Chapter 6
Conclusions

This book presented a complete framework for efficient 3D scene modeling and mosaicing. The objective was to develop an efficient and flexible tool for remote scientific studies that require 3D and visual information under any type of acquisition conditions and scene types. Using such a tool, where no additional sensor information and no special acquisition conditions are required, decreases the complexity and costs related to scientific visual studies.

During the presentation of this work, we mainly focused on underwater scene modeling due to the increased difficulty and additional challenges that are present in this environment. Nevertheless, we show successful results on applying the framework on other types of environments, including land-based and urban scenes.

The core of the framework is based on a novel SfM algorithm – DPR-SfM. The algorithm generates the scene model sequentially, in two stages. In the first stage, a seed model is created using camera motion estimation techniques. The seed model corresponds to the first few frames of the sequence, representing a small subregion of the scene. The second stage extends the seed model in order to cover the entire surveyed area. While the first stage uses classical SFM techniques, the second stage uses a direct camera-to-scene registration method which increases the accuracy, robustness and flexibility of DPR-SfM. Also, results show that direct camera registration enables the algorithm to quickly recover from scene occlusions and tracking errors (*e.g.* due to excessive blurriness induced by fast camera movements, camera temporary failures, etc.).

Generally, large scale scene modeling algorithms use additional sensor information such as camera position and attitude for accurate results. We show that DPR-SfM achieves the same accuracy with no additional information, increasing its flexibility with respect to other SFM techniques. DPR-SfM can be readily applied on image sequences acquired with any type of camera, both still and video, using natural or artificial lighting (*e.g.* strobe/focus lighting for deep waters).

T. Nicosevici & R. Garcia: *Efficient 3D Scene Modeling and Mosaicing*, STAR 87, pp. 143–146.
DOI: 10.1007/978-3-642-36418-1_6

The direct camera pose registration uses a novel dual RANSAC projective/homography approach which allows the DPR-SfM algorithm to accurately model both planar and non-planar scenes. This is particularly important in underwater and urban scenes, where parts of the scene can have significant parallax while other parts can be perfectly planar. The use of robust estimation methods was also extended to vertex position recovery. We show that using robust estimation for both camera pose and vertex position estimations increases the accuracy and robustness of the method.

DPR-SfM uses an efficient and flexible scene database that enables the parallel use of multiple feature extractors/descriptors, while allowing fast camera registration in case of complex and large scenes. In this context, we employed a Kd-tree scheme for efficient association between image features and model vertices.

The second part of the framework deals with the detection of cross-overs during visual surveys. The novel BoW method (OVV) is oriented towards online navigation and mapping, eliminating significant drawback of state of the art visual vocabulary algorithms such as strong *a priori* knowledge of the surveyed area and tedious user intervention.

OVV uses an incremental vocabulary building process that eliminates the need for the offline training stage. During the survey, the vocabulary is initialized from visual information extracted from a small number of images, corresponding to the beginning of the sequence. Using a novel vocabulary update criterion which takes into account the visual information present in the images, the vocabulary is updated in order to constantly represent the visual information contained in the scene.

The vocabulary is built using a novel data clustering method. The clustering, based on Fisher's linear discriminant, takes into account the global data distribution rather than local inter-cluster relations. We show that such an approach ensures a more efficient data distribution, increasing both the repetitiveness and the discriminative power of the resulting vocabularies. The discriminative power of the vocabularies is further improved using Linear Discriminant Analysis, which increases the separability of the visual words within the vocabularies. Also, the use of LDA enables data dimensionality reduction, decreasing the computational costs related to vocabulary building and image indexing.

In the context of a constantly changing vocabulary, we propose a new hierarchical feature-cluster association technique, that increases the stability of feature labeling. We show that stable feature labeling is critical in detecting visual similarities between images that are indexed at different vocabulary update steps. Also, to increase the computational efficiency of OVV, we propose a novel incremental image re-indexing method, eliminating the high cost of repeatedly indexing the images as the vocabulary changes.

Finally, we propose a novel Online Model Simplification algorithm oriented towards large scene reconstruction algorithms. OMS simplifies the 3D model sequentially, by analyzing the scene geometry locally, using plane-parallax approximations. During simplification, OMS selects only those 3D vertices that are geometrically representative for elements present in the scene (*e.g.* edges, corners, surface inflections, etc.). The vertex selection criteria not only takes into account the geometrical representativeness of the vertices but also their reconstruction accuracy.

We show through experimental results that model simplification using OMS greatly reduced the complexity of the models while having minimal impact on the accuracy, with results similar to state of the art offline model simplification algorithms.

6.1 Contributions of the Book

In this book, we have presented a complete framework for 3D scene modeling. Particularly, we focused on developing an accurate and flexible SfM algorithm, online cross-over detection and 3D model simplification. Experimental results presented in this book show the efficiency and accuracy of the three modules. Hereafter, we present the main contributions of this book:

- In Chapter 3 we proposed a novel SfM algorithm based on direct registration between camera and scene model. While the model is initialized using classical motion-based techniques, the scene model is extended using a new sequential two step approach: (i) the camera pose is obtained from camera-model registration and (ii) using the camera pose, the model is extended to comprise the new information extracted from the camera view. The camera registration uses a novel dual approach that allows reconstruction of both planar/non-planar scenes.
- In Chapter 4 we developed an online cross-over detection algorithm, based on visual BoW. OVV uses a novel incremental visual vocabulary that eliminates the need of *a priori* knowledge of the scene being surveyed. The vocabulary building process uses an automatic update criterion, based on image content, that reduces the number of vocabulary updates. Also, the vocabulary building uses a novel clustering approach that increases the quality of the vocabulary and data separability while allowing data dimensionality reduction. The natural convergence criterion used during vocabulary building eliminates any user intervention, increasing the ease of use of the method.

  Image indexing is carried out by means of a novel indexing method using hierarchical trees. The method increases the stability of the image indexing process in the context of dynamic vocabularies. Furthermore, as vocabularies constantly change, we avoid repeated complete indexing of frames using an efficient incremental re-indexing method that takes into account the changes in the vocabulary.
- In Chapter 5 we propose a novel method for 3D model simplification oriented towards online 3D modeling applications. The method analyzes the scene locally, using plane-parallax, hence not requiring knowledge of the

full scene model. Based on plane-parallax approximations, the method selects vertices that are geometrically representative. For this, we present a novel vertex selection criteria that takes into account both geometric relevance and reconstruction accuracy of the vertices.

6.2 Ongoing and Future Work

The work presented in this book can be improved and extended in several ways. We present hereafter ongoing and future work directions:

Direct Structure from Motion. After camera pose registration, image patches around features can be warped using camera-to-model transformations, reducing the effect of extreme geometric distortions on feature tracking. Also, the scene reconstruction accuracy can be improved, using cross-correlation as a refinement step after feature matching. Finally, the computational time of feature-to-model association can be highly decreased using recent developments in Graphics Processing Unit (GPU)-based parallel processing.

Online Loop Detection. The complexity of the visual vocabularies can be reduced by eliminating small, insignificant clusters at the bottom of the hierarchy. This would allow online cross-over detection for larger scenes in a more efficient way. Again, the use of GPU-based parallel processing would highly decrease the computational time related to vocabulary building and image indexing.

Online 3D Model Simplification. A new multi-resolution representation of the scene based on vertex geometrical relevance could be developed, similar to PM. Such a representation would allow the user to select the level of detail of the model, depending on the specific needs, hardware limitations, etc.