Proceedings Paper:

Evans, M.H., Aitken, J.M. orcid.org/0000-0003-4204-4020 and Anderson, S.R. orcid.org/0000-0002-7452-5681 (2022) Assessing the feasibility of monocular visual simultaneous localization and mapping for live sewer pipes : a field robotics study. In: 2021 20th International Conference on Advanced Robotics (ICAR). 20th International Conference on Advanced Robotics (ICAR), 06-10 Dec 2021, Virtual conference. Institute of Electrical and Electronics Engineers , pp. 1073-1078. ISBN 9781665436854

https://doi.org/10.1109/ICAR53236.2021.9659486

Assessing the feasibility of monocular visual simultaneous localization and mapping for live sewer pipes: a field robotics study

Mathew H. Evans[1], Jonathan M. Aitken[1], Sean R. Anderson[1]

***Abstract*— Sewer pipes are important to inspect for damage and blockages. Mobile robots with cameras are a natural choice for inspecting sewers, and indeed CCTV inspection using tethered mobile platforms is a well-established commercial approach. It therefore makes sense to also explore the use of camera data for localising defects for targeting subsequent repair. Visual odometry (VO) methods have been researched for robot localisation in pipes but the full visual simultaneous localisation and mapping (vSLAM) problem has received little attention. Whilst VO focuses on estimating the current pose of the robot, vSLAM focuses on building a map, as well as pose estimation, which should increase accuracy and robustness – both important for the future use of autonomous robots in sewer inspection. In particular, it is not known if one crucial element of vSLAM – loop closing using appearance-recognition methods – works effectively in sewer pipes due to problems of perceptual aliasing – where the high degree of visual similarity in image frames can lead to incorrect loop closures causing the vSLAM system to fail. The aim of this paper is to assess the feasibility of vSLAM for sewer pipes using real world data. The results demonstrate that whilst perceptual aliasing is a problem, appearance-recognition using bag-of-words methods can be used effectively. This demonstrates for the first time that full vSLAM systems are potentially useful for the sewer pipe environment.**

I. INTRODUCTION

Sewer pipes require inspection and maintenance throughout their life-span in order to detect blockages and damage [1]. Technologies for inspecting sewer pipes are reviewed in [2], and include closed-circuit television (CCTV) inspection [3]. In the future, untethered, autonomous, mobile robots have the potential to perform inspection but they would have to accurately localize themselves within the pipe network in order to locate damage and faults for repair. This is a challenging problem because sewer pipes are typically buried underground, preventing the reception of GPS signals. In addition, water utilities' maps of their pipe networks can often be incomplete and contain errors due to discrepancies arising between planned pipe replacement and actual work undertaken, lack of precise record keeping and loss of data (pipes can be tens of years old). Therefore, robots need accurate systems for simultaneous localization and mapping (SLAM) [4]–[6].

Cameras are usually included on pipe inspection robots so that human operators can visually inspect the pipes, including the robots MAKRO [7], KANTARO [8], MRINSPECT [9], PipeTron [10] and EXPLORER [11]. This makes them a natural sensor choice for navigation, and in fact cameras have long been used for localization in pipes, e.g. using image mosaicking [12], and more recently with modern keyframe optimization methods for monocular visual odometry (VO) in natural gas pipes [13], [14] and sewer pipes [15]. Depth cameras have also been used for VO in sewers, which slightly simplifies the problem [16], [17]. In related work, camera-inspection has been used in structure-from-motion (SFM) methods to perform 3D reconstruction of sewer pipes [18].

Whilst the VO problem and the SFM problem have been studied in sewer pipes, the full visual SLAM (vSLAM) problem including loop closing and appearance-based recognition has not yet been well investigated in this type of environment. VO is quite distinct from the full vSLAM problem because the primary goal in a VO system is to only estimate the current pose of the camera, for which a system will tend to use recent frames. In contrast, the goal of a vSLAM system is to estimate a map along with the pose, and the system will tend to use frames across the full time history to perform map and pose estimation. This has the potential to lead to increased accuracy but also increased robustness, because the vSLAM system enables matching to data observed much earlier in time.

The accuracy of robot mapping and localisation is important in sewer pipes because fault detection can lead to expensive and disruptive excavations for repair in busy streets, so it is critical to locate the fault at the first attempt when excavating. Also, robustness is critical because it is important that sewer robots do not become lost in pipes, leading to additional problems such as blockages and costly operations for robot recovery.

A key aspect of vSLAM compared to VO is loop closing, typically using appearance-based recognition methods [19]. Appearance-based recognition would appear to be a challenging problem in sewer pipes because of perceptual aliasing, which is where errors are made in place recognition due to 1. false positives where one place is mistaken for another due to high similarity in appearance, and 2. false negatives where there is a failure to recognise a previously visited place because it's appearance is indistinguishable from other places. However, appearance-based recognition in sewer pipes has not yet been investigated (to our knowledge), which is an important research gap to address, in order to characterise to what extent these methods can succeed.

The aim of this study is to investigate the feasibility of using monocular vSLAM for mapping and localization in

*This work is supported by the UK's Engineering and Physical Sciences Research Council (EPSRC) Programme Grant EP/S016813/1 Pervasive Sensing for Buried Pipes (Pipebots) `https://pipebots.ac.uk`.

[1]Mat Evans, Jonathan Aitken and Sean R. Anderson are with the Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, UK `mat.evans@sheffield.ac.uk`

Fig. 1. Experimental data. (a) The sewer inspection robot used to collect the modern CCTV data. (b) The inspection rig and view along the road where the sewer inspection took place for modern CCTV data. (c) The robot inside the manhole insertion point at the start of the inspection for the odern CCTV data. (d) An alternative view of the above-ground inspection route for the modern CCTV data. (e) A top-down map view of the sewer pipe route for the modern CCTV data. (f) An example image frame from the video recorded by the inspection robot for the modern CCTV data. (g)-(j) Examples of image frames from the historic sewer inspection videos.

sewer pipes, where we place particular emphasis on evaluating the use of appearance-based recognition in loop closing. We make use of state-of-the-art visual SLAM methods - ORB-SLAM3 [20], which is a recent update of the well-known ORB-SLAM2 algorithm [21].

In order to perform this study, we make use of a number of real-world data-sets from CCTV inspection of sewer pipes. These are a relatively unique data sets because they include data from different types of sewer pipe construction, camera technologies and image quality. In addition, it presents challenging problems for visual SLAM because the environment and inspection is real, not a synthetic, lab-based experiment. Hence, lighting, motion and visual features are not in the control of the experimenter and present a more stringent and realistic test of the SLAM system. The videos include varying levels of running water, pipe debris and corrosion.

A limitation of this study is that it uses monocular vSLAM methods, which are subject to scale ambiguity, which makes full quantitative assessment difficult at this stage. However, the setup is sufficient to test the accuracy of appearance-based recognition in terms of successful loop closures, which is the primary aim of the paper, and also to assess feasibility.

The sewer pipe data sets divide into two (see Fig. 1): The first data set is from a recent real-world CCTV inspection where we were able to accompany the sewer inspection team - this enabled us to perform proper camera calibration and make multiple runs for subsequent testing and analysis. The second data set is a collection of historical CCTV inspections, that are from a number of different types of sewer pipe collected over many years. This data set is advantageous for evaluating the robustness of vSLAM methods across different types of sewer pipe environment but due to its historical nature, the opportunity for camera calibration was not available, so the absolute accuracy would be less than normally expected.

II. METHODS

A. *Modern CCTV data collection*

Modern CCTV data was collected by a professional sewer surveying team from Severn Trent Water at a location in Derbyshire under the supervision of the authors. A tethered Mini-Cam Proteus ATEX camera robot crawler system (Fig. 1(a), minicam.co.uk) was lowered into live sewer pipes. Camera images could be inspected in real time from a computer terminal in an inspection van, from which the camera robot was remotely controlled. The robot was driven along

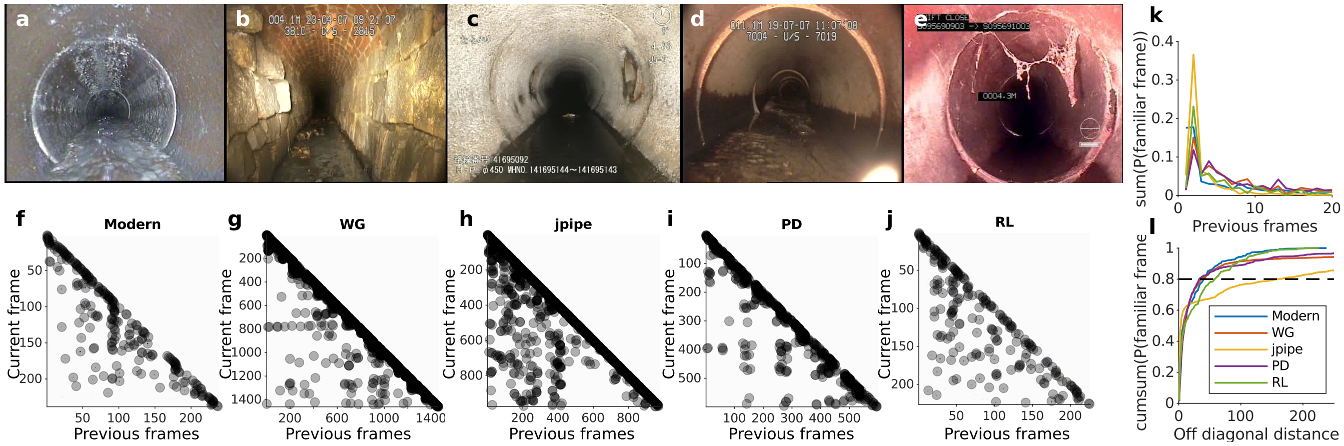

Fig. 2. Appearance-based recognition and mapping in modern and historical CCTV data. (a-e) Example stills from pipe inspection CCTV showing the diverse visual environments, lighting conditions and image quality. (f-j) FAB-MAP results for each example video shown in (a-e). Each row in the image plot shows the probability that the current image has been seen previously. Maximum probability frame for each row is marked with grey dot. Dark line of dots along the diagonal show new frames being correctly identified as novel. Off-diagonal entries show confusion due to aliasing. (k) Summed probability of previous frames being most similar to the current frame (f-j) for sequential frames back in time. (l) Cumulative sum of results in (k).

the pipe from one manhole location to the next manhole inspection chamber, as determined through visual inspection of the camera images (40.48m distance). Distance travelled as measured by the tether encoder was written on to the camera images in real time. This process was repeated twice in the same pipe to provide data to test place recognition and loop closure.

Camera calibration was performed using OpenCV chess-boards (9x6, 71mm and 42mm square diameter) [22] and Matlab (Mathworks) calibration routines. Images for calibration were collected above ground during the site visit by the authors.

B. Historical CCTV data collection

Historical CCTV data was gathered from a number of different sewer inspection teams across many years by colleagues in the Department of Civil Engineering at the University of Sheffield. Survey locations were withheld but are broadly from the South Yorkshire area as well as one video from Japan. These surveys were conducted by experienced camera operators using professional grade equipment for the time, but vary in camera image quality, lighting conditions and camera movement - some involve simple linear movement along the pipe while others feature frequent pans and tilts to allow close visual inspection of pipe defects. In all cases the camera operators made notes of the pipe condition which were ‘burned’ as text on to the videos, posing a further challenge for VO and vSLAM.

C. Data preparation

All videos were manually inspected for occlusion and errors. Selected videos were transcoded to high-quality grayscale PNG images using FFMPEG (`ffmpeg.org`). For ORB-SLAM3 results videos were transcoded at the camera acquisition rate (25fps). FAB-MAP expects images to be from distinct places and not consecutive video frames, therefore videos were down-sampled to 1 frame per second to reduce redundancy from frame to frame.

D. The visual SLAM system

To perform visual SLAM in this paper we use the ORB-SLAM3 algorithm, which consists of three parallel threads of 1. tracking, 2. local mapping, and 3. loop closing. The algorithm uses the ORB feature for both tracking and mapping, which is fast to compute, rotation invariant and provides good invariance to different viewpoints.[1] The ORB-SLAM3 algorithm uses the same main elements as ORB-SLAM2 [21], whilst providing additional functionality for features such as multiple maps - the core functionality is described below.

1) Tracking: The tracking part of ORBSLAM3 performs pose estimation using feature matching (with ORB features) and bundle adjustment.

Initial pose estimation is either performed using the previous frame (if the previous frame was successfully tracked) using a constant velocity motion model, or if tracking has been lost, using global relocalization.

After the initial camera pose estimation, a local map is projected into the frame, where the local map is defined as the set of keyframes that shares map points with the current frame. The pose is then optimised with respect to all the local map points found in the frame.

In the last step of this stage, the current frame can be defined as a new keyframe if it satisfies a number of conditions, i.e. 1. that more than 20 frames have passed since last global relocalization, 2. that more than 20 frames have passed since last keyframe insertion, 3. that the current frame contains at least 50 points and 4. the current frame tracks less than 90% of the last keyframe.

[1]The results in this paper were obtained using the open source implementation of ORB-SLAM3 at `https://github.com/UZ-SLAMLab/ORB_SLAM3`

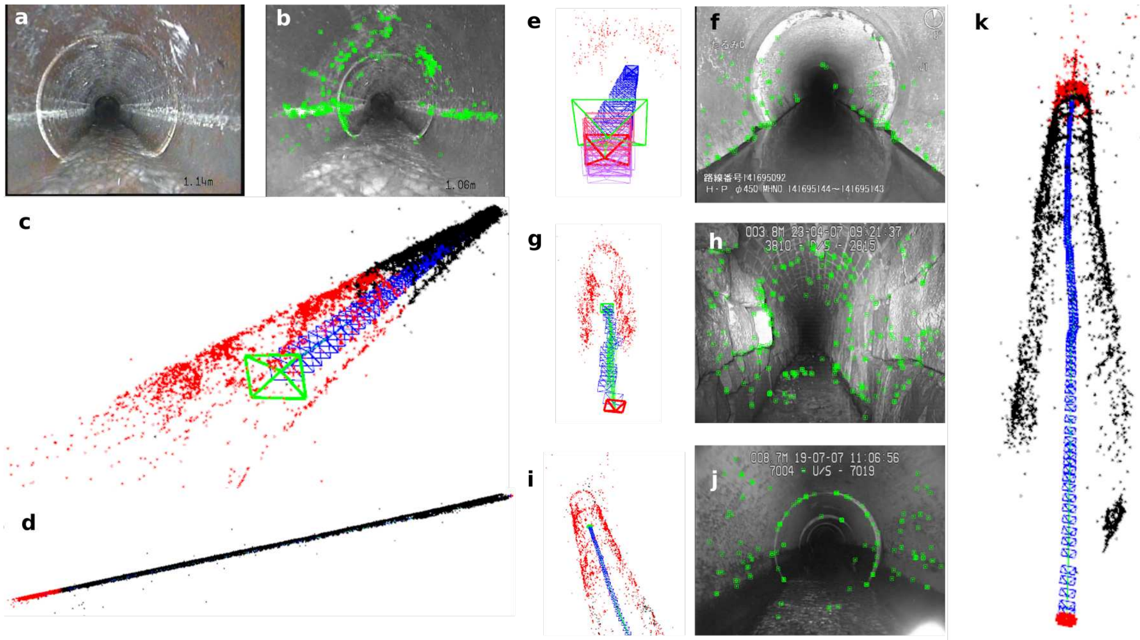

Fig. 3. Visual odometry on modern and historical CCTV data. (a) Example video frame from the modern CCTV dataset. (b) Frame in (a) with detected ORB features overlaid in green. (c) Example pointcloud (black/red points) and keyframe trajectory (blue/green) from the video in (a). Red points are 'active' reflecting points from the current and most recent frames, black points are older and saved as a map for future use. (d) Point cloud at the end of tracking the full 40m length of pipe, showing a straight cylinder shape reflecting good performance. (e-k) Evaluating ORB-SLAM3 performance on historical CCTV data. (e,g,i) Point clouds and trajectories vary in quality from video to video. Purple keyframes in (e) reflect failed and re-initialized tracking for video 'JP'. (f,h,j) ORB features were successfully extracted from all examples. (k) Example high-quality straight trajectory and cylindrical point cloud map from historical video 'PD'.

2) Local mapping: The local mapping stage processes a new keyframe denoted by K_i.

First, the new keyframe K_i is transformed into a bag of words (BoW) representation, which is then added to a covisibility graph. The covisibility graph is an undirected, weighted graph, where each node is a keyframe and an edge between two keyframes exists if they share a minimum number of map points (at least 15).

In order to make the map more robust and remove outliers due to incorrect data assocation, map points are deleted if they do not pass a test during the first three keyframes after creation. The test conditions are: 1. the tracking stage must find the point in more than 25% of frames where the point is predicted to be visible and 2. the point must be observed from at least three keyframes after map point creation.

New map points are created by triangulating the ORB features from the new keyframe K_i and the set of keyframes $\mathcal{K}_c$ that are connected to K_i in the covisibility graph. In the case of unmatched points in K_i, previous keyframes are searched for matches.

Local bundle adjustment is used to process the new keyframe K_i as well as the set of all keyframes connected to it $\mathcal{K}_c$ in the covisibility graph and all map points included in those keyframes.

In order to maintain a compact set of keyframes the local mapping stage attempts to detect redundant keyframes and delete them. Redundancy is assessed by checking the number of repeated map points: those keyframes in $\mathcal{K}_c$ that contain 90% of map points that are in at least three other keyframes are discarded.

3) Loop closing: Loop closing is performed on the newest keyframe K_i using appearance-based recognition with the BoW method and proceeds through four substages: loop candidate detection, a similarity transformation test, fusion of duplicated map points and pose graph optimization.

4) Implementation: In practice ORB-SLAM3 was executed with default parameters with the exception of camera properties (derived through calibration where possible, or manual search) and the number of ORB image features. On the modern CCTV data feature numbers could be varied from 700 up to 10000 with little affect on performance (data not shown). On the historical CCTV in some cases higher feature numbers (2000-5000) were necessary for successful initialization. For some videos no combination of camera or execution parameters could be found to allow successful initialization.

E. Appearance-based mapping

In order to test appearance-based visual recognition and mapping in isolation from the full visual SLAM system, we used a BoW method, similar to that used in ORB-SLAM3, which is FAB-MAP [23]. This enabled us to study the effectiveness of appearance-based recognition without the complicating factors of visual odometry, and therefore better understand the sensitivity of this approach to the sewer environment. Analysis of FAB-MAP performance was done in Matlab, using the open source code implementation available at `robots.ox.ac.uk/~mjc/Software`.

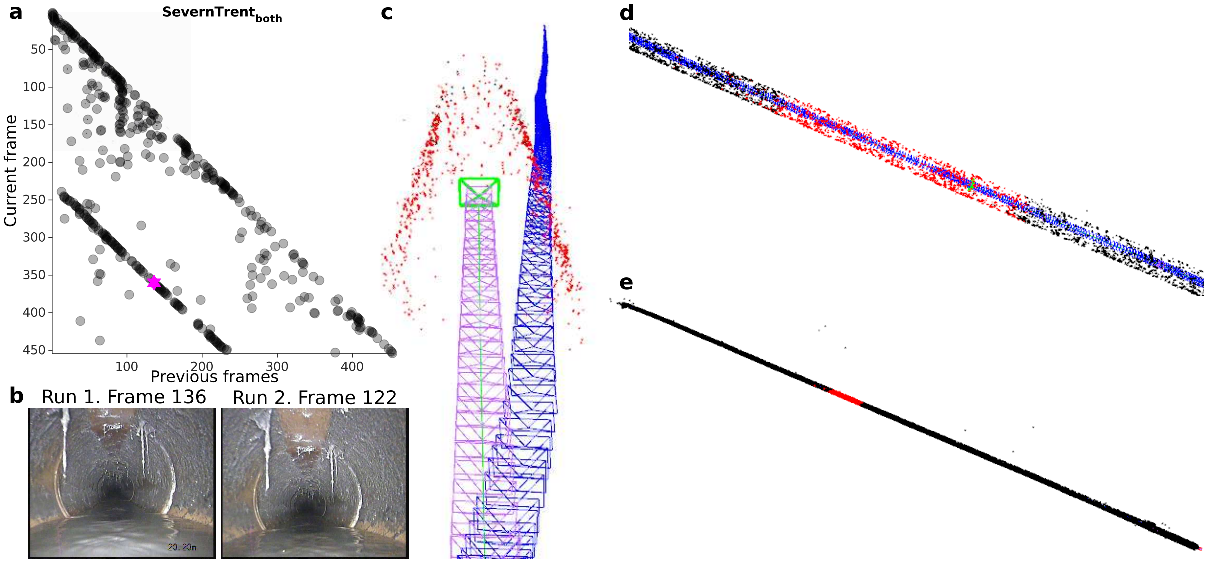

Fig. 4. Full vSLAM with loop closure on modern CCTV data. (a) Appearance-recognition result matrix (as in 2) for two separate camera runs down the same pipe (238 images from run 1, 216 in run 2). Dark grey circles indicate the most likely previous frame encountered. Strong off diagonal line in the bottom left corresponds to loop closure - frames identified on run 2 that are very similar to frames in run 1. (b) Example loop closure frames corresponding to the magenta star in (a). (c) Keyframe trajectory (purple) and feature point cloud (red) of run 2 prior to recognition of a loop closure. Run 1 keyframe trajectory shown in blue. (d) vSLAM example following successful loop closure on run 2. Currently 'active' feature points (red) are aligned to previously encountered points (black). (e) As in (d) zoomed out to show the full extent of the point cloud map.

III. RESULTS

A. Appearance-based mapping

FAB-MAP was used to determine the feasibility of appearance-based recognition for localization and mapping in buried pipes. FAB-MAP returns the probability that the current image (video frame) is a new place or one that has been seen before. If aliasing is a significant problem in this environment, FAB-MAP will fail to recognize new image frames as novel places. Fig 2 (f-j) shows the results of applying FAB-MAP to pipe inspection videos from both of our datasets. In each case the majority of images are recognised as novel places, as indicated by the prominent dark grey line along the main diagonal. To quantify this result further we computed the summed probability that a given frame was novel, or the same as a previous frame for each time point sequentially into the past (Fig 2(k). Perfect performance would result in high probability for current/previous frame 1 and zero probability for older frames, as is broadly shown in the figure. The cumulative probability that a new frame is correctly identified as a novel place exceeds 80% within 50 frames for most videos (Fig 2(i)).

B. Visual odometry

We next tested the performance of visual odometry with ORB-SLAM3 on CCTV data (Fig 3). First we applied ORB-SLAM3 to the modern CCTV dataset. Following camera calibration ORB-SLAM3 successfully extracted ORB features along the pipe wall and ignored the running water (see Fig 3 (b)). The algorithm constructed a cylindrical point cloud and a straight trajectory for the full 40m length of pipe, corresponding to the ground truth data - note that due to the scale-free monocular setup exact quantification of the estimated distance travelled was not possible here. However, the results demonstrate the feasibility of using VO in this environment (Fig 3(c,d)).

Next we tested whether visual odometry would be possible on historical CCTV data where camera calibration is not possible, and camera motion and lighting is highly variable. Here ORB-SLAM3 showed mixed results. In some cases good camera and execution parameters could be found allowing point cloud maps of the environment to be made and camera trajectories to be recovered (Fig 3(g-k)). In other cases, due to erratic camera movement, image occlusion or poor image quality, no combination of camera or execution parameters could be found. Tracking would either fail to initialize, or tracking would fail a short way into the video (see Fig 3(e)). It may be possible to 'stitch' failed trajectories together in post-hoc analysis but that is beyond the scope of this work.

C. Loop closure and vSLAM

Finally we tested the feasibility of loop closure and full vSLAM in live sewer pipes. First we tested whether loop closures would be detected in FAB-MAP by concatenating two videos from the modern CCTV dataset corresponding to two runs of the CCTV camera robot down the same pipe (238 images from run 1, 216 images from run 2). Fig 4(a) shows the results of this analysis. Images from run 2 are correctly recognised as the same place as the corresponding images in run 1 (see the prominent off-diagonal line of dots in the lower left segment of Fig 4(a)). An example of two matched images from the two runs - corresponding to the magenta star in Fig 4(a) - are shown in Fig 4(b).

ORB-SLAM3 also successfully detected loop closures on the modern CCTV dataset. Following accurate mapping of the pipe on run 1 the algorithm re-initializes at the start of run 2 (Fig 4(c)). Following a short duration of tracking the algorithm recognizes frames in run 2 as being the same place as in run 1 and aligns the two point clouds and trajectories (Fig 4(d)). This close correspondence is maintained throughout the remainder of the 40m trajectory, as shown by the alignment of currently 'active' points from run 2 (red) within the previously constructed point cloud map from run 1 (black) in Fig 4(d,e).

IV. Discussion and Concluding Remarks

In this paper we set out to determine whether camera-based monocular vSLAM approaches could be used for inspecting real-world sewer pipe systems, with a goal of using autonomous mobile robots for localising defects in the future. We tested three approaches - appearance-based mapping, visual odometry and visual SLAM - on a range of modern and historical pipe inspection videos. We showed that across different lighting and camera conditions there are sufficient visual features to allow recognition and separation of different images to afford robust localization and mapping.

We have shown that perceptual aliasing is a problem in these pipe environments - FAB-MAP performance is not perfect - but that perhaps surprisingly appearance-based recognition can work successfully the majority of the time on the data used here. Combining ego motion with appearance based loop closure (as ORB-SLAM3 does) seems necessary for robust localization in these challenging pipe environments.

Going forward it will be important to optimize lighting and camera equipment choice for the pipe environment. Hardware choices will become especially important for successful deployment on mobile robots in smaller diameter pipes. Finally, there is a pressing need to quantify the performance of vSLAM methods in sewer pipes. In future work we aim to resolve the scale ambiguity problem and determine localization accuracy using a combination of geometric information, and multi-sensor data fusion, e.g. using IMUs and wheel odometry. These methods will need to be applied on multiple runs along the same pipes to assess the limits of vSLAM accuracy in real-world buried pipe environments over realistic survey lengths.

Acknowledgments

We thank Sue Coulthard and her team at Severn Trent for help collecting the modern CCTV survey data, and we also thank Prof. Simon Tait and Dr Will Shepherd at the Department of Civil Engineering, University of Sheffield, for providing the historical CCTV data.

References

[1] P. Rizzo, "Water and wastewater pipe nondestructive evaluation and health monitoring: A review," *Advances in Civil Engineering*, vol. 2010, pp. GEN–97–1–10, 05 2010.

[2] O. Duran, K. Althoefer, and L. D. Seneviratne, "State of the art in sensor technologies for sewer inspection," *IEEE Sensors Journal*, vol. 2, no. 2, pp. 73–81, 2002.

[3] R. Wirahadikusumah, D. M. Abraham, T. Iseley, and R. K. Prasanth, "Assessment technologies for sewer system rehabilitation," *Automation in Construction*, vol. 7, no. 4, pp. 259–270, 1998.

[4] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robotics and Automation Magazine*, vol. 13, no. 2, pp. 99–108, 2006.

[5] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): Part II," *IEEE Robotics & Automation Magazine*, vol. 13, no. 3, pp. 108–117, 2006.

[6] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[7] E. Rome, J. Hertzberg, F. Kirchner, U. Licht, and T. Christaller, "Towards autonomous sewer robots: the MAKRO project," *Urban Water*, vol. 1, no. 1, pp. 57–70, 1999.

[8] A. A. F. Nassiraei, Y. Kawamura, A. Ahrary, Y. Mikuriya, and K. Ishii, "A new approach to the sewer pipe inspection: Fully autonomous mobile robot "KANTARO"," in *Proceedings of the 32nd Annual Conference on IEEE Industrial Electronics (IECON)*. IEEE, 2006, pp. 4088–4093.

[9] S. Roh, J.-S. Lee, H. Moon, H. R. Choi, and Others, "Modularized in-pipe robot capable of selective navigation inside of pipelines," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008, pp. 1724–1729.

[10] P. Debenest, M. Guarnieri, and S. Hirose, "PipeTron series - Robots for pipe inspection," in *Proceedings of the 3rd International Conference on Applied Robotics for the Power Industry (CARPI)*, 2014, pp. 1–6.

[11] H. Schempf, E. Mutschler, A. Gavaert, G. Skoptsov, and W. Crowley, "Visual and nondestructive evaluation inspection of live gas mains using the explorer™ family of pipe robots," *Journal of Field Robotics*, vol. 27, no. 3, pp. 217–249, 2010.

[12] D. Krys and H. Najjaran, "Development of visual simultaneous localization and mapping (VSLAM) for a pipe inspection robot," in *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA)*, 2007, pp. 344–349.

[13] P. Hansen, H. Alismail, P. Rander, and B. Browning, "Pipe mapping with monocular fisheye imagery," *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, pp. 5180–5185, 2013.

[14] P. Hansen, H. Alismail, P. Rander, and B. Browning, "Visual mapping for natural gas pipe inspection," *International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 532–538, 2015.

[15] J. Künzel, T. Werner, P. Eisert, J. Waschnewski, R. Möller, and R. Hilpert, "Automatic Analysis of Sewer Pipes Based on Unrolled Monocular Fisheye Images," in *Proc. 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018*, vol. 2018-Janua, no. March, 2018, pp. 2019–2027.

[16] D. Alejo, F. Caballero, and L. Merino, "RGBD-based robot localization in sewer networks," in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*. IEEE, 2017, pp. 4070–4076.

[17] D. Alejo, F. Caballero, and L. Merino, "A robust localization system for inspection robots in sewer networks," *Sensors*, vol. 19, no. 22, 2019.

[18] X. Zhang, P. Zhao, Q. Hu, H. Wang, M. Ai, and J. Li, "A 3D reconstruction pipeline of urban drainage pipes based on multiviewimage matching using low-cost panoramic video cameras," *Water (Switzerland)*, vol. 11, no. 10, p. 2101, 2019.

[19] S. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, "Visual place recognition: A survey," *IEEE Transactions on Robotics*, vol. 32, no. 1, pp. 1–19, 2015.

[20] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An accurate open-source library for visual, visual-inertial and multi-map SLAM," *IEEE Transactions on Robotics*, DOI: 10.1109/TRO.2021.3075644.

[21] R. Mur-Artal and J. D. Tardos, "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, oct 2017.

[22] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[23] M. Cummins and P. Newman, "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *The International Journal of Robotics Research*, vol. 27, no. 6, pp. 647–665, 2008.