

Thesis Title

Sourav Garg

May 22, 2017

Contents

1	Introduction	2
1.1	Motivation, Challenges and Research Gap	2
1.1.1	Motivation from practical applications	2
1.1.2	Challenges involved in development of the related technology	2
1.1.3	Research Gap	2
2	Research Problem	2
2.1	Problem Statement	2
2.2	Research Questions	3
3	Literature Review	4
3.1	Problem Overview	4
3.2	Related Work	4
4	Research Plan	4
4.1	Characterization of visual SLAM system	4
4.1.1	Visual Odometry and Place Recognition	5
4.1.2	Adapting to Environment	6
4.1.3	Egomotion	7
4.2	Utilizing Semantic Information	8
4.2.1	Environment Semantics	8
4.2.2	Semantic Segmentation <i>Across</i> Places	8
4.2.3	Semantic Segmentation <i>Within</i> Places	9
4.3	3D Condition- and Viewpoint-Invariant Visual SLAM	9
4.3.1	Condition-Invariant Relative Pose Estimation	9
4.3.2	Semantics in 3D	9
5	Work Progress	9
5.1	Performance Evaluation Using High Fidelity Simulation	10
5.1.1	Place Recognition - SeqSLAM	10
5.1.2	Visual SLAM - ORB-SLAM	10
5.1.3	Paper Accepted at IROS 2016	11

5.2	Place Recognition Using Semantics	11
5.2.1	Semantic Segmentation of Environment	11
5.2.2	Paper Submitted to IROS 2017	12
5.3	Motion Estimation under Unfavorable Conditions	12
5.3.1	Low Light and High Speed	12
5.3.2	Speed-Normalized Data Sampling	13
5.3.3	Planned Submission for ICRA 2018	13
6	Timeline	14
7	Conclusion	14

1 Introduction

1.1 Motivation, Challenges and Research Gap

1.1.1 Motivation from practical applications

- Start with practical applications that motivate the research problem that is to be solved with the PhD thesis.
- Some examples of such applications in a broader perspective.
Autonomous vehicles and ADAS, surveillance, indoor navigation (big retail stores, malls, auditoriums etc.) etc.
- Gradually converge to the exact research problem to be solved.
Ability to recognize places and estimate ego-motion visually under extreme changes in environmental conditions.
- Some examples of potential practical applications of the solution that will be developed with the proposed research work.

1.1.2 Challenges involved in development of the related technology

1.1.3 Research Gap

- Briefly review the literature to highlight the research gap quoting the related work.
- Mention exactly where the research gap is.

2 Research Problem

2.1 Problem Statement

- Elaborate on the research problem after formally stating it.

- Developing a condition- and viewpoint-invariant visual semantic SLAM system
- Highlight the key aspects of the problem
- Describe the challenges related to the problem that we are going to solve in proposed research.
 - Characterization of the components of such a competency for a detailed understanding of what needs to be fixed
 - Effective use of semantic information for robustness and meaningful interaction within the system
 - Creating a 3D SLAM system integrating different useful components from 1D or 2D and scale them accordingly.

2.2 Research Questions

The overarching research question following the literature review and the identified research gap is -

How can we develop a general purpose visual place recognition or visual SLAM system which is condition- and viewpoint-invariant and adapts to changes in environment as well as egomotion?

In order to answer this and to enable such a system, we need to answer the following questions:

- How can we identify and model the key characteristics of visual SLAM like competencies in robotic applications?
*The three main components of any such competency are: the **robot** enabled with sensors (like camera), the **environment** where the robot operates and the **algorithm** that enables effective interaction between the former and the latter.*
 - How can we characterize the place recognition, visual odometry and visual SLAM algorithms for robustness against variations in camera viewpoint and environmental conditions?
 - How can we characterize the environment so as to enable adaptive behavior in robots by dynamically determining the changes in the environment?
 - How can we estimate the camera motion under extreme environmental conditions and speedy camera movements?
- How can we use semantic information related to images in order to increase the robustness of visual place recognition and SLAM methods?
 - How can we encode environmental conditions information into semantic labels within and across the images in an image sequence?

- Can we use semantic image labels to categorize the environment into meaningful segments for adaptive robot behavior?
- Can we learn image pixel-level semantic mask for robust place recognition under extreme environmental conditions?
- What are the key unsolved challenges in developing a condition- and viewpoint-invariant 3D (metric) visual SLAM system?
 - How can we estimate 3D relative pose between images with vast appearance variations?
 - How well does the characterization of 1D place recognition system and semantic labeling across and/or within the images scale in 3D?
- Elaborate on each question.

3 Literature Review

3.1 Problem Overview

- Provide an overview of the research problem quoting the literature.

3.2 Related Work

- Describe in details the related body of work with respect to the research problem and the research questions.

4 Research Plan

Elaborate following points in subsequent subsections.

- Characterization of environment and camera sensor with respect to visual place recognition and visual SLAM algorithms.
- Exploring possible ways and extent of using semantic information to improve visual place recognition and SLAM system.
- Developing a 3D metric visual SLAM system which is condition- and viewpoint-invariant and adapts to changes in environment and egomotion.

4.1 Characterization of visual SLAM system

The visual SLAM system or any other similar competency developed as a robotic application can broadly be understood to comprise these three components:

1. *Environment* where the robot operates or learns about its surroundings.
2. *Robot and Sensors* for interacting with the environment and enabling sensing and responding in some form.

3. *Algorithm Interface* is what enables the interaction between the robot and the environment and is designed specific to the application.

Add figure?

In order to develop a well-performing robotic application, it is important to understand the aforementioned components and their characteristics that significantly impact the performance. Therefore, it remains one of the research questions to identify such characteristics specific to the application and appropriately model them. The research plan for exploring the characterization of each component is detailed in the following subsections.

4.1.1 Visual Odometry and Place Recognition

Visual Simultaneous Localization and Mapping constitutes: *Visual Odometry*, *Visual Place Recognition* and *Mapping* of environment. The state-of-the-art visual SLAM systems or any of its individual constituents are often developed based on certain assumptions with respect to its environment, its sensors or the operating characteristics. Though it is not possible to estimate all the possibilities of operating characteristics beforehand, it is vital to understand the ones that significantly impact the performance.

SLAM systems at disposal: We plan to explore the state-of-the-art methods like SeqSLAM [1], ORB-SLAM [2], LSD-SLAM [3] etc. in order to understand their certain characteristics, for example, invariance to camera viewpoint and environmental conditions - both of which lie at the core of a general purpose visual SLAM system.

Data Gathering: The simultaneous condition- and viewpoint-invariance of a 3D visual SLAM system has yet not been achieved due to challenges involved. Therefore, it is important to understand the performance sensitivity of the existing algorithms towards these variations. Such an analysis will also require the appropriate data to conduct tests. However, the availability of real or simulated data corresponding to condition and viewpoint variations is limited. This is mainly because it is difficult, time-consuming and expensive to collect real data with a sound ground-truth generation system especially for condition variations - which essentially means traversing places at different times of day, in different weather conditions and across seasons. Similarly, capturing images of a place for all possible camera viewpoints is not always feasible. It seems relatively easier to do the same in simulation, given there exists a 3D model of a city or alike with sufficient rendering, so that condition- and viewpoint-varied traverses could be infinitely generated, but such a ready-to-use model does not exist and needs a lots of effort.

Simulation: We plan to collaborate with peers to simulate a city-like environment in 3D and obtain repeated traverses with varying viewpoint and conditions

of the environment. Along with that, we will also make use of some of the existing and newly collected real world data which may not be sufficiently large as compared to the simulated one, but be rich enough with respect to the variations in the environment such that the performance curves so obtained will look similar to those using simulated world. The choice of relying on simulation than real world for characterizing the visual SLAM systems is merely due to large amount of effort required in the latter and collaboration opportunities at disposal for the former.

Conclusion: The characterization in this form will immensely help in performance benchmarking using simulated data for different robotic applications. This will also enable the enhanced understanding of parameters of the system that particularly drive its performance for the variations induced in the test data.

4.1.2 Adapting to Environment

Motivation: The environment plays a key role in determining the performance of any robotic application because it is often the case that many of the applications are designed specific to certain environment types and therefore are brittle towards significant variations in the environmental settings. In context of visual place recognition and visual odometry, the variations in environment are induced mainly due to *scene structure* and *transient conditions*. The scene structure is actually the basis for differentiating places from each other, but frequent transitions within different environments call for different parameter settings for improved performance, for example, urban canyons versus highways or forests, or outdoor vs indoor etc. On the other hand, transient conditions like time of day, weather and season, though representing the same place, remarkably change the visual appearance of the scene. Therefore, it is necessary to understand the environment and allow possible adaptive robot behavior for variations in the environment.

Recognizing places in varying environment: The condition-invariant place recognition methods such as SeqSLAM [1], SMART [4], and others [5, 6, 7] have been shown to be able to recognize places under the influence of extreme variations in conditions of the environment. These variations are mainly *global* which means that they are static with respect to space but not time, for example, change in season, weather or time of day will only affect the appearance of a particular place in the environment when it is visited after a certain interval of time. On the other hand, the spatial neighborhood of that place will be similarly affected by the changes in the environmental conditions.

We are rather interested in local variations in the environment where global conditions changes may or may not occur. Examples of such local variations are mainly related to transitioning from outdoor to indoor environments, textureless to cluttered scenes, urban canyons to forest roads etc. The seamless tran-

sition between such environmental settings is only possible by segmenting the environment into different chunks based on their appearance attributes.

Segmenting the environment: The place recognition methods generally employ a measure of matching places which can be either whole-image based as in SeqSLAM [1] or point features based collective score as in FAB-MAP [8]. These place matching scores effectively discriminate places from each other and hence could also be used to create segments based on these scores. The idea is to build a self-similarity matrix for a given dataset and then statistically perform segmentation or data clustering based on the affinity scores. The segments so formed can then be used to define neighborhood region for a place such that place recognition performance may be improved based on collective matching of different segments. This would also require us to collect datasets which possess transitions from one type of environment into the other quite often.

Conclusion: The characterization of environment with respect to local variations in overall appearance of environment will help in dynamically adapting the robot behavior accordingly. For example, place matching and motion estimation can be improved by tuning the system parameters in order to accommodate the variations in environment.

4.1.3 Egomotion

Motivation: The egomotion estimation is an important competency for robotic tasks that involve movement. Visual Odometry is the means of estimating egomotion using visual cues. The state-of-the-art visual odometry solutions are often prone to failures because of fast camera motion, motion blur, photometric and radiometric changes in appearance of environment etc. The failure of visual odometry is catastrophic for a visual SLAM system as the robot immediately loses its localization information and cannot relate current motion estimates with those collected earlier. The camera motion also plays a significant role in place recognition especially when repeated traverses of an environment exhibit different motion pattern, which makes it harder to recognize places. This is because a fixed-size temporal neighborhood window centered at a place will no longer contain similar places in different traverses.

Speed Normalization: We will look into the scope of improving place recognition using visual odometry. Most of the practical visual place recognition applications for example, an autonomous vehicle, do not exhibit uniform motion during the journey. The motion estimation can help in speed normalizing the collected imagery, so that the places are separated by a constant physical distance instead of constant number of frames.

Motion Prediction: We also plan to develop a motion estimator or predictor as a backup switch for state-of-the-art visual odometry methods such that it

provides some source of information to roughly localize the robot even if such an estimate is less accurate or less precise. It is important because such estimates can be improved at a later stage by either using visual place recognition or improved visual odometry.

4.2 Utilizing Semantic Information

The use of semantic information has recently become easier because of deep-learned classifiers and regressors that can be trained on very large image data to precisely predict the semantics of the test image. These semantics can be used to achieve a meaningful interpretation of places and maps used in a visual SLAM system. We plan to make use of this semantic information to improve visual place recognition and visual SLAM system in different ways as described in following subsections.

4.2.1 Environment Semantics

One of the simplest use of semantics seems to be imparting meaning to the environment where robot operates. The place categorization [9] and scene attributes detection [10] using semantics has not been explored within a place recognition framework so far. Unfortunately, these semantic classifiers do not provide sufficient information regarding environmental conditions like time of day, season or weather etc., but mostly characterize the structure of the scene leading to particular place category labels. These transient conditions of environment have been semantically explored in [11].

We plan to use the semantic attributes based on scene structure as well as transient conditions from different classifiers to create ground truth semantic labels of both types for a huge database. These labeled images will then be used to train (or retrain) a deep neural network to learn the semantics of images for both scene structure as well as transient conditions. This will make place categorization more robust and easy to incorporate in a place recognition system.

4.2.2 Semantic Segmentation *Across* Places

The segmentation of places in an environment has been explored in earlier sections based on place matching scores. The use of semantics for segmenting the environment can be advantageous because firstly, deep-learned features are more discriminative as compared to hand-crafted ones and secondly, it can help in filtering or shortlisting of the place matching candidates by performing a semantic matching first.

The semantic segmentation within an environment will be explored in context of place recognition. The filtering or shortlisting of place matching candidates leading to reduction in computation time for searching places will be benchmarked for large databases. The performance improvement in place recognition using place matching score as described earlier and using semantics will

be also be compared.

4.2.3 Semantic Segmentation *Within* Places

4.3 3D Condition- and Viewpoint-Invariant Visual SLAM

4.3.1 Condition-Invariant Relative Pose Estimation

4.3.2 Semantics in 3D

5 Work Progress

- Describe the work done till date with respect to the research problem and how it relates to the main objective
 - Formal characterization of place recognition and visual SLAM algorithms like SeqSLAM and ORB-SLAM
 - * Using simulated 3d environment to understand the performance variations in place recognition, visual odometry and SLAM algorithms by testing repeatedly generated image sequences that vary with respect to the viewpoint (6-DoF) and environmental conditions (time of day etc.).
 - Characterization of camera-motion as well as the environment where the camera operates.
 - * Developing methods to improve place recognition by segmenting the environment into meaningful chunks
 - Using semantic place categorization to segment data into semantic chunks like indoor-outdoor, light-dark etc. in order to define neighborhood for a reference image for effective place matching within that chunk.
 - Using SAD scores within consecutive frames instead of a semantic categorizer to achieve the same as above.
 - Dynamically defining such segments by using cues from query images.
 - * Developing methods to improve place recognition, visual odometry and visual SLAM by estimating camera speed under poor environmental conditions
 - Improving place recognition performance by speed-normalizing the image sequence for effective place matching using camera speed as a function of SAD scores between consecutive frames.
 - Developing hybrid visual odometry by use of state-of-the-art method backed with a SAD score based camera speed estimator to prevent failing of visual odometry and visual SLAM algorithms due to factors like low-light environment, fast camera motion, motion blur etc.

- Mention publications or submitted work (if any)
- IROS'16

5.1 Performance Evaluation Using High Fidelity Simulation

The use of high fidelity simulation is an attractive option for most of the researchers primarily because it gives access to infinite data generation which is very close to real world. A simulated environment enables detailed performance evaluation of algorithms which helps in understanding the data-dependent characteristics of the system and optimizing its parameters. We evaluated various robotic vision algorithms like place recognition, visual odometry, visual SLAM and object recognition to study their performance variations with respect to variations in the input data. The high fidelity simulation of a city-like environment was initially developed to generate different datasets and then fed to different algorithms for performance evaluation. The work was done with collaborative efforts from peers and the research components relevant to this report are described in subsequent subsections.

5.1.1 Place Recognition - SeqSLAM

The visual place recognition is the capability of a mobile robot to recognize a place during a revisit solely using visual cues. Thus, it requires repeated traverses of the environment with no restriction on viewpoint or environmental conditions like time of day, weather or season etc. during the subsequent visit. We evaluated performance of SeqSLAM - a condition-invariant place recognition algorithm using different datasets generated from the simulated environment. These datasets varied with respect to 5 different times of day, that is, Dawn, Morning, Noon, Afternoon and Sunset as well as different camera viewpoints with variations in lateral offset, vertical orientation (pitch) and horizontal orientation (yaw). We observed the performance curves which as per the expectations showed a decrease in performance with extreme viewpoint variations and a constant high performance with varying conditions of environment. We also performed some experiments on real world data along with its simulated version to observe the performance trends which were found to be similar. However, the results on simulated data were consistently better than the real world data which is also often the case with any simulation.

5.1.2 Visual SLAM - ORB-SLAM

ORB-SLAM is a state-of-the-art visual SLAM system proven to work well for both monocular and stereo or depth-based input. Similar to performance evaluation of visual place recognition, evaluating a visual SLAM system also requires repetitive traversing of an environment, but in a continuous loop. This is necessary to make sure that all the components of a SLAM system that is

visual odometry, place recognition and mapping are effectively tested. Therefore, different traverses of the environment at different times of day and varying viewpoints as also described in previous subsection were always appended by a noon baseline dataset. Such an arrangement made sure that visual odometry component gets tested in the first part of the data along with its capability of continuity across the appended data in terms of local changes. The variation of first part across different datasets and second part within the dataset made sure that the place recognition component is effectively tested. We used average trajectory error as a performance measure for generating the performance curves. It was observed that the performance of ORB-SLAM did not show any consistent pattern with variations to either viewpoint or environmental conditions. In general, in a visual SLAM system, occurrence of even a single false loop closure or a failure in visual odometry is catastrophic. The former may lead to an absurd trajectory with high error and the latter will lead to an incomplete trajectory. Moreover, most of the visual SLAM algorithms use initialization methods that use random numbers, therefore, expecting a consistent pattern in a performance curve is probably not viable.

5.1.3 Paper Accepted at IROS 2016

The research work was accepted as a conference paper at IROS 2016.

5.2 Place Recognition Using Semantics

The use of semantics with respect to places has been rediscovered recently with deep-learned CNN models after its successful experiments with object recognition. The semantics for places are generally with respect to a single image, and provide labels about the scene structure, texture, environmental conditions etc. **“semantic labels fig”** Though, it enables place recognition in a coarse manner, recognizing *specific* places in an environment which may have similar semantic labels is the *traditional* place recognition problem dealt in robotics. We explored the use of semantic labels for individual places within an image dataset in improving place recognition performance.

5.2.1 Semantic Segmentation of Environment

One of the apparent ways of using semantic labels in place recognition is to coarsely filter the places that may match the query place. This will definitely reduce the computation time for place search within the database, and may also improve performance depending on the robustness of features used to semantically categorize places. Another way of utilizing the semantic information in context of place recognition is to exploit the temporal nature of places data encountered in both reference and query databases. A mobile robot traversing an environment is most likely to witness variations in its environment, both minor and major, especially in the applications of a visual SLAM system. We

developed a system that performed semantic place categorization on an incoming stream of images to form temporal semantic segments of places which are generally similar. This was followed by a more *specific* (or *traditional*) place recognition system that utilized these temporal segments to define the neighborhood region around reference query places to bias their matching scores accordingly. **“block dia”** The effectiveness of this system is more pronounced for the image datasets having significant variations in the environment, for example, transiting between indoor and outdoor environments like a train running through tunnels as well as open areas or a mobile robot within a university campus; transiting between regions of varying illumination or texture like an urban canyon versus highway within a forest etc. We tested our system on a wide variety of datasets ranging from a 23 km train journey to short campus traverses, all exhibiting variations in environment with respect to scene structure, illumination, environmental conditions, texture etc. **”dataset example fig”** We observed a considerable performance gain using the proposed system for datasets with medium to extreme variations in their environment. **“results fig”**

5.2.2 Paper Submitted to IROS 2017

The proposed research work has been submitted as a conference paper to IROS 2017.

5.3 Motion Estimation under Unfavorable Conditions

Visual Odometry being one of the important competencies for mobile robotics is also an integral part of a visual SLAM system. The correct estimation of egomotion depends on the characteristics of both the camera motion and the environment. The state-of-the-art visual odometry methods are, though able to generate a 6-DoF pose, are sensitive to extreme changes in the camera motion or the operating environment. We explored the applicability of these methods on some of the real world datasets that exhibit variations in camera speed, for example, a vehicle on a heavy traffic route with frequent halts, and variations in illumination, for example, transiting from an artificially-lit to unlit environment at night etc.

5.3.1 Low Light and High Speed

The low light environment means less visual features as compared to the scene captured in broad daylight. Such a gradual drop of visual features, for example, while moving from an artificially-lit, cluttered scene to an unlit bland environment at night, leads to failure of visual odometry. We performed experiments on different datasets exhibiting such variations in environment using visual odometry component of ORB-SLAM and LSD-SLAM. The test datasets also had varying camera speeds in different parts of environment which further

led to failures because capturing large and small motion signals requires different parameters settings for the system.

We developed a method to generate a motion signal for unfavorable conditions as described above, using sum of absolute difference (SAD) score between down-sampled and patch-normalized consecutive images. This is similar to place recognition method used in SeqSLAM. The patch-normalization of images is able to handle the variations due to environmental conditions. It was also observed that SAD scores between images drops gradually with increasing frame separation between the images in an image sequence. Hence, we used a polynomial fitting method to estimate the camera speed according to the SAD score patterns. We were able to estimate the high-speed instances of camera motion within the test dataset where state-of-the-art methods failed. We intend to use a hybrid approach that can make use of the proposed motion estimator to either tune the state-of-the-art method parameters to correctly estimate a 6-DoF pose or to club the output of both the systems to prevent failure and generate a consistent odometry information.

5.3.2 Speed-Normalized Data Sampling

We explored the scope of improving place recognition performance using state-of-the-art visual odometry methods. The motion information is often required in the place recognition framework in order to sample the images at a constant distance as opposed to constant frame separation. A robot with uniform motion will not be affected by such a choice, but in practice, the place recognition and visual SLAM applications contain reference and query imagery captured at varying camera speeds, for example, a vehicle on road with varying traffic conditions on same route at probably different times of day. The speed-normalized imagery can be either obtained using a separate sensor or a visual odometry solution. The challenges involved in using a state-of-the-art visual odometry solution are described in previous subsection. Hence, we resort to the use of proposed motion estimator, also described above, to speed-normalize the image databases for both query and reference. The motion estimator determines the ideal frame separation between the images based on the SAD score between them. This dynamic frame separation corresponds to the physical camera motion and is estimated high when camera moves slowly and low when camera moves fast. We observed a significant improvement in SeqSLAM's place recognition performance using the proposed speed-normalized data sampling before starting to match the places. **“fig-results”**

5.3.3 Planned Submission for ICRA 2018

We plan to submit the proposed research work to ICRA 2018 with some additional research work on top of it.

6 Timeline

7 Conclusion

References

- [1] Michael J. Milford and Gordon F. Wyeth. SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights. *Proceedings - IEEE International Conference on Robotics and Automation*, pages 1643–1649, 2012.
- [2] Raul Mur-Artal, J M M Montiel, and Juan D Tardos. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Transactions on Robotics*, pages 1–17, feb 2015.
- [3] Jakob Engel, Thomas Schöps, and Daniel Cremers. LSD-SLAM: Large-Scale Direct Monocular SLAM,. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *European Conference on Computer Vision (ECCV)*, volume 8690 of *Lecture Notes in Computer Science*, pages 834–849, Cham, 2014. Springer International Publishing.
- [4] Edward Pepperell, Peter Corke, and Michael Milford. All-environment visual place recognition with SMART, jun 2014.
- [5] Tayyab Naseer, Luciano Spinello, and Cyrill Stachniss. Robust Visual Robot Localization Across Seasons using Network Flows. *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 2014.
- [6] S Niko, Sareh Shirazi, Adam Jacobson, Feras Dayoub, Edward Pepperell, Ben Upcroft, and Michael Milford. Place Recognition with ConvNet Landmarks: Viewpoint-Robust, Condition-Robust, Training-Free. *Robotics Science and Systems*, 2015.
- [7] Will Maddern, Alexander D Stewart, and Paul Newman. LAPS-II: 6-DoF Day and Night Visual Localisation with Prior 3D Structure for Autonomous Road Vehicles.
- [8] Mark Cummins and Paul Newman. Highly scalable appearance-only SLAM - FAB-MAP 2.0. In *Robotics: Science and Systems*, Seattle, United States, 2009.
- [9] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. Learning deep features for scene recognition using places database. In *Advances in neural information processing systems*, pages 487–495, 2014.
- [10] Genevieve Patterson and James Hays. SUN Attribute Database: Discovering, Annotating, and Recognizing Scene Attributes. In *Proceeding of the 25th Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

- [11] Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, and James Hays. Transient attributes for high-level understanding and editing of outdoor scenes. *ACM Transactions on Graphics (TOG)*, 33(4):149, 2014.