International Journal of Pure and Applied Mathematics
Volume 120 No. 5 2018, 1117-1125
ISSN: 1314-3395 (on-line version)
url: http://www.acadpubl.eu/hub/
Special Issue
http://www.acadpubl.eu/hub/

3D FEATURE DESCRIPTORS AND IMAGE RECONSTRUCTION TECHNIQUES: A REVIEW

E.T.Jaba Jasphin[a], C.Sheeba Joice[b]
[a]Assistant professor,
Department of Electronics and Communication Engineering,
Saveetha Engineering College, Chennai
jabajasphin@saveetha.ac.in
[b]Professor,
Department of Electronics and Communication Engineering,
Saveetha Engineering College, Chennai
sheebajoice@saveetha.ac.in

ABSTRACT

The image reconstruction is being widely used in many medical and engineering applications because of the low cost 3D scanners. 3D feature descriptor play a major role in image reconstruction. The Objective of this paper is to compare different 3D feature descriptors and image reconstruction techniques. 3D image reconstruction techniques reconstruct 3D shapes from images captured using depth sensor and label the object, based on the reconstructed 3D shapes. From the set of images 3D models are created, refers to 3D reconstruction. Normally, 3D scenes are captured to get 2D images which is the reverse processes of 3D reconstruction. This paper gives review about different feature descriptors and techniques used for 3D image reconstruction.

Keywords:3D reconstruction, Sensors, RGB-D, Kinect, GPU, Local Feature descriptors, camera, datasets.

1. INTRODUCTION

Feature point identification is the basic step in image reconstruction. Thispaper compares different3D feature descriptors and methods to reconstruct the 3D shapes from RGB-D images captured from unstructured environments.Most of image processing algorithms works well in controlled environments.Also the depth information is lost when the image is captured using 2D cameras.So we need an efficient image reconstruction algorithm.

Section 2 comparesthe performance of 3D Local Feature descriptors based on different data sets.Section 3 explains different 3D image reconstruction techniques with different cameras and datasets.

2. 3D LOCAL FEATURE DESCRIPTORS

The local feature descriptor must be able to describe local image region so that it can be distinguished from other regions and it can be matched effectively with those images that are similar. After a feature has been detected using a local feature detector the descriptor is built. Other terms used for local features are interest points, interest regions and key points.

Different 3D Local Feature descriptors are compared by Y. Guo et al. (2016)[1]. With key point detection and feature description, Local feature descriptors are calculated. First key points with high information content are identified. Then local geometric information is extracted near the key points. To get point to point feature correspondence, feature descriptors of two surfaces are matched. Important attributes for 3D local feature descriptors are robustness and descriptiveness. If a feature descriptor is capable to distinguish between two local surfaces it is descriptive. If a feature descriptor is not sensitive to a number of disturbancesit is robust. Comparison of 3D local feature descriptors based on descriptiveness is listed in Table 1.

Table1: Comparison of 3D local feature descriptors based on descriptiveness

3D Local Feature Descriptors	**Descriptiveness**	**Applications**
Fast Point Feature Histogram (FPFH)	i)Descriptive, computationally efficient and light weight method with small number of points ii) FPFH requires less memory	Used in both Real time and space crucial applications
Signature of Histogram of Orientations (SHOT)	It has descriptiveness and computational efficiency with large number of points.	Used in Real time systems
Rotational Projection Statistics (RoPS)	i)It provides good descriptiveness at the cost of high memory requirement ii) RoPS produce very stable performance.	Used in Space crucial applications

2.1. Histogram Based Descriptors

Comparison of 3D local feature descriptors based on datasets are listed in Table 2. Some of the spatial distribution histogram based descriptors are 3D Shape Context (3DSC), Spin image(SI),Rotational Projection Statistics(RoPS), Tri-Spin-Image (TriSI), Unique Shape Context (USC), 3D Tensor descriptors, etc. 3D local feature descriptors depending on geometric attributes are Local Surface Patch(LSP), 1D histogram, Fast Point Feature Histogram (FPFH), SHOT, Point Feature Histogram(PFH), Variable-Dimensional Local Shape Descriptors (VD-LSD) etc. For large dataset models, Scalability is best for USC,TriSI and 3DSC. Descriptiveness is low, computational and storage cost are high for USC and 3DSC. For applications on large datasets TriSI is the best choice.

Table2: Comparison of 3D local feature descriptors based on datasets

Performance	**Datasets**	**3D local feature descriptors**[1]
High-resolution Datasets.	Laser scanner, Random views, Retrieval, and 2.5D views	FPFH, RoPS and PFH provides best performance.
Low resolution datasets.	Dense stereo, Space time, Kinect, and LIDAR	RoPS, USC, TriSI and PFH provides best performance.
Stability across datasets	Retrieval, Random views, Laser scanner, Space time, Kinect, LIDAR, Dense stereo, 2.5D views	i)RoPS is very stable. ii)FPFH, SI, TriSI and PFH are Stable.
Overall performance using all the datasets.	Retrieval, Random views, Laser scanner, Space time, Kinect, LIDAR, Dense stereo, 2.5D views	i)FPFH, SHOT, TriSI and PFH provides good overall performance. ii)RoPS provides best overall performance.

2.2. Point Pair Features (PPFs)

L. Kiforenko et al. (2017) compared the point pair features (PPFs)[2], with local histogram features like SHOT, SI, Equivalent Circumference Surface Angle Descriptor (ECSAD), PFH/FPFH and 3DSC/USC. Other types of feature descriptors are 3D-SURF(3Dimentional

Speeded Up Robust Features) or SI-SIFT (Shape Index-Scale Invariant Feature Transform), Normal Aligned Radial Feature (NARF), Heat Kernel Signature (HKS) descriptor and Color Point Pair Features (CPPF). The Speeded Up Robust Features(SURF) became more robust and faster feature descriptor. SURF computes the Haar wavelet response of the feature point neighborhood. The SURF detector can be used as interest point detector which is rotation and scale invariant.

The point pair features use distance and angle as relationships between any two points. PPF outperformed other features for high resolution data. SHOT and USC showed good performance for noisier data. Towards noise FPFH is unstable and USC is more robust. PPF performance degrades faster under occlusion and clutter when compared with histogram features.

3. 3D RECONSTRUCTION

Here the different techniques used in 3D reconstruction is discussed. A. Garcia-Garcia et al. (2016) explained about the robust system in [3], in which objects can be recognized in poor light conditions and scenes with occlusion. Running on a mobile GPU (Graphical processing Unit) computing platform exhibit a reasonable performance, and it consumes less power. Microsoft Kinect or Primesense carmine provide color and depth (RGB-D) data streams for the acquisition system. The mobile GPU computing platform (NVIDIA Jetson Tk1) recognize objects within 7 seconds.

The object recognition pipeline is divided Keypoint detection, Descriptor Extraction, Feature matching, Correspondence grouping, pose refinement and hypothesis verification stages. To detect keypoints, keypoint extraction and uniform sampling are used. The descriptor extraction used FPFH, 3DSC, USC, SHOT, CSHOT and RoPS. The search structures like k-d trees are often used because they are efficient. For correspondence grouping, Geometric Consistency Grouping (GCG) is implemented in Point Cloud Library (PCL). The Instance alignment or pose refinement uses Iterative Closest Point (ICP) algorithm and Global Hypothesis Verification (GHV) algorithm for Hypothesis verification.

K. Lu et al. (2015) used both view-based and model-based method for 3D Object retrieval [4]. The 3D objects are described by low level and high level features to explain the relationship among 3D objects called as model based method. The group of images taken from different directions represent the 3D objects is called view based method. Here both view and model-based are jointly used for 3D Object retrieval. An Object graph is created using Model based features. The relevance among 3D models is estimated by learning on the Joint View-Model graphs.

The Model based method preserve the 3D objects global spatial information. The View based method is highly discriminative and provide better retrieval of the 3D objects method than model based method. In view based method, the spatial relationship among different views is described when the camera array information is available.

Qu et al. (2017) proposed a RGBD salient object detection via deep fusion[5]. Here instead of giving raw pixel as an input to CNN (Convolutional Neural Networks), various flexible and interpretable saliency feature vectors are given as input. Thus the CNN predict more effectively by learning a combination of existing features. Then trained CNN is integrated with a superpixel-based Laplacian propagation framework. By exploiting the intrinsic structure of the input, a spatially consistent salient map is extracted.

Y. Gao et al. (2012) used camera constraint-free view 3D object retrieval algorithm [6]. Most approaches depend on their own camera array settings to capture views of 3D objects. To overcome camera array restriction, Camera Constraint-Free View (CCFV) based 3D object retrieval algorithm is used. Without camera constraint each object in an image captured from any direction is represented by a free set views. Cluster all query views to build the query model. A positive and negative matching model are trained separately using positive and negative matched samples for correct 3D object comparison.

A. Agudo et al. (2016) describes a real-time sequential method to recover the camera motion and the 3D shape of deformable objects from a calibrated monocular video simultaneously [7]. The Navier-Cauchy equations in 3D linear elasticity is used to model the time-varying shape per frame. These equations are embedded in an extended Kalman filter, resulting in sequential Bayesian estimation approach. The proposed approach performs data association over the whole sequence and can run in real time at frame rate for small maps. This approach is particularly relevant for medical imaging, where rich priors and accurate models are often available. This is also appropriate for robotic tasks involving the manipulation of non-rigid objects where an estimation in real time is mandatory.

By exploiting shading cues captured from an IR camera, G. Choe et al. (2017) used the method to refine geometry of 3D meshes depth camera[8]. With natural indoor illumination the IR images are robust because it filters undesired ambient light. In IR spectrum, natural objects in the visible spectrum with colorful textures have uniform albedo. To estimate the surface details, multiview information and 3D mesh from the Kinect fusion is used. On the mesh model this approach directly operates for geometry refinement.

By projecting a structured light pattern, an active range sensing is utilized in the Kinect I. To capture the images in dark environments Kinect II holds the IR projector and IR camera. For depth map acquisition Time-Of-Flight (ToF) technology is used. For depth measurement and capturing shading cues for high quality reconstruction IR camera of the Kinect is used. But for 3D reconstruction only the estimated depth map is used in Kinect fusion. The proposed method operates on the 3D mesh directly and the geometry refinement process is optimized. The result is a high quality mesh model. This approach is robust to indoor illumination. It also works well in natural lighting environments and dark rooms. For estimating initial geometry, the Kinect fusion provided in the Kinect SDK 1.7 and 2.0 is used.

S. Hadfield et al. (2017) used a standard stereo reconstruction with a wide range of classic top-down cues from urban scene understanding [9]. Cues which are reformulated includes recognizing concave, convex, occlusion boundaries, coplanar and collinear edges. The top-down cues reduce issues relating to the baseline (i.e. the separation of the two cameras). The accuracy of these matching and triangulation based systems is strongly limited by the baseline. The top-down approach use information about surface directions and types of edges present in the scene. This approach significantly improve the robustness of obtained reconstructions.

Haltakov et al. (2016) proposed a geodesic pixel neighborhoods for 2D and 3D scene understanding[10]. Here based on the concept of pixel neighborhoods two stage classification is used. Every pixel is classified based on its appearance in the first stage. This pixel neighborhood is defined by using the geodesic distance. This is able to capture both local image and more global object relations. The first stage is summarized by a voting histogram feature. This is given as the input for the second classifier. It is geodesically smoothed to get the final segmentation.

Salman H.Khan et al. (2015) proposed a model for labeling of scenes using RGBD Images [11]. The Conditional Random Field (CRF) model is used for labeling of indoor scenes by effectively utilizing the depth information. CRF defines local, pairwise and higher order interactions between image pixels. Local level interactions combine the energies from appearance, depth and geometry based cues. Pairwise interactions are learning the spatial discontinuity of object classes across the image. Higher order interactions treats smooth surfaces as cliques and all pixels on the surface to take the same label. This model uses both appearance and geometric information. The geometry of indoor surfaces was approximated using region growing algorithm for segmentation. This was combined with appearance based information. The learned boundaries was used to define the image spatial discontinuity. This method also captures long range interactions on the dominant planar surfaces by defining cliques. By using a single slack formulation of the rescaled margin cutting plane algorithm, the parameters of the model were learned.

Thomas Morwald et al. (2013) addressed advances in real time object tracking [12]. More accurate pose estimation and faster convergence can be achieved with confidence dependent variation together with iterative particle filtering. The fixed particle poses removes jitter and ensures convergence. These provides basis for tracking systems performing in the real world. The qualitative states of tracking are convergence, quality, loss and occlusion. If already the pose has found, the state of tracking is convergence. The quality gives confidence of the currently tracked pose and loss detects when the algorithm fails. Occlusion determines the degree of occlusion if only parts of the objects are visible.

The task is to find the position and orientation (pose) of an object in space, then projection of a geometric model together optionally with texture information compared to the current image. The measure from the comparison is minimized with respect to the pose by applying a Monte Carlo Particle Filter (MCPF). So these approach is based on Tracking State Detection (TSD), texture mapping, Pose recovery, online learning and model completeness. In online learning, these feature points and surface texture of the object are learned automatically.

A. Amamra et al (2014) proposed a GPU based real time RGBD Data filtering[13].An innovative adaptation of Kalman filtering schemeis proposed to improve the precision of Kinect as a real time RGBD capture device. The Microsoft Kinect (RGBD) sensor works in real time at the frequency of 30fps. Kinect outputs RGB stream, depth stream and audio stream. Kinect sensor includes an IR projector, an IR camera and RGB camera. An IR projector projects the IR pattern on the scene, an IR camera captures the reflected light of the projected pattern and RGB camera works as an ordinary color camera. The captured disparity map is projected on a set of discrete parallel planes using kinect. To clean the raw depth measurement output by Kinect, GPU implements the parallel design.

The depth data are naturally rough and noisy. Filtering approaches removes noisy data (outliers), clean the useful zones (inliers) and preserve the edges. To accurately stabilize the Kinect output over time Kalman filter is used. The kalman filter works based on a recursive prediction of the next state and its correction. With no additional history of systems behavior, it runs in real time because, it uses the present measurement and the previously estimated state as input.

Sanchezet al (2016) proposed a method to solve the problem in place recognition using a portable single 3D sensor[14]. It identifies and tracks the position even it is the known area or significant changes in the environment occurs. Here place recognition is considered as a classification problem and considered only controllable area for efficient search space reduction. By using temporal consistency with respect to relative tracker, classification hypothesis are discarded. The compact classifier scales with the map size is used.The user is tracked continuously following localization by manipulating the known environment using efficient data structure.By selecting geometrically stable points, filtering the outliers and integrating the relative tracker robust results are achieved.

Boukamcha et al (2017) presented a robust method for real object 3D shape reconstruction[15]. The composite format is proposed for projecting the light pattern on the object. It combines primary color coded channels into one composite format which reduces the number of patterns. For both calibration and construction phase, spatial and temporal intensity variation is used. High quality depth map is obtained without complex calculations from the linear light reflected by the shape of the object. The requirement is a digital camera, flashlight and mask of pattern. Compared with structured light scanning system calibration procedures the proposed system hardware cost is minimized.

Panet al (2016) proposed a dense 3D reconstruction by combining both depth and RGB information[16]. Mostly in the depth methods the camera object distance are constrained from 0.4m to 4m. This 3D reconstruction method is also used,ifthe camera object distance less than 0.4m.When the camera fails to obtain the correct depth information, this method uses RGB information with depth to refine the reconstruction results. When the camera is close to the object, RGB information along with feature detection and triangulation method is used. This provides accurate camera poses and 3D points.

Hofer et al (2017) proposed an efficient 3D scene abstraction using line segments[17]. From different images the geometric constraints are used to match the 2D line segments. Graph clustering problem is used as a reconstruction procedure. With a small amount of data and very short time,significant amount of 3D information about a scene can be encoded in contrast to point-clouds.This method is an alternative for all scenarios in which 3D edge information is preferred over a point-cloud or surface reconstruction.

Schops et al (2017) proposed a large-scale outdoor scenes 3D reconstruction on a mobile device[18]. The fisheye camera in the device enables a user to reconstruct large scenes in few

minutes. To compute depth maps the device's GPU is used. To detect and discard unreliable depth measurements a set of filtering steps are proposed.This method enable live reconstruction of large outdoor scenes on a mobile device.
The comparison of different 3D reconstruction methods are given in Table 3.

Table3: Comparison of 3D Reconstruction methods

Method Used	**Sensors**	**Computing platform**	**Data sets/ Database used**
3D object recognition pipeline on mobile GPGPU (General Purpose computation on GPUs)[3].	Low cost 3D sensors like Microsoft Kinect or primesense carmine	Mobile GPU computing platform (NVIDIA Jetson Tk1)	Not Used.
3D Object retrieval used both view-based and model-based method[4].	Not Used	PC with i52.4GHz CPU and 16GB memory	National Taiwan University 3D Model database (NTU) , Princeton Shape Benchmark (PSB) and Shape Retrieval Content (SHREC)
RGBD Salient Object Detection via Deep Fusion[5].	Not Used	MATLAB	NLPR RGBD salient object detection dataset, the NJUDS2000 stereo dataset, and the LFSD dataset.
3D object retrieval using Camera Constraint-Free view (CCFV) [6].	Not Used	Not Mentioned	NTU 3-D model database and the ETH database
Real-time 3D reconstruction of non-rigid shapes with a single moving camera [7].	Hand-held camera	Visual rigid SLAM software.	Not mentioned
Refining Geometry from Depth Sensors using IR Shading Images[8].	Kinect I and II Kinect I is based on a structured-light technique. Kinect II is based on a time-of-flight technology.	Not Mentioned	Not Used
Standard stereo reconstruction with a wide range of classic top-down cues from urban scene understanding [9].	Not Used	MATLAB Single Intel Sandy Bridge core at 2.4 GHz and required a maximum of 4 GB of memory.	Middlebury 2014 and KITTI datasets

Geodesic pixel neighborhoods for 2D and 3Dscene understanding[10].	Not used	Desktop machine equipped with two Intel Xeon X5690 processors with 6 cores each running at 3.46 GHz.	CamVid, MSRC-21, StanfordBackground, eTRIMS, Daimler Urban, KITTI Segmentation Dataset.
Labeling of scenes using RGBD Images[11].	Microsoft Kinect	Matlabrunning on single core, thread.	NYU-Depth and SUN3D
Real time object tracking [12].	Logitech Webcam Pro 9000	NVIDIA GeForce GTX 285 GPU	Not Used
Real Time RGBD data filtering[13].	Microsoft Kinect	NVIDIA GeForce 2GB GTX 680 GPU	Not Used
Localization and trackingportablereal-time 3D sensors[14]	Velodyne HDL-32E sensor	Computer with an Intel Xeon CPU @ 2.80 GHz with 8 GB of RAM and a 64 bit operating system.	Datasets acquired with a LIDAR scanner
Dense 3D reconstruction combining depth and RGB information[16]	RGB-D Kinect camera	Desktop PC with an Intel quad core 3.2 GHz processor and a GTX 660 graphics card with 4 GB of RAM	Not Used
Efficient 3D scene abstraction using line segments[17].	Not used	Desktop machine equipped with an Intel Core i5 CPU (4 ×3.4 GHz), 12 GB of main memory, and an nVidia GeForce GTX 580 Ti GPU.	Publicly available datasets with ground truthHerz-Jesu-P8 dataset (8 images), Timberframe dataset (240images),Castle-P30 (30 images)
Large-scale outdoor 3D reconstruction on a mobile device[18].	Tango Tablet's depth sensor	A standard desktop PC with an Intel Core i7-4770K processor (3.5 GHz) and an Nvidia GeForce 780 GTX graphics card	ICL-NUIM dataset

4. CONCLUSION

In this paper, we presented the need for 3D object reconstruction. We have done the comparative analysis of different 3D local feature descriptors based on descriptiveness and based on datasets. In 3D reconstruction, the methods used in different conditions were analyzed. This analysis will be used for further implementing suitable 3D reconstruction technique in engineering, agricultural and medical applications.

5. REFERENCES

[1] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, J. Wan, and N. M. Kwok, “A Comprehensive Performance Evaluation of 3D Local Feature Descriptors,” *Int J Comput Vis*, vol. 116, no. 1, pp. 66–89, Jan. 2016.

[2] L. Kiforenko, B. Drost, F. Tombari, N. Krüger, and A. Glent Buch, “A performance evaluation of point pair features,” *Computer Vision and Image Understanding*, Sep. 2017.

[3] A. Amamra and N. Aouf, “GPU-based real-time RGBD data filtering,” *J Real-Time Image Proc*, pp. 1–18, Sep. 2014.

[4] C. Sánchez, P. Taddei, S. Ceriani, E. Wolfart, and V. Sequeira, “Localization and tracking in known large environments using portable real-time 3D sensors,” *Computer Vision and Image Understanding*, vol. 149, pp. 197–208, Aug. 2016.

[5] H. Boukamcha, A. Ben Amara, F. Smach, and M. Atri, “Robust technique for 3D shape reconstruction,” *Journal of Computational Science*, vol. 21, pp. 333–339, Jul. 2017.

[6] H. Pan *et al.*, “Dense 3D reconstruction combining depth and RGB information,” *Neurocomputing*, vol. 175, pp. 644–651, Jan. 2016.

[7] M. Hofer, M. Maurer, and H. Bischof, “Efficient 3D scene abstraction using line segments,” *Computer Vision and Image Understanding*, vol. 157, pp. 167–178, Apr. 2017.

[8] T. Schöps, T. Sattler, C. Häne, and M. Pollefeys, “Large-scale outdoor 3D reconstruction on a mobile device,” *Computer Vision and Image Understanding*, vol. 157, pp. 151–166, Apr. 2017.

[9] A. Garcia-Garcia, S. Orts-Escolano, J. Garcia-Rodriguez, and M. Cazorla, “Interactive 3D object recognition pipeline on mobile GPGPU computing platforms using low-cost RGB-D sensors,” *J Real-Time Image Proc*, pp. 1–20, Jun. 2016.

[10] K. Lu, N. He, J. Xue, J. Dong, and L. Shao, “Learning View-Model Joint Relevance for 3D Object Retrieval,” *IEEE Transactions on Image Processing*, vol. 24, no. 5, pp. 1449–1459, May 2015.

[11] L. Qu, S. He, J. Zhang, J. Tian, Y. Tang, and Q. Yang, “RGBD Salient Object Detection via Deep Fusion,” *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2274–2285, May 2017.

[12] Y. Gao *et al.*, “Camera Constraint-Free View-Based 3-D Object Retrieval,” *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2269–2281, Apr. 2012.

[13] A. Agudo, F. Moreno-Noguer, B. Calvo, and J. M. M. Montiel, “Real-time 3D reconstruction of non-rigid shapes with a single moving camera,” *Computer Vision and Image Understanding*, vol. 153, no. Supplement C, pp. 37–54, Dec. 2016.

[14] G. Choe, J. Park, Y.-W. Tai, and I. S. Kweon, “Refining Geometry from Depth Sensors using IR Shading Images,” *Int J Comput Vis*, vol. 122, no. 1, pp. 1–16, Mar. 2017.

[15] S. Hadfield, K. Lebeda, and R. Bowden, “Stereo reconstruction using top-down cues,” *Computer Vision and Image Understanding*, vol. 157, no. Supplement C, pp. 206–222, Apr. 2017.

[16] V. Haltakov, C. Unger, and S. Ilic, “Geodesic pixel neighborhoods for 2D and 3D scene understanding,” *Computer Vision and Image Understanding*, vol. 148, pp. 164–180, Jul. 2016.

[17] S. H. Khan, M. Bennamoun, F. Sohel, R. Togneri, and I. Naseem, “Integrating Geometrical Context for Semantic Labeling of Indoor Scenes using RGBD Images,” *International Journal of Computer Vision*, vol. 117, no. 1, pp. 1–20, Mar. 2016.

[18] T. Mörwald, J. Prankl, M. Zillich, and M. Vincze, “Advances in real-time object tracking: Extensions for robust object tracking with a Monte Carlo particle filter,” *Journal of Real-Time Image Processing*, vol. 10, no. 4, pp. 683–697, Dec. 2015.

6. ABOUT THE AUTHORS

E.T.Jaba Jasphin received her M.E degree with the specialization Digital Communication and Networking in 2006. She is currently a research scholar and working as Assistant Professor in ECE Department. Her research interest includes computer vision, machine learning and Image processing. She is an IETE member.

Dr.C.Sheeba Joice received her Ph.D degree from Anna University in 2012. She is currently working as Professor in the Department of ECE, Saveetha Engineering College. She is the Former Honorary Secretary of IETE Chennai Centre. She published papers in reputed journals. Her research interest includes Embedded systems and Image processing.