Proceedings Paper:
Haggart, R. and Aitken, J.M. (2021) Online scene visibility estimation as a complement to SLAM in UAVs. In: Fox, C., Gao, J., Esfahani, A.G., Saaj, M., Hanheide, M. and Parsons, S., (eds.) Towards Autonomous Robotic Systems : 22nd Annual Conference, TAROS 2021, Lincoln, UK, September 8–10, 2021, Proceedings. TAROS 2021 : Towards Autonomous Robotic Systems, 08-10 Sep 2021, Lincoln, UK. Lecture Notes in Computer Science (13054). Springer Cham, pp. 365-369. ISBN: 9783030891763. ISSN: 0302-9743. EISSN: 1611-3349.

https://doi.org/10.1007/978-3-030-89177-0_38

Online Scene Visibility Estimation as a Complement to SLAM in UAVs*

Rory Haggart, Jonathan M. Aitken

Department of Automatic Control & Systems Engineering, University of Sheffield
{rdehaggart1, jonathan.aitken}@sheffield.ac.uk

Abstract. Simultaneous localisation and mapping (SLAM) relies on low-cost on-board sensors such as cameras and inertial measurement units. It is crucial that the surroundings are visible to the cameras to maximise the accuracy of the system. An estimation strategy is proposed to augment ORB-SLAM2 that considers feature extraction capability, distribution of the extracted features in the image frame, and the ability of the algorithm to track features over time. The method is tested on challenging datasets, and the output is evaluated against different visibility conditions. The proposed method is shown to react appropriately and consistently to 'less visible' conditions such as fog, sunlight, and rapid motion in real time, with minimal computational load.

Keywords: Simultaneous Localisation and Mapping · Visibility

1 Introduction

In the field of robotic navigation, simultaneous localisation and mapping (SLAM) uses low-cost, on-board sensors to build up a three-dimensional representation of the local surroundings and localise the robot relative to points in this map.

Semi- and fully-autonomous systems are on the rise. Between 2011 and 2017 the number of patents relating to automated driving that were filed at the European Patent Office rose by 330% [2]. In vehicles employing these systems, the role of action and reaction is assumed by on-board sensors and actuators interfacing with decision making systems to control the vehicle. If a system using SLAM with visible-light cameras can 'understand' how visible the scene is to those cameras, it could adapt - re-orienting the cameras, or adjusting how many features should be extracted from the incoming image stream.

* Supported by the Department of Aerospace Engineering and the Department of Automatic Control and Systems Engineering at the University of Sheffield. Also this work is supported by the UK's Engineering and Physical Sciences Research Council (EPSRC) Programme Grant EP/S016813/1

2 Related Work

2.1 SLAM

ORB-SLAM2 [5] is an indirect visual SLAM technique, meaning features are extracted from preprocessed images and tracked between frames. The features are described using binary descriptors and used to perform global bundle adjustments and loop closures that allows for a consistent position estimation.

VINS-Mono [8] is a popular and sophisticated visual-inertial algorithm that has consistent and accurate tracking of sensor pose. The authors note that whilst their technique may operate in poor visibility, improvements to investigate observability properties of the online camera data would be beneficial.

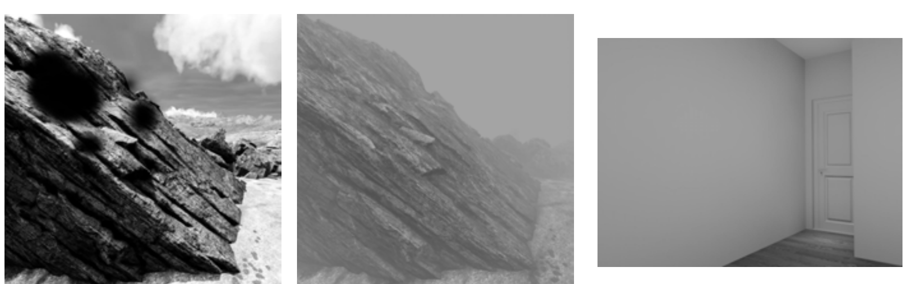
2.2 Scene Visibility Estimation

The authors of [6] present a model that accounts for the multiple scattering of light in the atmosphere due to conditions such as fog and rain. This is based on the glow surrounding light sources in inclement weather. It is one of several attempts to estimate dynamic visibility distance based on the presence of fog.

In [7], the authors develop a technique that uses the observed contrast of road markings. The system is tested for a variety of conditions - e.g. when sunrise causes glare in the image, which interestingly resulted in a lower visibility estimate than the more frequently studied case of fog. The method is shown to be robust to a range of conditions, but relies on the presence of known features.

3 Methodology

The ORB-SLAM2 code was modified to use information about extracted features in each frame to calculate the visibility estimation metric components (see Table 1). Additional processes, such as ones to save the outputs, were also added.



(a) (b) (c)

Fig. 1: Example frames of (left to right) partially occluded, foggy, and featureless scenes from the Midair [3] and InteriorNet [4] datasets

	Equation	Description
a	$S_a = \frac{N_F}{N_{F,max}}$	S_a is defined as the ratio of the number of extracted features (N_F) to the target, defined by the user ($N_{F,max}$). This is how well the camera can 'see' the scene
b	$S_b = 1 - \frac{\chi^2}{\chi^2_w}$ $\chi^2 = \sum_{i=0}^{N_B} \frac{(O_{b_i} - E_{b_i})^2}{E_{b_i}}$ $E_{b_i} = \frac{N_F}{N_B}$	Each frame is divided into N_B 'bins', each containing some number (O_{b_i}) of the extracted features. A chi-square value (χ^2) of this binned distribution is calculated, where E_{b_i} is the 'expected' number of features in each bin b_i if the distribution of features was homogeneous. S_b is defined as the complement to the chi-squared value when normalised against a 'worst-case' value (χ^2_w), representing the condition of all extracted features being positioned exclusively in $1/8^{\text{th}}$ of the frame. This evaluates the homogeneity of the distribution
c	$S_c = \frac{N_T}{N_{L_v}}$	S_c is defined as the number of features that are tracked (N_T) as a fraction of the number of features that are theoretically located within the frustum of the camera (N_{L_v}). For more dynamic visibility, features may be lost even whilst they remain within the cameras line of sight

Table 1: Three ORB-SLAM2 SVE Calculated Components.

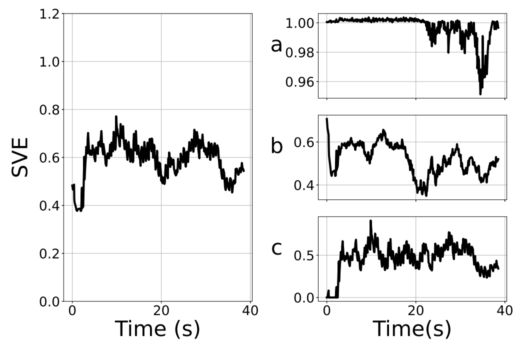
4 Results and Discussion

Trajectory	**Condition**	**Mean** S	**Mean % Tracked**
VO_test 0	Sunny	0.737	96.92%
	Sunset	0.723	94.76%
	Foggy	-0.324	2.59%

Table 2: The mean values of S ($S = 0.2S_a + 0.4S_b + 0.4S_c$) and of the percentage of the trajectory that was successfully tracked by ORB-SLAM2 over 15 tests of three conditions in a trajectory from the MidAir dataset.

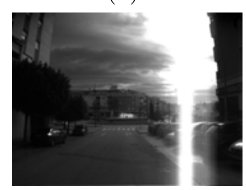
With ORB-SLAM2 augmented to become ORB-SLAM2 SVE (ORB-SLAM2 with Scene Visibility Estimation), tests were performed using the MidAir [3], InteriorNet [4], and Malaga [1] datasets. The set of visibility impairments that could be tested were fog (Figure 1b), partial lens soiling (Figure 1a), direct sunlight (Figure 2c), rapid motion, featureless scenery (Figure 3c), and planar scenery. Tracking sustainability and visibility for some of the MidAir data are shown in Table 2 with a sample visibility output from ORB-SLAM2 SVE in Figure 2a.

An assessment of the execution time using a tool developed by the authors of ORB-SLAM3 revealed that the additional components had minimal impact on the computational load, and that the implementation was efficient allowing the algorithm to perform a high accuracy estimation.



(a)

(b)



(c)

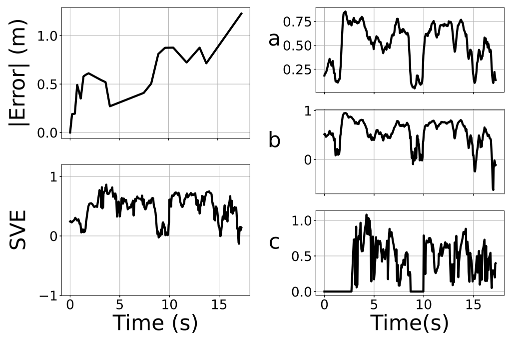
Fig. 2: (a) shows the visibility outputs from ORB-SLAM2 SVE for trajectory 15 of the Malaga dataset (b) shows the image frame with the highest associated S_b ($t \approx 0.4\,\mathrm{s}$) (c) shows the frame with the lowest associated S_b ($t \approx 22.3\,\mathrm{s}$)

5 Conclusions

Refinement is still required. Across all tests, the strategy responded appropriately in real time to qualitatively less visible frames as a result of factors including fog, direct sunlight, and featureless scenery, improving on existing methods that account for single factors. S_b proved the most intuitive metric, but no direct correlation between S_b and tracking accuracy was observed. However, using the ORB-SLAM2 visualiser, it was recognised that a poor distribution of tracked features - rather than extracted features - in the frame led to a worsened pose estimate. A detailed assessment was not completed. Additionally, S_b did not always show adequate sensitivity in conditions such as partial lens soiling (see Figure 1a), and this could indicate the need for tuneable parameters.

S_c should have been useful - as the number of tracked features decreases, tracking accuracy should worsen. The expectation was that before tracking is lost, S_c should start decreasing, though this was not always observed. The value was also highly variable between frames, and trends were hard to decipher - applying this calculation to keyframes rather than all frames may be a solution.

After these problems have been addressed, the visibility information could be fed to the system to adapt performance or re-orient the hardware, as discussed in Section 1.



(a)

(b) $S_a = 0.77$, $S_b = 0.81$, $S_c = 0.60$

(c) $S_a = 0.05$, $S_b = -0.02$, $S_c = 0.00$

Fig. 3: (a) shows the visibility outputs from ORB-SLAM2 SVE for trajectory 'original_3_3' in the '3FO4K7I2Q0PG' subset of the InteriorNet dataset. (b) shows the image frame with the highest associated visibility ($t \approx 4\,\mathrm{s}$) (c) shows the frame with the lowest associated visibility ($t \approx 9.19\,\mathrm{s}$)

References

1. Blanco, J.L., Moreno, F.A., Gonzalez-Jimenez, J.: The málaga urban dataset: High-rate stereo and lidars in a realistic urban scenario. International Journal of Robotics Research **33**(2), 207–214 (2014)
2. European Patent Office: Patents and self-driving vehicles. Tech. rep., European Patent Office (2018)
3. Fonder, M., Droogenbroeck, M.V.: Mid-air: A multi-modal dataset for extremely low altitude drone flights. In: Conference on Computer Vision and Pattern Recognition Workshop (CVPRW) (2019)
4. Li, W., et al.: Interiornet: Mega-scale multi-sensor photo-realistic indoor scenes dataset. In: British Machine Vision Conference (BMVC) (2018)
5. Mur-Artal, R., Tardós, J.: ORB-SLAM2: An open-source slam system for monocular, stereo, and RGB-D cameras. IEEE Transactions on Robotics **33**(5), 1255–1262 (2017)
6. Narasimhan, S.G., Nayar, S.K.: Shedding light on the weather. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. vol. 1, pp. I–I (2003)
7. Pomerleau, D.: Visibility estimation from a moving vehicle using the RALPH vision system. In: Proceedings of Conference on Intelligent Transportation Systems. pp. 906–911 (1997)
8. Qin, T., Li, P., Shen, S.: VINS-Mono: A robust and versatile monocular visual-inertial state estimator. IEEE Transactions on Robotics **34**(4), 1004–1020 (2018)