An Evaluation of Local Shape Descriptors in Probabilistic Volumetric Scenes

Maria I. Restrepo
http://vision.lems.brown.edu/students/restrepo
Joseph L. Mundy
http://vision.lems.brown.edu/faculty/mundy

LEMS Laboratory
School of Engineering
Brown University
Providence, RI. USA

Motivation

Understanding the three-dimensional world from images is a long standing goal in computer vision, with numerous applications in the areas of robotics, autonomous navigation, city mapping and surveillance. Multi-view stereo can be thought of as the initial step towards this goal and it has been widely studied by the scientific community. However, few of the available methods can perform image-based 3-d modeling of outdoor, crowded, large scale scenes, where accurate modeling is difficult due to severe occlusion, varying illumination conditions, the presence of highly reflective surfaces and sensor errors. In this realm, probabilistic volumetric methods offer a dense representation for the solution of the multi-view stereo problem, modeling explicitly the scene's uncertainty. In particular, Pollard and Mundy [5] propose a volumetric Bayesian framework that is robust to large variations in appearance from a video sequence. This model is extended into a continuous framework by Crispell *et al*. [1] allowing for an octree subdivision of space. Finally, a recent GPU implementation [4] of this model, capable of processing one HD-resolution frame per second, guarantees scalability to large urban scenes and encourages its application to higher level 3-d computer vision tasks.

Inspired by the growing number of applications of the probabilistic volumetric model (PVM) to 3-d scene modeling and understanding, this work aims to provide the first evaluation of the performance of several local shape descriptors extracted from the PVM in terms of accuracy for object classification. Descriptors based on local histograms are of particular interest as they are the most popular and most successful for many image indexing applications. While the performance of many 3-d shape descriptors has been studied in point cloud data (from range sensors, or computer generated meshes) it is unclear that their descriptiveness and robustness to noise successfully extends to the diffuse surface probability distributions of the PVM. Surfaces can be poorly localized due to the presence of highly reflective materials, large regions of constant intensity and challenging illumination conditions in the input image data. This work takes a step towards the characterization of the PVM as a new representation for 3-d scene understanding that provides a dense continuous representation of scene geometry despite of all these sources of ambiguity.

Evaluation Framework

This papers evaluates four popular shape descriptors, namely Spin Images (SI) [3], 3-d Shape Contexts (SC) [2] , Signatures of Histograms of OrienTations (SHOT) [8] and Fast Point Feature Histogram (FPFH) [7]. For evaluation, the proposed framework starts by learning the probabilistic models for 17 urban scenes from publicly available aerial imagery collected from a helicopter flying around Providence, RI USA. After learning the PVM, surface normals are computed by convolving three-dimensional Gaussian derivative kernels with the volumetric surface probabilities. Surface normals and their locations are used to compute the local 3-d descriptors. Finally, the descriptors are used to learn *Bag of Words* models for five object categories: planes, cars, houses, buildings and parking lots. During learning and categorization, the objects are manually segmented and labeled using the bounding boxes provided in [6]. The *Bag of Words* framework follows common practices, where a common vocabulary is learned for all categories using k-means clustering. Naive Bayes, SVM and Nearest Neighbors classifiers are compared during the classification stage. To ensure robustness of the evaluation, the results are reported for various vocabulary sizes, descriptor parameters and classifiers across multiple splits of the train/test data. A special effort is made to report all parameters used through out this work, such that this evaluation can provide precise guidance to future works using the PVM. An overview of the proposed pipeline is presented in Figure 1.

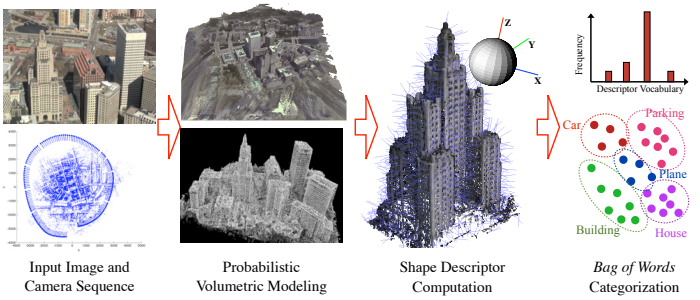

Figure 1: Framework overview. First, an input image and camera sequence are used to learn the PVM. Then, surface normals and shape descriptors are computed at highly likely surface locations. Finally, a *Bag of Words* model is used to learn and classify 5 object categories.

Results

Some of the results presented in this paper are shown in Figure 2. Under the different test scenarios, the FPFH obtained high recall while having the advantage of being compact and fast to compute. Spin Images underperformed, in particular when recognizing buildings. The SVM classifier was the more effective classifier. Shape contexts achieve adequate classification performance, but have very high storage requirements, posing run-time challenges during batch k-means. The results indicate that distribution-based descriptors effectively extract salient characteristics of the shape information in the PVM for object categorization. This work provides guidance on the selection of descriptors and parameters for characterization of the PVM, making a fundamental step on the understanding of the shape information in the PVM.

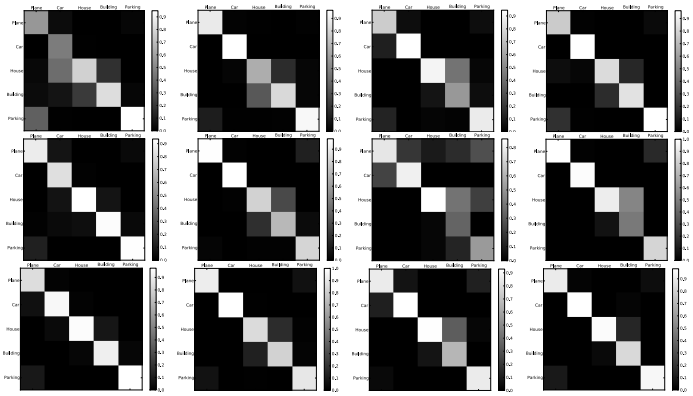

Figure 2: Average confusion matrices (across trials) . Columns (left to right): FPFH, SHOT, SI, SC. Rows (top to bottom): Nearest Neighbor, Bayes, SVM. In all cases $r_{supp} = 30$ and $K = 500$

[1] D Crispell, J Mundy, and G Taubin. A Variable-Resolution Probabilistic Three-Dimensional Model for Change Detection. *IEEE Trans. Geosci. Remote Sens.*, 2011.

[2] Andrea Frome, Daniel Huber, Ravi Kolluri, Thomas Bülow, and Jitendra Malik. Recognizing Objects in Range Data Using Regional Point Descriptors. In *ECCV*, 2004.

[3] A.E Johnson and M Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *PAMI*, 1999.

[4] Andrew Miller, Vishal Jain, and Joseph Mundy. Real-time Rendering and Dynamic Updating of 3-d Volumetric Data. In *Workshop on GPGPU*, 2011.

[5] T Pollard and J.L Mundy. Change Detection in a 3-d World. In *CVPR*, 2007.

[6] Restrepo, M.I, Mayer, B.A, Ulusoy, A.O. and Mundy, J.L. Characterization of 3-d Volumetric Scenes for Object Recognition. *IEEE J. Sel. Topics Signal Process.*, 2012.

[7] R.B Rusu, N Blodow, and M Beetz. Fast Point Feature Histograms (FPFH) for 3D Registration. In *ICRA*, 2009.

[8] Federico Tombari, Samuele Salti, and Luigi Di Stefano. Unique signatures of histograms for local surface description. In *ECCV*, 2010.