 3D domain by 3D feature detection and registration. As a result, evaluation was on multi-modal 3D to 3D data registration where 3D data may be extracted from either 2D images or direct 3D measurement.

SC, SI, SHOT and FPFH descriptors were evaluated for various test sets. The performances of most descriptors are acceptable for indoor datasets with stable material, lighting

condition and background, but FPFH works slightly better in terms of accuracy and speed. For outdoor scenes with a more variable environment, SHOT and FPFH show better performance. The SHOT descriptor is good at registering dense reconstructions with high accuracy. However, it is poor at proxy model registration and sometimes shows unstable behaviours. The FPFH descriptor failed in pseudo-symmetric structure registration, but it shows relatively stable performances in general registration.

For a simple planar proxy model or symmetric structure registration, a wider area with enough feature points should be considered because local features have limited information about the geometry. Considering other properties of the data such as colour information or geodesic distance between feature points can be helpful for successful registration if they are available.

The feature detector and all descriptors evaluated exhibit problems with errors in reconstruction resulting from transparent and reflective surfaces. This is a fundamental problem from capture devices and reconstruction methods, but it is still required to develop a feature detector and descriptor which produce feature sets consistent in their locations and descriptions regardless of local geometric sampling, errors and noise.

This work is still in progress rather than a definitive evaluation. Our future works will include testing more state-of-the-art descriptors for matching between various multi-modal datasets and analysing relationships between capture errors, reconstruction errors, feature detectors and descriptors to influence the whole registration framework. Novel detectors and descriptors may be required to achieve robust matching and registration of multi-modal data.

ACKNOWLEDGMENT

This research was supported by the UK TSB project SyMMM and the European Commission, FP7 IMPART project (grant agreement No 316564).

REFERENCES

[1] P. Besl and N. McKay. A method for registration of 3-d shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[2] H. Dutagaci, C. P. Cheung, and A. Godil. Evaluation of 3d interest point detection techniques via human-generated ground truth. *The Visual Computer*, 28(9):901–917, 2012.

[3] D. Eberly. Least Squares Fitting of Data. http://www.geometrictools.com/Documentation/LeastSquaresFitting.pdf, 2008. [Online; accessed 01-Mar-2013].

[4] A. Frome, D. Huber, R. Kolluri, T. Bulow, and J. Malik. Recognizing objects in range data using regional point descriptors. In *Proc. ECCV*, 2004.

[5] Y. Furukawa, B. Curless, S. Seitz, and R. Szeliski. Manhattan-world stereo. In *Proc. CVPR*, 2009.

[6] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2010.

[7] A. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):433 – 449, 1999.

[8] H. Kim and A. Hilton. 3d modelling of static environments using multiple spherical stereo. In *Proc. RMLE workshop in ECCV*, 2010.

[9] M. Restrepo and J. Mundy. An evaluation of local shape descriptors in probabilistic volumetric scenes. In *Proc. BMVC*, pages 46.1–46.11, 2012.

[10] R. B. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (fpfh) for 3d registration. In *Proc. ICRA*, pages 3212–3217, 2009.

[11] R. B. Rusu and S. Cousins. 3D is here: Point Cloud Library (PCL). In *Proc. ICRA*, 2011.

[12] T. Sattler, B. Leibe, and L. Kobbelt. Improving image-based localization by active correspondence search. In *Proc. ECCV*, 2012.

[13] N. Snavely, S. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3d. In *Proc. ACM SIGGRAPH*, pages 835–846, 2006.

[14] I. Stamos, L. Liu, C. Chen, G. Wolberg, G. Yu, and S. Zokai. Integrating automated range registration with multiview geometry for the photorealistic modeling of large-scale scenes. *International Journal of Computer Vision*, 78(2-3):237–260, 2008.

[15] J. Sun, M. Ovsjanikov, and L. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *Proc. SGP*, pages 1383–1392, 2009.

[16] C. Tomasi and T. Kanade. Detection and tracking of point features. *Pattern Recognition*, 37:165–168, 2004.

[17] F. Tombari, S. Salti, and L. Di Stefano. Unique signatures of histograms for local surface description. In *Proc. ECCV*, pages 356–369, 2010.

[18] F. Tombari, S. Salti, and L. Di Stefano. Performance evaluation of 3d keypoint detectors. *International Journal of Computer Vision*, 102:198–220, 2013.