University of Strathclyde

Department of Engineering

Image-based 3-D reconstruction of constrained environments

Kristofer Law

A thesis submitted for the degree of
Doctor of Philosophy

July 29, 2020

Copyright

Signed: KRISTOFER LAW Date: 19/06/2020

Acknowledgements

I would like to take the time to first thank my primary supervisor Dr Graeme West for giving me the opportunity to work under his wing, teaching me the intricacies of Nuclear Engineering, continually helping me throughout the never-ending Ph.D process and helping review this thesis. Secondly, I would like to thank Dr Paul Murray for his contributions and his expansive image processing knowledge which helped direct the research direction undertaken. As a whole, I would like to thank the researchers and academic staff in the Advanced Nuclear Research Centre for helping me extend my knowledge into different domains with thought provoking discussions and providing a good laugh every now and then when it was very much needed.

Outside the realms of academia, I would first like to thank my family - William, Margaret and Craig - for their long and unyielding support throughout all my long and hard years in education. I would also like to thank Alyshia for her undying and never-ending support during the Ph.D process, providing a much needed source of motivation and distraction during the trying times. Furthermore, I would like to give a thanks to my friends Calum, Gavin and Stuart for their help and support. Finally, I would like to extend a final thanks to EDF Energy, in particular Chris Lynch, and the Ronald Miller scholarship for their technical assistance with regards to the complexities of nuclear reactor inspection and reviewing my publications in a timely manner and for funding the work documented in this thesis.

I was once asked what the abbreviation “Ph.D” stood for during discussions with a previous Masters supervisor Professor John Soraghan. He proclaimed that it stood for “Permanent Head Damage”. In hindsight, he was most definitely right.

Abstract

Nuclear power plays a important role to the United Kingdom electricity generation infrastructure, providing a reliable baseload of low carbon electricity. The Advanced Gas-cooled Reactor (AGR) design makes up approximately 50% of the existing fleet, however, many of the operating reactors have exceeding their original design lifetimes. To ensure safe reactor operation, engineers perform periodic in-core visual inspections of reactor components to monitor the structural health of the core as it ages. However, current inspection mechanisms deployed provide limited structural information about the fuel channel or defects. This thesis investigates the suitability of image-based 3-D reconstruction techniques to acquire 3-D structural geometry to enable improved diagnostic and prognostic abilities for inspection engineers. The application of image-based 3-D reconstruction to in-core inspection footage highlights significant challenges, most predominantly that the image saliency proves insufficient for general reconstruction frameworks. The contribution of the thesis is threefold. Firstly, a novel semi-dense matching scheme which exploits sparse and dense image correspondence in combination with a novel intra-image region strength approach to improve the stability of the correspondence between images. This results in a percentage increase of 138.53% of correct feature matches over similar state-of-the-art image matching paradigms. Secondly, a bespoke incremental Structure-from-Motion (SfM) framework called the Constrained Homogeneous SfM (CH-SfM) which is able to derive structure from deficient feature spaces and constrained environments. Thirdly, the application of the CH-SfM framework to remote visual inspection footage gathered within AGR fuel channels, outperforming other state-of-the-art reconstruction approaches and extracting representative 3-D structural geometry of orientational scans and fully circumferential reconstructions. This is demonstrated on in-core and laboratory footage, achieving an approximate 3-D point density of 2.785 - 23.8025N_X/cm^2 for real in-core inspection footage and high quality laboratory footage respectively. The demonstrated novelties have applicability to other constrained or feature-poor environments, with future work looking to

producing fully dense, photo-realistic 3-D reconstructions.

Contents

List of Figures

List of Tables

Nomenclature

Acronyms

3-D	Three Dimensional
ABWR	Advanced Boiling Water Reactor
ACR	Advanced CANDU Reactor
AGR	Advanced Gas-Cooled Reactor
ASIST	Automated Software Image Stitching Tool
BA	Bundle Adjustment
BWR	Boiling Water Reactor
CANDU	Canada Deuterium Uranium reactor
CBIU	Channel Bore Inspection Unit
CBMU	Channel Bore Measurement Unit
CRVW	Core Restraint Viewing Winch
DMS	Dense Motion Statistics
DSLR	Digital Single-Lens Reflect
EPR	Evolutionary Power Reactor
EXIF	Exchangeable Image File
FBR	Fast Breeder Reactor
FGLT	Fuel Grab Load Trace
FNR	Fast Neutron Reactor

FoR	Field-of-Regard
FoV	Field-of-View
FPS	Frame(s) Per Second
GCR	Gas-Cooled Reactor
GPU	Graphics Processing Unit
LWGR	Light Water Graphite Reactor
MAGNOX	MAGNesium OXide reactor
MAP	Monitoring Assessment Panel
MIFM	Motion-based Iterative Feature Matching
MKER	Mnogopetlevoy Kanalynyi Energeticheskiy Reaktor
MVS	Multi-View Stereo
MVS	Multi-View Stereo
NDI	Nondestructive Inspection
NICIE	New In-core Inspection Equipment
PAL	Phase Alternating Line
PHWR	Pressurised Heavy Water Reactor
PWR	Pressurised Water Reactor
RANSAC	RANdom Sampling and Consensus
RBMK	Reaktor Bolshoy Moshchnosti Kanalniy
ROI	Regions of Interest
RPV	Reactor Pressure Vessel

RVI	Remote Visual Inspection
RVI	Remote Visual Inspection
SAD	Sum of Absolute Differences
SfM	Structure-from-Motion
SfS	Shape-from-Shading
SIFT	Scale-Invariant Feature Transform
SLAM	Simultaneous Localisation and Mapping
VO	Visual Odometry
VVER	Water-Water Energy Reactor
ZNCC	Zero-mean Normalised Cross Correlation

Chapter 1

Introduction

1.1 Motivation

In the United Kingdom (UK), there are 7 Advanced Gas-cooled Reactor (AGR) nuclear power stations currently in operation (see Figure 1.1), which are close to or have already exceeded their original, conservative, lifetime design estimations. Around 50% of these operating reactors planned to be shut down within the next decade [1,2]. The life limiting factor of these reactors is the condition of the graphite reactor core bricks, which act as structural housing for the fuel and perform moderation of the nuclear reactor [3]. The primary method of assessing the graphite bricks is through Remote Visual Inspection (RVI). RVI is deployed throughout the nuclear industry for a variety of tasks such as; defect detection [4,5] and repair [6,7], accident response and reactor decommissioning [8,9] across many reactor designs such as the AGR, the Canada Deuterium Uranium (CANDU) reactor, Boiling Water Reactor (BWR) and the Pressurised Water Reactor (PWR). Each reactor has unique challenges which requires the development of application-specific equipment for deployment in physically constrained and radioactive environments.

Inspection in the nuclear industry is implemented with a diverse set of methodological approaches such as the use of ultrasonic tools to investigate weld joints [10] in the Reactor Pressure Vessel (RPV) and defects in CANDU pressure tubes [4], bespoke underwater robotic camera systems to investigate the PWR RPV [7] and lower core plates [6], pipe inspection robots for feeder pipes in the Pressurised Heavy Water Reactor (PHWR) and tethered inspection tools used to gather video and sensory data about the AGR fuel channels and control rods [2,3,11]. All of the antecedent approaches share a common goal; to gather evidence and facilitate the understanding of operational

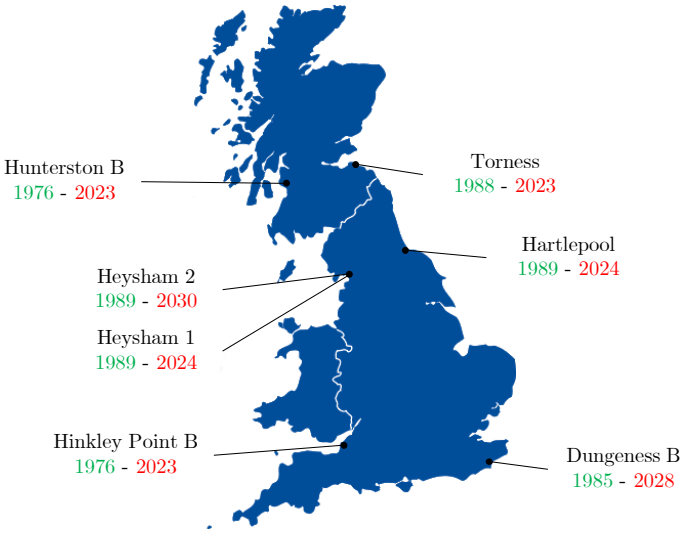


Figure 1.1: Estimated operational lifespan of the United Kingdom AGR fleet

degradation of reactor components. Inspection and the importance of gaining an understanding of rigid structures extends beyond the nuclear industry and applications are found in the oil and gas industry to inspect pipelines [12–14], sanitation for sewer pipe inspection [15, 16], archaeological and cultural heritage preservation [17, 18] and reconstructions of urban environments [19, 20]. As technology continues to advance, computationally complex tasks such as rendering a scene or environment in 3-D is becoming common-place and is being increasingly adopted as part of routine health assessment strategies.

Deriving 3-D structural geometry of a scene can be performed using image processing techniques such as Structure-from-Motion (SfM) [21] and Simultaneous Localisation and Mapping (SLAM) [22] . Both methods derive 3-D structural information of the observed environment in addition to an estimation of the camera pose relative to the reconstructed 3-D scene either off-line or in real-time respectively. A simplified overview of this approach can be seen in Figure 1.2. These techniques fundamentally operate by determining geometrically verified correspondences in patterns between the unordered or spatially coherent input images which observe the same object or scene. Derivation of the camera pose and the corresponding 3-D points is then calculated, refined via

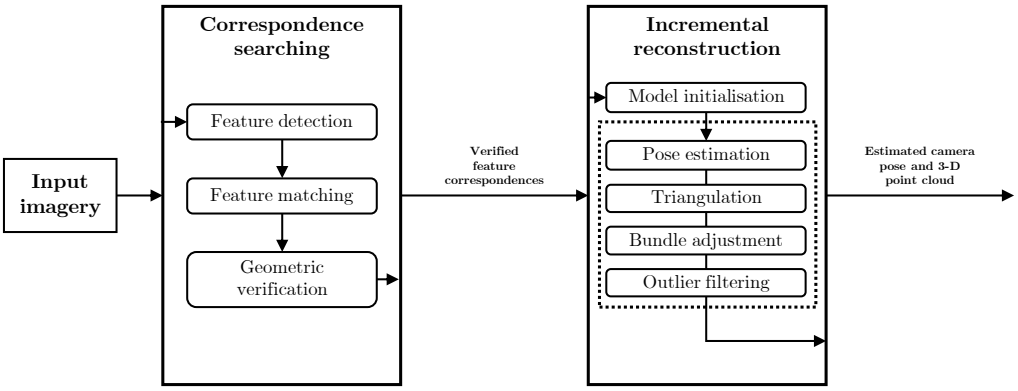


Figure 1.2: Generalised pipeline for an incremental SfM framework

bundle adjustment to produce a representative 3-D point cloud and camera motion. Due to the sequential nature of the pipeline, errors induced during the correspondence stage will result in a compounded error at reconstruction, resulting in techniques such as SfM and SLAM techniques being unable to triangulate 3-D locations precisely resulting in an ambiguous and misrepresentative 3-D point cloud. This is a key research challenge tackled in this work.

Due to the inherent complexity of 3-D scene reconstruction techniques, there is an extensive criterion of input requirements in order to produce an optimal 3-D reconstruction, independent of the target scene or application area. When applying techniques such as SfM, the key considerations and challenges can be classified as issues with the acquisition or processing of the source image data. First and foremost, the image capture process and the associated hardware and sensory equipment must be sufficient for the operational environment. The second consideration is to establish that the utilised hardware is extensively calibrated to ensure that bias and distortion originating from the hardware can be minimised. Furthermore, the capture of ground truth data using RGB-D passive image sensors or Light Detection and Ranging (LiDAR) active sensors [23] is extremely valuable as it can be used to quantify reconstruction and the estimated camera pose accuracy. With these acquisition considerations taken into account, the process of reconstructing the target scene can be greatly simplified.

However, even if all the above considerations are implemented at the acquisition phase, there is still many challenges with regards to processing and subsequently deriving 3-D scene geometry from the image data. The fundamental challenge of producing an authentic 3-D reconstruction is to obtain an robust and accurate set of correspondences between input images. If the acquired image data is *feature-rich* (i.e. each image observes an environment which comprises of unique, salient textural characteristics), the correspondence searching mechanism can reliably identify correspondences between images culminating in a dense and representative 3-D point cloud reconstruction of the rigid object or scene. Conversely, when these techniques are applied to *feature-poor* imagery (i.e. each image observes an environment that is non-descript, repetitive or contains textural uniformity), the correspondence mechanism becomes unreliable, often mismatching detected features and consequently producing erroneous estimations of the 3-D geometry and the reciprocal camera pose and displacement. In some severe cases, techniques such as SfM and SLAM may fail to initialise the reconstruction phase altogether and this is fundamentally a secondary technical challenge which is tackled in this work.

1.2 Research Objective

The research objective of the undertaken work is to determine if an approach could be developed to produce 3-D reconstructions from source footage which is feature-poor and deficient in features. This objective was demonstrated through application to AGR in-core inspection footage gathered within physically constrained and poorly lit environments, where the source footage is unsuitable for assessment using current state-of-the-art image-based 3-D scene reconstruction techniques. Current visualisation techniques deployed to inspect and analyse the AGR fuel channels compose of manually stitched image montages of Regions of Interest (ROI) [24] or 2-D panoramic images of the full AGR fuel channel [2]. A time-line of the inspection approaches is shown in Figure 1.3. Nonetheless, both techniques provide limited 3-D structural information which can be extremely valuable when assessing the fuel channel structure and quantifying

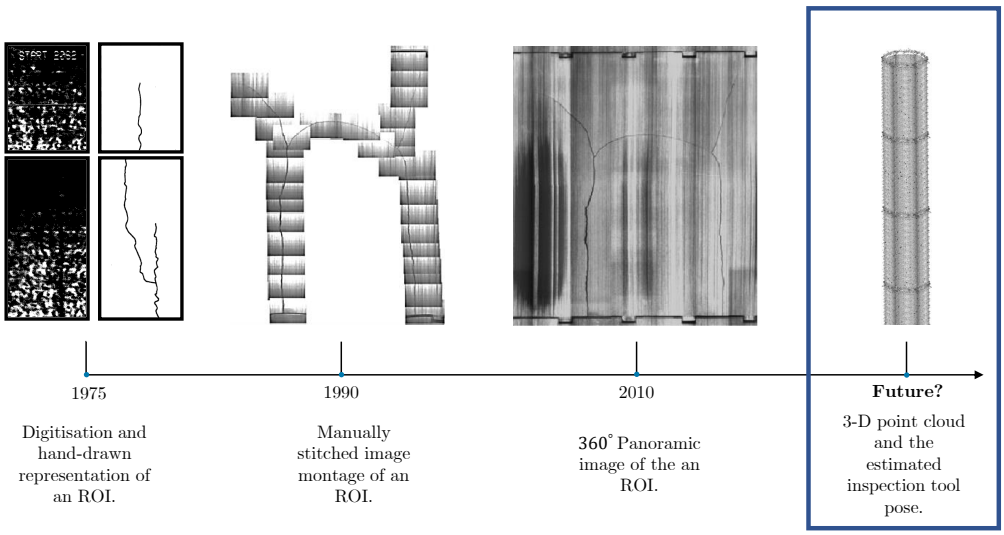


Figure 1.3: Estimation of the timeline of adopted AGR visualisation methodologies. The future approach highlighted in blue is demonstrated within this thesis

structural defects.

However, the existing RVI setup presents a number of challenges; Firstly, the inspection tool has not been designed for the application of 3-D scene reconstruction techniques such as SfM and therefore lacks the hardware calibration or ground truth measurement necessary for 3-D scene reconstruction. Secondly, the hazardous environment within the core severely limits the hardware which can be deployed due to the radiation tolerance required limiting the quality and resolution of the captured data. Furthermore, installation of new inspection equipment is prohibitively expensive. Thirdly, access to the AGR fuel channel is tightly constrained and only a small part of the fuel channel is visible at any given time which in turn limits the applicability of 3-D scene reconstruction techniques such as SfM. Moreover, the interior texture of Gilsocarbon graphite bricks is repetitive and contains a high degree of textural uniformity which leads to feature-poor footage.

In this thesis, the research objective is approached methodically by addressing many of the previously discussed challenges presented by the application of scene reconstruction techniques such as SfM, and allows for the recreation of representative 3-D geo-

metric models of constrictive, non-descript environments as observed within the RVI footage. An in-depth investigation into current state-of-the-art approaches within image processing with particular emphasis on disambiguation of image feature correspondence mechanisms is carried out in Chapter 4.1 - 4.2, with a novel matching scheme called the Motion-based Iterative Feature Matching (MIFM) proposed in Chapter 4.3 to address these issues whilst providing a robust candidate set of feature correspondences. A novel SfM framework, anglicised as the Constrained Homogeneous Structure-from-Motion (CH-SfM) framework, is proposed in Chapter 5 and Appendix A which integrates the MIFM approach with an incremental SfM approach [21, 25] to mitigate challenges associated with determining 3-D geometric models of physically constrained and ambiguous objects or scenes. The CH-SfM framework is designed primarily to function using pre-existing RVI footage acquired within the AGR fuel channels and is demonstrated to operate on other datasets with similar, challenging environments in addition to laboratory footage obtained by a novel experimental apparatus to replicate a sub-section of the AGR fuel channel as shown in Chapter 6.3. The framework also comprises of an additional reconstruction refinement procedure described in Chapter 6.4, which unifies reconstructions obtained from multiple orientations into a singular, cohesive circumferential reconstruction which is subsequently denoised to provide a representation of the AGR fuel channel structure. The proposed MIFM scheme is evaluated via a distortion analysis provided in Chapter 6.2, providing a significant improvement on deficient imagery against a similar state-of-the-art matching scheme. In Chapter 6.3, the reconstructions produced by the CH-SfM framework demonstrate a considerable improvement over state-of-the-art incremental reconstruction frameworks with regards to extracting representative 3-D structural geometry of the AGR fuel channel from different imaging sensors. The contributions provided within this work are shown visually in Figure 1.4 and described in the next section.

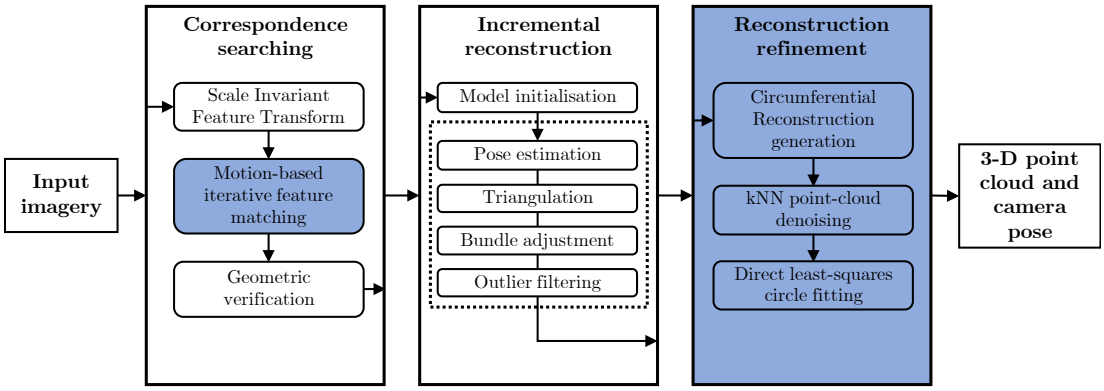


Figure 1.4: Contextualisation of the proposed contributions to knowledge (highlighted in blue) within the developed CH-SfM framework.

1.3 Contributions to Knowledge

- *A novel incremental SfM framework which specialises within constrained, feature-poor environments.*

  The proposed CH-SfM framework is designed to find accurate feature correspondences between homogeneous imagery captured in constrained environments as shown in Figure 1.4. The framework incorporates a novel semi-dense matching scheme as described in Chapter 4 called Motion-based iterative feature matching (MIFM), which spatially discretises the image into regions of stability and in combination with an adaptive thresholding relaxation scheme, produces a robust set of feature correspondences which allows for improved reconstruction quality. The CH-SfM reconstruction performance is quantified in Chapter 5.

- *The first application of SfM to support condition monitoring and visual inspection of nuclear reactor cores.*

  The main contribution is the application and the evaluation of the CH-SfM framework developed using source imagery from real life visual inspection data gathered within the AGR core, in addition to footage gathered from experimental apparatus which emulates three layers of a AGR fuel channel. Consequently, the research

has contributed ideas on how to modify not only the inspection procedure of the AGR fuel channels but to propose possible modifications to the inspection tools themselves so that inspections can be performed and subsequently quantified in a robust and time efficient manner. The CH-SfM framework also incorporates an novel application-specific reconstruction refinement procedure (see Figure 1.4) which allows fully circumferential 3-D reconstructions of the AGR fuel channels to be produced and refined described in Chapter 6.4.

- *An in-depth and comparative analysis of incremental SfM frameworks with a focus on how the proposed framework compares to the current State-of-the-Art.*
  The contribution investigates image-based 3-D reconstruction applied to challenging environments where there is a lack of distinct, salient features either in subsets or throughout the source images. The MIFM approach proposed in Chapter 4.3 is quantified via a distortion analysis performed in Chapter 6.2, comparing its performance against a similar state-of-the-art matching approach on incrementally distorted imagery to determine the "fall-over" point of not only the matching scheme but the CH-SfM framework. Additionally, a comparative analysis is presented using both real in-core and laboratory inspection footage in Chapter 6.3 to compare the performance of the CH-SfM framework to the state-of-the-art with an emphasis on the reproduction of a representative point-cloud of the fuel channel interior surface with particular emphasis on disambiguation. The reconstruction accuracy of the CH-SfM framework is quantified in Chapter 6.3 and 6.4 using an estimation of the AGR fuel channel geometry.

1.4 Publications arising from this work

1. P. Murray, G. M. West, K. Law, S. Buckley-Mellor, G. Cocks, C. Lynch, "Automated video processing and image analysis software to support visual inspection of AGR cores," *in Proceedings of the 5th EDF Energy Generation Ltd Nuclear Graphite Conference*, Southampton, U.K., 2016.

2. K. Law, G. West, P. Murray, and C. Lynch, "3-D advanced gas-cooled nuclear re-

actor fuel channel reconstruction using Structure-from-Motion," in *Nuclear Plant Instrumentation, Control and Human Machine Interface Technologies* (NPIC & HMIT 2017), San Francisco, U.S.A., 2017.

3. K. Law, G. West, P. Murray, and C. Lynch, "Towards extracting 3-D structural representations of AGR core fuel channels from 2-D in-core inspection videos," in International Symposium on Future I&C for Nuclear Power Plants (ISOFIC 2017), Gyeongju, South Korea, 2017.
   [**Awarded the Best Student Paper award.**]

4. K. Law, G. West, P. Murray, and C. Lynch, "3-D Reconstruction of AGR Fuel Channels using RVI footage," *in Proceedings of the 6th EDF Energy Generation Ltd Nuclear Graphite Conference*, Lake District, U.K., 2018.

5. K. Law, G. West, P. Murray, and C. Lynch, "3-D visualization of AGR fuel channel bricks using Structure-from-Motion", *Nuclear Engineering and Design*, Vol. 359, pp. 110472, 2020

1.5 Structure of Thesis

The remainder of this thesis is structured as follows:

Chapter 2 develops the problem definition and the primary application of the research - visualisation of AGR fuel channels to supplement or augment pre-existing visualisation mechanisms. The chapter discusses the operation of an AGR reactor and highlights the need for visual inspection as a direct result of reactor operational degradation, and goes into greater detail about how the reactor fuel channels have been historically visually inspected. The chapter also provides a small discussion on the potential future avenues of nuclear reactor inspection. The chapter concludes with a brief discussion into the issues of applying state-of-the-art visualisation techniques to AGR RVI footage, prefacing the direct need for the research carried out in this document.

Chapter 3 introduces the concept of image-based 3-D scene reconstruction to the reader, discussing the key computational steps in a typical SfM framework. A discussion on contemporary SfM frameworks is provided to the reader, in addition to reviewing application-specific approaches deployed in literature in similar problem spaces. An investigation is then carried out in Chapter 3.4 , applying previously reviewed image-based 3-D reconstruction approaches to AGR RVI footage. The investigation indicates that general approaches in literature are unsuitable for imagery that are largely homogeneous.

Chapter 4 establishes a core and fundamental issue within not only 2-D image registration but the subsequent issues that stem from it. This chapter contextualises the issues faced by generic contemporary frameworks by proposing scenarios where degeneracies occur as a direct result of the characteristics of the input image dataset - such as repetitive texture and structures and image or feature-based ambiguity. The novel Motion-based Iterative Feature Matching (MIFM) approach is proposed in Chapter 4.3, which combines sparse and dense image correspondence in combination with an intra-image, regionalised matching approach to obtain robust feature matches that subsequently improves the resulting reconstruction accuracy.

Chapter 5 introduces a reconstruction framework called the CH-SfM framework and performs validation of the reconstruction framework using synthetic data. This in combination with Appendix A discusses the framework in detail to the user.

Chapter 6 concentrates on the application of the aforementioned CH-SfM framework described in Chapter 5 applied to image data acquired both using bespoke apparatus emulating a sub-section of the AGR fuel channel and real in-core footage. The chapter begins by introducing a distortion analysis in Chapter 6.2 to quantify the performance of the novel MIFM

approach proposed in Chapter 4.3 and the limitations of the encompassing CH-SfM framework in Chapter 5 with regards to source image data. The CH-SfM framework is then evaluated objectively against state-of-the-art incremental reconstruction frameworks in Chapter 6.3, with the reconstruction density and accuracy being quantified using suitable metrics. The premise of circumferential reconstructions are also introduced in Chapter 6.4.

Chapter 7 summarises the work performed within the thesis and its associated contributions, discusses the performance and the associated limitations of the proposed novelties and provides a discussion of potential avenues of future work in order to build upon the work presented in this thesis.

Appendix A acts as a user guide to the bespoke image-based reconstruction framework CH-SfM which integrates the motion-based iterative matching approach described in Chapter 4. The user guide provides a low-level guide on how the parameters used and how the framework operates at a functional level with code-based examples. This appendix serves as an accompaniment to future readers using the CH-SfM framework.

Chapter 2

Visual inspection and condition monitoring of an AGR

2.1 The Advanced Gas-Cooled Reactor

2.1.1 AGR Design & Operation

In the U.K, 7 AGR stations were constructed - Dungeness B, Hinkley Point B, Hunterston B, Hartlepool, Heysham 1 & 2 and Torness - between 1965 to 1980 with each station except Dungeness B being constructed in pairs housing 2 reactors within each station subsequently generating approximately 1000-1300 megawatt electric (MWe) at the start of life. The AGR is regarded as a 2^{nd} generation British reactor design, evolving from the 1^{st} generation MAGNOX (MAGNesium OXide) reactor, which uses a graphite core as a neutron moderator. The graphite moderator comprises of 6000 cylindrical high purity, Giloscarbon graphite bricks, 3000 of which form a lattice structure of approximately 300 vertical channels which house the fuel stringers as visualised in Figure 2.1. These graphite fuel channels are interlocked with smaller cylindrical, hollow interstitial channels of which around 50% are designated as control rod channels with the rest being gas-cooling channels and "filler bricks". To ensure vertical stability and pitch of the reactor design, the lattice structure is interconnected by graphite keys that are intended to limit stress and maintain the structural integrity of the core [26]. Each Giloscarbon carbon brick that make up the fuel channels is approximately 825mm in height with an internal diameter of 263mm at the start of life.

Within each fuel channel is a fuel stringer which comprises of fuel elements that hold 36 stainless steel pins containing enriched Uranium Dioxide (UO_2) ceramic fuel pellets. This is used to superheat high pressure Carbon Dioxide (CO_2) gas using a

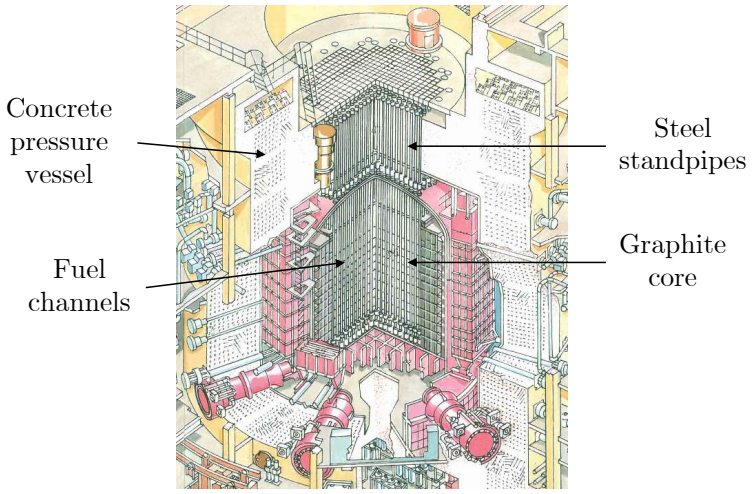


Figure 2.1: Illustrative cross-section highlighting the key components of an AGR core. Courtesy of EDF Energy.

sustained, controlled thermal reaction. The superheated gas passes into the top of the boiler section and through the gas circulator which in turn evaporates the water inside the pressure boiler into high pressure, superheated steam that is used to rotate the turbine blades. With the turbine blades moving, this drives a generator and produces power.

2.1.2 Operational Degradation of the Graphite Core

The graphite core is subject to operational degradation during reactor operation primarily due to two distinct processes; neutron irradiation and radiolytic oxidation which inherently limits the lifespan of the reactor [2,27]. Neutron irradiation, radiolytic oxidation and prolonged exposure to high pressure (CO_2) gas coolant results in dimensional change to the brick in addition to strength and weight loss of the graphite brick which consequently can result in structural defects such as cracks appearing. These changes, if severe enough, could impact neighbouring bricks such as the control rod channels in addition to restricting the movement of fuel within the fuel channels. To monitor the health of the graphite core, online condition monitoring and offline visual inspection is performed to detect and monitor structural defects or changes within the graphite core.

2.2 AGR Graphite Core Monitoring & Inspection

2.2.1 Condition Monitoring

Condition monitoring is one of the 2 approaches deployed understand the health of the reactor and has become increasingly important as the reactors begin to exceed their initial estimated design lifetime. Throughout the AGR lifespan, the reactor condition is assessed by an Monitoring Assessment Panel (MAP) which aggregate many sources of monitoring data and partition them into "Class 1" and "Class 2" parameters [27,28]. The Class 1 parameters are as follows and more information on these parameters can be found in [3,28]:

Fuel Grab Load Trace (FGLT) – As the fuel stringer is inserted and removed (denoted in literature as the charge or discharge) during the refuelling process which occurs every 6-8 weeks, the weight and height of the fuel assembly are measured to infer an approximate structural geometry of the fuel channel [3,28].

Fuel Movements – This parameter is an encapsulation of several properties tied specifically to the refuelling process such as the condition of the spent fuel and the fuel stringer in addition to the monitoring the refuelling machine and the days in which fuel is inserted or removed from the fuel channel.

Control Rod Analysis – Monitoring the control rods is extremely important since it is the reactors primary mode of shut-down in-case of a critical operational failure. Control rods within the AGR can be partitioned into *bulk rods* and *regulating rods* where bulk rods are used for long-term hold down or tripping the reactor and regulating rods are automatically driven rods which aim to balance the output power of the core [29]. By monitoring the height of the regulating rods, it can be used to ascertain if the motion of the control rods is restricted due to damage to the encompassing brick channel as a result of reactor operation.

Channel Power Analysis – This parameter monitors the thermal and neutronic power of each fuel channel within the core. The thermal power is gauged by the

temperature of the (CO_2) coolant and the neutronic power is calculated based on the grade of enriched fuel in combination with the predicted burn-up rates.

2.2.2 Inspection

Manual visual and bore inspection is the second approach to understand the health of the reactor. Manual inspection occurs during planned, periodic outages of the AGR where the status of the reactor is evaluated based on Remote Visual Inspection (RVI) using TV cameras, trepanning/sample removal and bore estimation of a small subset of fuel channels, typically around 10% of the 300 AGR channels which are pre-selected using genetic algorithms [30, 31] based on the parameters described in Section 2.2.1. With this approach, inspection engineers can evaluate the core condition to understand both visually and structurally how the fuel channels are changing and degrading during reactor operation.

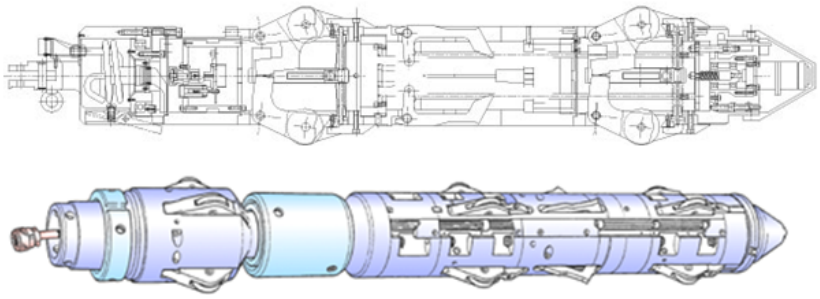


Figure 2.2: Cross-sectional and rendered image of the Channel Bore Measurement Unit (CBMU). Diagram courtesy of EDF Energy.

Bore Measurement and Trepanning

To acquire the approximate structural geometry of the AGR fuel channel, devices such as the Channel Bore Measurement Unit (CBMU) or the New In-core Inspection Equipment (NICIE) are deployed. The following characteristics are extracted using data acquired from the CBMU/NICIE tools:

Channel Diameter – The channel diameter is calculated through the use of 4 feelers placed circumferentially around the inspection tool as seen in Figure 2.2. The scan di-

ameter is calculated as the mean of the diameter measurements of feelers perpendicular to each other.

Ovality – The ovality of the fuel channel is obtained as the difference between the diameter measurements of 2 perpendicular directions.

Tilt – The tilt of the channel is calculated through the use of gyroscopic measurements obtained at the bottom and top of the tool to calculate inclination.

Shrinkage – Shrinkage is calculated as a percentage value change with respect to the measured channel diameter to the Start of Life nominal diameter value.

Visual Inspection

Visual inspection is a comprehensive method of investigating channels of interest. Using specialist tools such as the Channel Bore Inspection Unit (CBIU) and NICIE which contain radiation tolerant cameras, engineers can capture footage to visually assess the appearance of degradation in the channel and further investigate possible anomalies detected from on-line condition monitoring.

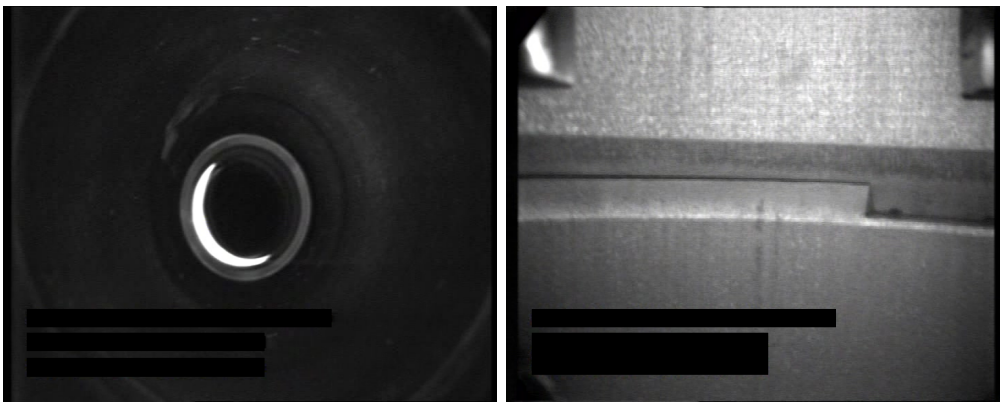


Figure 2.3: ***Left***. RVI footage obtained from the forward-facing camera using inspection tool, ***Right***. RVI footage inspecting a brick interface using the same camera (with a mirror placed at 45°)

The inspection process begins with the engineers first lowering the inspection tool into the channel. During this stage, the forward facing camera captures circumferential footage of the channel walls. As the camera reaches the bottom, a mirror is mechanically

manoeuvred 45° in-front of the camera and changes the Field-of-Regard (FoR) onto the fuel channel surface wall. An example frame from both the downward facing and channel wall footage is shown in Figure 2.3.

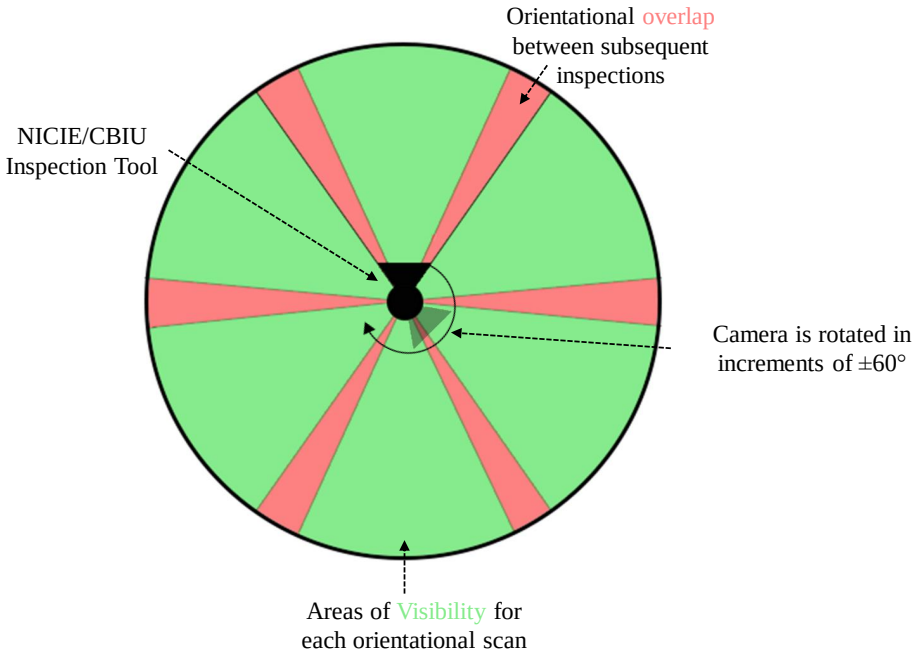


Figure 2.4: Diagram visualising the fuel channel inspection protocol. Green denotes the visibility of each independent vertical scan of the AGR fuel channel and the red highlights areas of visible overlap between neighbouring scans.

Due to the limited Field-of-View (FoV) of approximately 70° provided by the inspection tools, to provide full channel coverage, video must be recorded in successive scans of ±60°. To ensure full coverage, the video strips contain an overlap of 10° to offset any slight rotation introduced whilst traversing the fuel channel. A visualisation of the inspection process can be seen above in Figure 2.4. During the inspection, engineers will take a note of any structural defects such as a crack or general areas of interest which are then further investigated after the initial orientational vertical scans are complete. This footage is termed "crack following footage" in which the inspection engineers will closely evaluate the area of interest and inspect it varying the view-points, lighting conditions and camera parameters to thoroughly inspect the area in question. After inspection, the footage is taken off-site where it is further investi-

gated and quantified using visualisation techniques in order to support the safety case to allow the reactor to be returned to power.

2.2.3 AGR Visualisation - A Historical Overview

During a reactor life-time, the frequency of reactor visual inspections has increased as the core ages. Similarly, the approaches to evaluating and processing the captured data has also evolved. Figure 2.5 shows how the visualisation approaches have evolved over time.

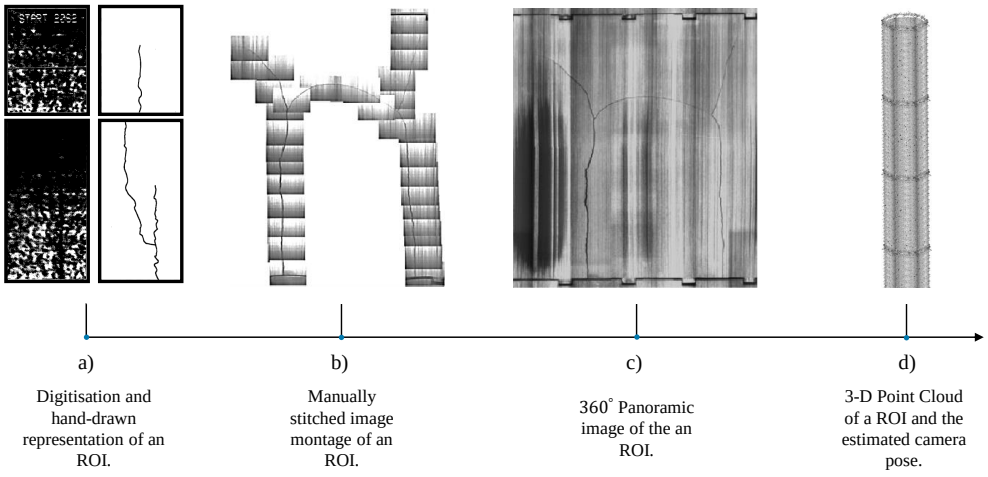


Figure 2.5: A timeline of AGR fuel channel defect visualisation. a) Footage from video tape alongside corresponding hand-drawn interpretations [24].; b) Manually stitched image of a region of interest.; c) A panoramic image of the same defect (cropped to the individual brick layer) [2]; d) 3-D geometry of an AGR fuel channel [32]

Manual interpretation and image stitching was one of the first methods of visual inspection deployed within the AGR reactors by inspection engineers. Approximately 25 years ago, visual inspection engineers would use image stills printed from analog video tape and hand-drawn representations of any defect within the channel to analyse and characterise it [24] . These images (as seen in Figure 2.5.a)) however could take up to a week to generate due to the digitisation process and were limited to defects picked up on individual orientation scan of the AGR fuel channel. As the reactor

has aged, inspection tools such as CBIU, NICIE and NICIE2 have been commissioned to capture visual and other sensory data such as bore diameter measurements. The inspection cameras on these tools provide increased video resolution allowing for a more comprehensive visual inspection which is still used to this day. With the improved visual fidelity, engineers manually stitch images to form individual montages of defects (as seen in Figure 2.5.b)) which can subsequently be studied and analysed to quantify the defect before the station is returned to power, provided it is safe to do so. This method enhanced the visualisation of defects whilst reducing the turn-around time from a week to approximately one working day to generate the images required for further analysis. This is the current method of visualization used during AGR fuel channel inspection.

Recent work has seen to the development of the Chanorama image (Channel Panorama), a 360° panorama generated using video footage captured from within the AGR fuel channels. Chanoramas can be created using the ASIST (Automated Software Image Stitching Tool) software tool to provide inspection engineers with an efficient, repeatable, automatic method for generating defect montages while using all of the available video data to allow full channel visualisation in a snapshot [2]. With ASIST, inspection engineers can automatically generate a full 2-D chanorama of the AGR fuel channel interior in approximately 20 minutes. The ASIST process begins by identifying the first and last frames of each orientation scan. Then, a horizontal window of 5 pixels (corresponding to those which have changed due to the motion of the inspection tool) is extracted from each frame and vertical image strips are formed through accretion of the windowed pixels. Afterwards, each vertical strip is aligned using binary edge detection and cross correlation and subsequently merged to form a chanorama. An image of a single brick layer cropped from a full chanorama can be observed in Figure 2.5.c)). ASIST has now been successfully evaluated in parallel with the existing manual image stitching process and it is anticipated that soon the case will be made to switch from the existing manual process to an automatic one using the software. This represents the current state-of-the-art in reactor core visual data inspection, however in broader inspection approaches outwith nuclear, there is considerable value to be gained

by leveraging structural information from inspection footage. It is at this point where there is the open question of whether similar approaches could be adopted for in-core inspection, given the known constraints on captured image quality.

3-D interpretation of AGR RVI footage was first introduced in West et al [33] through the use of anaglyphic imagery and pivot videos which both produce an illusionary 3-D stereoscopic effect. These techniques can be useful, especially by allowing a user to effectively simulate the act of pivoting around a region of interest within the channel to provide a view from multiple angles and allow inspection engineers to make a more informed decision. However, this technique does not provide any depth information which could prove useful for visualisation of the brick structure or examining the characteristics of defects within the channel. There now follows a concise review of state-of-the-art inspection and monitoring approaches in nuclear, before returning to address the challenges and limitations of extracting 3-D information from the imagery.

2.3 Inspection and Visualisation in the Nuclear Industry

Remote Visual Inspection (RVI) and Nondestructive Inspection (NDI) techniques are effectively deployed across the nuclear industry due to the hazardous and often constrained environments which makes the placement of human operators in Nuclear reactors to perform tasks unsafe. Around the world, there is 455 operational reactors across 31 countries [1] which can be classified under 5 different fundamental reactor types - Pressurised Water Reactor (PWR), Boiling Water Reactor (BWR), Pressurised Heavily Water Reactor (PHWR), Gas-Cooled Reactor (GCR) and Fast Breeder Reactor (FBR) and is tabulated in Table 2.1. Due to structural and reactor component dissimilarity, there is no global solution for reactor inspection and visualisation, with each form of reactor requiring bespoke inspection apparatus and associated visualisation methodologies.

Therefore, in order to inspect and visualise critical reactor components, an assortment of Nondestructive Testing (NDT) equipment has been deployed within the nuclear

[1] This information is accurate as of 03/09/2018 on World Nuclear

Reactor Type	Reactor Designs	Operational	Permanent Shutdown	Under Construction
PWR	Light Water Graphite Reactor (LWGR) Water-Water Energy Reactor (VVER) Evolutionary Power Reactor (EPR)	314	59	45
BWR	Reaktor Bolshoy Moshchnosti Kanalniy (RBMK) Advanced Boiling Water Reactor (ABWR) Mnogopetlevoy Kanalynyi Energeticheskiy Reaktor (MKER)	75	40	4
PHWR	Canada Deuterium Uranium (CANDU) Advanced CANDU Reactor (ACR)	49	8	4
GCR	MAGNOX (MAGNesium OXide) Advanced Gas-cooled Reactor (AGR)	14	38	0
FBR	Fast Neutron Reactor (FNR) Gas/Sodium/Lead cooled Fast Reactor (GFR/SFR,LFR)	3	8	1
		455	**153**	**54**

Table 2.1: Tabulation of civil nuclear reactors currently in service, under construction or permanently shutdown.

industry for a variety of different tasks and applications with regards to NDT to help facilitate a greater understanding of the reactor condition and underlying defects that can arise.

2.3.1 Nondestructive Reactor Inspection

Due to the variance of nuclear reactor designs, there is a large degree of diversity with regards to the life-limited components structure, operating environment and the accessibility. Furthermore, assets of interest for different reactor designs have application specific criterion in order to successfully extend the operational life-time of the component. Within nuclear, inspection is generally used to capture the structural health of the reactor component, but it can also be complimented with other applications such as servicing and repair of components using remote-controlled robots [10, 34, 35] and accident response based inspection [8, 9, 36]. Within literature, many of the reactor inspection approaches are heavily dominated by ultrasonic inspection approaches used to assess a variety of different scene geometries, with many approaches adopting ultrasonic transducer arrays to inspect defects in RPV weld joints [7, 10], CANDU pressure tubes [4, 37] and feeder pipe geometry [38]. This is influenced by many of the reactors currently in operation have a presence of water, filtering out visual inspection as a viable route. Other approaches are also adopted, such as optical radar for 2-D and 3-D reconstruction of the RPV inner surface [39, 40], laser-scanning of a gas baffle in

a MAGNOX reactor [41] and image-based inspection approaches [2, 9, 32] for inspection irradiated in-door environments and image-based assessment of nuclear reactor environments. Generally, image processing approaches are adopted for the harshest of environments where passive sensory equipment is highly susceptible to failure, or used to help guide robotic equipment itself [6, 38, 42] rather than a primary mode of inspection.

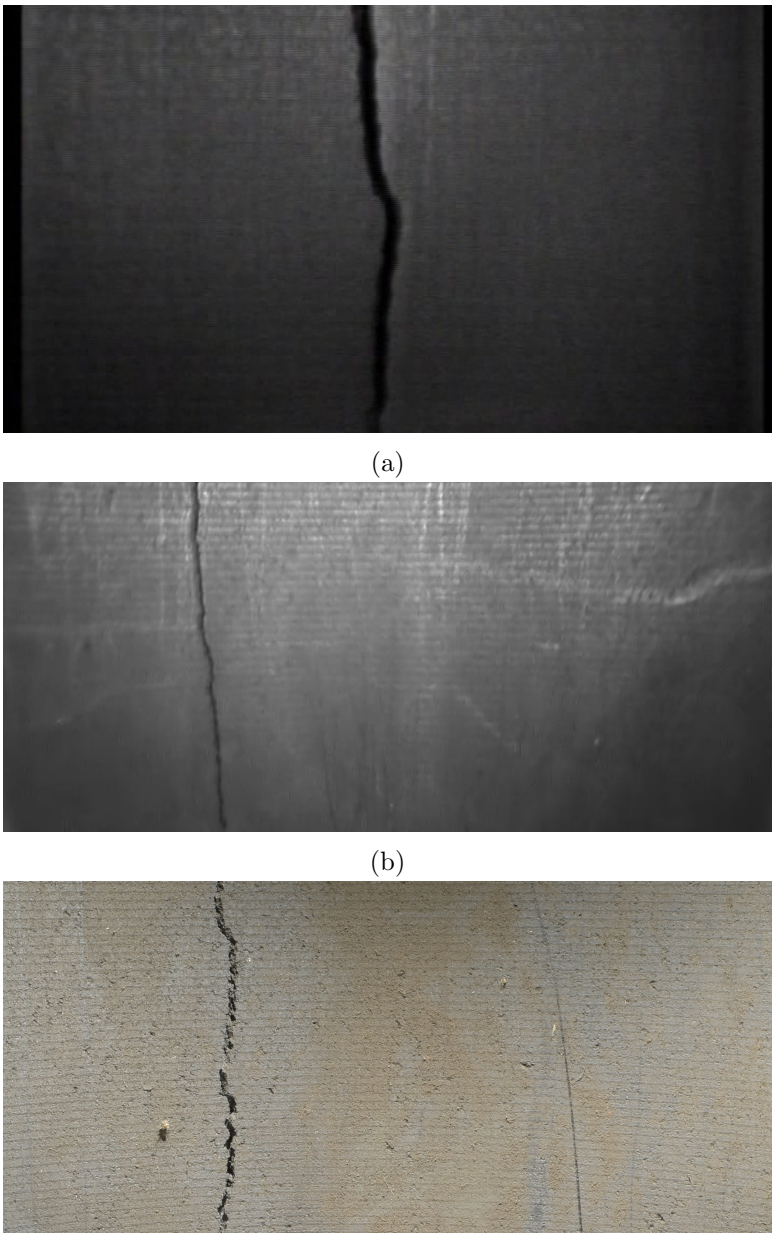
2.4 Limitations and Challenges of 3-D Visualisation of AGR Fuel Channels

2.4.1 Hardware Limitations

The purpose of this section is to describe in detail the inherent limitations from both a hardware and theoretical standpoint. The 1^{st} point of discussion is the most restrictive of the 2 limitations and that is the hardware that is used to inspect the reactor. Fundamentally, the main limitation of deriving 3-D structural geometry of the AGR fuel channel is that the hardware inspection tools such as CBIU or NICIE2 used to inspect the reactor are simply not designed with 3-D image-based reconstruction in mind. The original purpose of these tools was to allow for visual inspection of the AGR fuel channel in its entirety and footage capture of structural defects to provide additional information for monitoring, investigative and categorisation purposes. The CBIU and NICIE2 tools fulfil this purpose, however, they are limited by the technology available at the time of their development. With regards to obtaining 3-D reconstruction, the limitations presented by the hardware are listed in the subsections below:

Image Quality

The 1^{st} and most prominent limitation of the footage obtained using the RVI footage is that the captured image quality. The reason as to why image quality plays an important part is that it dictates the textural quality of the image. If the image resolution is insufficient, appears blurry and suffers from noise or compression artifacts, extracting



(a)

(b)

(c)

Figure 2.6: Visual image quality comparison of AGR fuel channel bricks inner surface using a) Original, compressed CBIU image; b) Original, minimally compressed image from CBIU using bespoke experimental apparatus at Strathclyde; c) Extremely high-quality image from same apparatus using a DSLR camera with mounted mirror.

reliable features from the image is challenging and can cause issues for most image processing applications. Conversely, if the image quality is high, with a high image

resolution with sharp, crisp textures visible within the image, it provides a much better basis for processing.

The image resolution obtained by either the CBIU or NICIE2 inspection tools is of a "576p25" analog Phase Alternating Line (PAL) format with horizontal and vertical image dimensions of 720×576 captured at a frame-rate of 25 frames per second (FPS). The footage is obtained in an analog format where it is then processed and digitised so that further analysis can be performed on the footage off-site. A significant issue with using a DVD recording device is that it introduces image compression to the captured footage. Image compression therefore becomes an additive problem, resulting in distortion of the image data as a direct result of quantisation noise and the compression artefacts from video containers such as MPEG [43]. To demonstrate the effect of this, Figure 2.6a) and b) are recorded from the same tool however Figure 2.6b) has a high degree of compression quality.

As observed, the image still in Figure 2.6b) gathered in laboratory conditions with minimal video compression produces images with a higher degree of textural information whereas the image still gathered from the in-core inspection suffers from motion compression based artefacts, directly resulting in the textural content being blurred. In Figure 2.6c), the image still has been taken by a modern Digital Single-Lens Reflect (DSLR) camera with a mounted-mirror lens to closely emulate the capture platforms such as NICIE/CBIU as described in Section 2.2.2 with no compression applied, resulted in an extremely high quality image. The tangible difference in image quality between the inspection tool and the modern camera can be observed and acutely demonstrates the limitations of the hardware with regards to the image quality.

With the NICIE inspection tool used in the laboratory, tests were carried out using the inspection tool with more information about this in Chapter 6 to confirm the source of possible image artefacts and lighting issues. The primary observation was that the RVI footage carried out within the reactors contain a large degree of non-uniform illumination, especially around the bottom left and right corners of the image. To exhibit this, the image still from Figure 2.6b) contrast values were increased by 50% which can be observed in Figure 2.7. This is a direct result of the LED placement

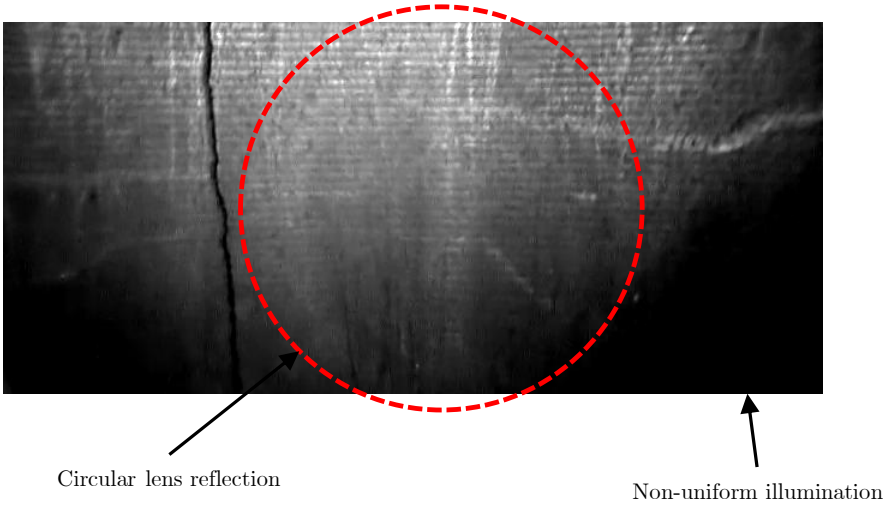
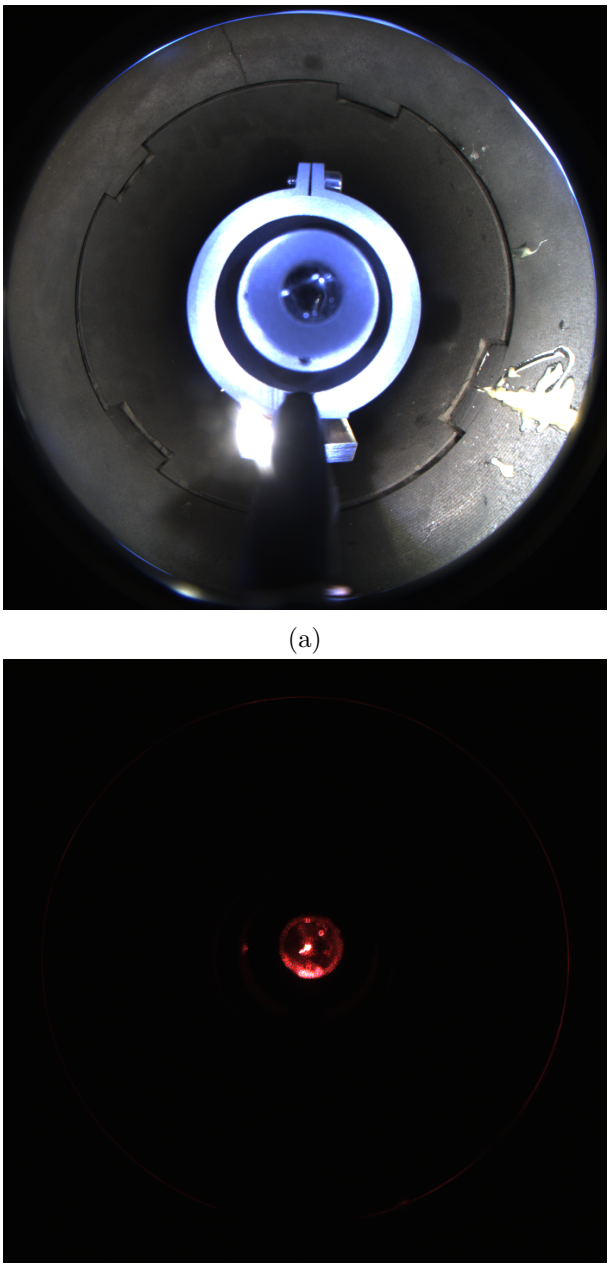


Figure 2.7: NICIE tool image still from 2.6b) with 50% contrast increase which highlights non-uniform illumination and a camera lens reflection (Highlighted using a black circle). Note: This is more apparent when viewed in video.

within the inspection tool, which causes the light to reflect off the mirror resulting in a reflection of the lens onto the channel surface. This is problematic when trying to determine 3-D information as this will skew the correspondence process as the lens reflection will appear in every image. Furthermore, as the camera moves up the channel, the previously bright surface at the top of the image becomes dark and determining a correspondence between these respective image stills also becomes very challenging. Therefore, image processing algorithms that are invariant to the changes in lighting must be taken into consideration so that accurate correspondences between each image can be obtained. This is discussed further in Chapter 4 and 6.

Inspection procedure

There are challenges with using the in-core inspection footage for the extraction of 3-D information. The first challenge is the use of a pin-hole camera model in order to inspect the cylindrical fuel channel. Although the pin-hole camera with an engaged mirror is excellent for thoroughly inspecting cracks, it is not suitable nor is it efficient to derive the 3-D structure of the fuel channel. One possible approach is through the utilisation



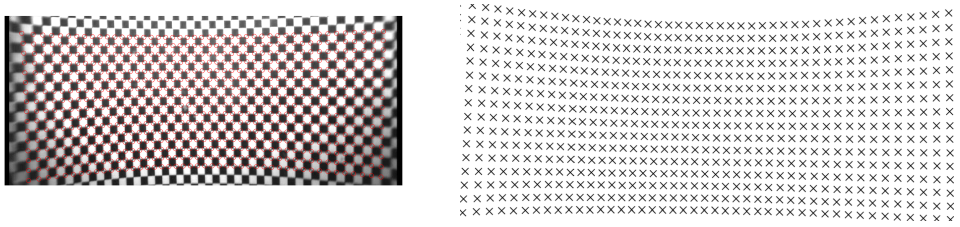
(a)

(b)

Figure 2.8: a) AGR fuel channel image still using MAPS probe [14] which has a fisheye lens; b) Laser projection onto brick surface. Note: The MAPS probe was not designed for volumetric scanning of the diameter of an AGR brick therefore the laser is near indistinguishable.

of an inspection tool with circular fisheye lens optics which allows the capture of an entire 180° Field of View as it is manoeuvred down the fuel channel. An example of this

can be seen in Figure 2.8a) where the MAPS probe developed by Summan et al [14] was centrally translated through an AGR brick in laboratory conditions. With a fisheye lens, instead of implementing 6 vertical scans at differing angular view-points, a single scan could be employed to capture the entire surface which could vastly accelerate the inspection process and as a result, reduce the critical outage path of the reactor.



(a) (b)

Figure 2.9: a) Flat checkerboard image with detected corners necessary for calibration process; b) Detected corner locations without underlying image. Note the large amount of tangential distortion and blurring from the camera which complicates calibration.

Additionally, another fundamental component that is missing for the derivation of 3-D structural geometry is the relevant ground-truth data. Before the inspection begins, calibration takes place ,however, this is only for the lighting intensity and the focal length of the camera. Important factors such as the intrinsic and extrinsic parameters of the camera which are critical to understanding the relationship between the image space and the world space is not obtained. Consequently, accurately extracting geometry of the source environment becomes very complex. Furthermore, without a proper camera calibration process, lens distortion deforms the source images which in turn obscures potential depth cues that can be used for 3-D reconstruction. To illustrate this, Figure 2.9 demonstrates the lens distortion of the NICIE inspection tool by observing a flat checkerboard, an approach commonly used to calibrate cameras and rectify the distorted images so that lens distortion can be minimised. In Figure 2.9b), by plotting the detected corner locations, it can observed that the inspection tool suffers from a high degree of lens distortion as each row should be aligned in a straight-line. With this level of distortion, RVI footage can misrepresent the captured structure.

Another component of ground-truth data that is not considered yet important is the capture of rigid-body motion of the inspection tool within the reactor. The only

quantification of motion within the reactor is when the inspection tool has retracted back to the top of the channel where a gauge can establish the angular off-set from its initial angle. Therefore, there is no account of how and when the inspection tool is subjected to unexpected rigid-body motion whilst the tool is inside the fuel channel. Without this information, there can be no validation of how the inspection tool moves within the fuel channel and this was the principal motivation for the development of a bespoke apparatus set-up explained in Chapter 6

2.4.2 Image-based 3-D scene reconstruction challenges

In addition to the hardware limitations, there are also some theoretical challenges when trying to perform 3-D scene reconstruction in environments such as the interior of a AGR fuel channel. The inspection tool gathers 2 tangible sets of data that can be used to perform the reconstruction. A video which can be decomposed into a series of contiguous images and the capture of the fuel channel diameter using the bore feelers as detailed in Chapter 2.2.2. To produce a 3-D reconstruction of the AGR fuel channels, the best approach was the use of an image-based 3-D scene reconstruction framework based on the principles of SfM. The output of SfM is two-fold, the derivation of the observed environments structural geometry in the form of a sparse point cloud and the associated rigid-body motion of the capture device often depicted as interconnected cameras.This process is explored in much greater detail in the following chapter.

Determining reliable correspondences between AGR RVI images becomes very challenging, even with state-of-the-art image registration techniques [44, 45] due to very little textural information and challenging capture conditions as seen in Figure 2.3, 2.6 and 2.7. Without a robust and reliable set of correspondences between images, subsequent processes such as triangulation of the correspondences results in ambiguous and erroneous representations of the source environment. Since accuracy of the point cloud is essential, not necessarily the completeness [9], it is pertinent that an approach is developed in order to operate within challenging feature spaces as observed within the AGR fuel channel. In industries that operate within a similar problem space, additional methods of extracting 3-D representations of the constrained com-

ponents revolves around the use of additional sensory equipment, such as structured lighting [13] and laser-based profiling [46] to effectively augment the 3-D reconstruction procedure. Within the nuclear industry however, there is a reliance on ultrasonic methods, especially within water-based reactor designs, to produce 3-D representations of the component surface as discussed in Section 2.3.1. Generally, there has been an introduction of new and different means of inspecting constrained rigid structures and producing 3-D reconstructions using Time-of-Flight (ToF) and RGB-D camera systems, and LiDAR (Light Detection and Ranging) [47], with a in-depth review of such approaches presented in Santoso et al [23]. The application of such methods are not of use with regard to inspecting AGR fuel channels due to being cost prohibitive, however these approaches could be potentially adopted as potential inspection methods for the future. In the following section, a small discussion will be introduced to identify potential imaging solutions for visual inspection is discussed.

2.5 Future reactor condition monitoring and inspection

Within the nuclear industry, most inspection methods were based on the use of phased ultrasonic sensory equipment due to the dominance of water-based reactor designs. In this section, it looks to the future to determine what other possible approaches that could be adopted in the future for general nuclear inspection. This is motivated by the introduction of "Generation IV" nuclear reactor designs [48][2] and what this can mean for the future of reactor inspection within the U.K. The 4^{th} generation of nuclear reactor inspection involves the advancement of two core fundamental measurement and inspection approaches:

Condition Monitoring Continual real-time data acquisition and analysis of the in-core operation parameters.

Periodic Inspection Nondestructive inspection of the reactor components during periodic reactor shut-down and maintenance.

[2]This is not to be confused by Industry 4.0 - termed the fourth industrial revolution

The 4^{th} generation looks to deploy some of the latest and most sophisticated methods of monitoring and inspection to simultaneously extend reactor operational lifetimes, reduce the critical outage path and improve operator safety.

Passive Inspection Approaches

There are increasing number of passive inspection approaches and methodologies that could be adapted in order to improve nuclear reactor inspection that to the best of the authors knowledge is currently unused or under-utilised within the nuclear industry. The fundamental reason for this lag for the nuclear industry is due to the sensory equipment having to be intolerant to radioactive and hazardous environments, resulting in considerably longer periods of development. The first considerations would be involving the camera system itself and the optical camera model that is employed for inspection. The first consideration is the use of lenses within the inspection; deploying a camera with a very large field of view ensures much greater coverage which fundamentally reduces inspection times within the reactor. The use of the typical pinhole camera model has a highly restrictive FoV which suffers from distortion above approximately 120° whereas the use of the unified camera model in use by catadioptric lenses [49] or the use of the novel double sphere camera model [50] for fisheye lenses allows for a far greater FoV and would vastly improve 3-D reconstruction. A simple extension of this would be implement a secondary camera system to produce stereo vision, exploiting the concept of stereopsis [51] to ascertain 3-D depth from the scene.

Active Inspection Approaches

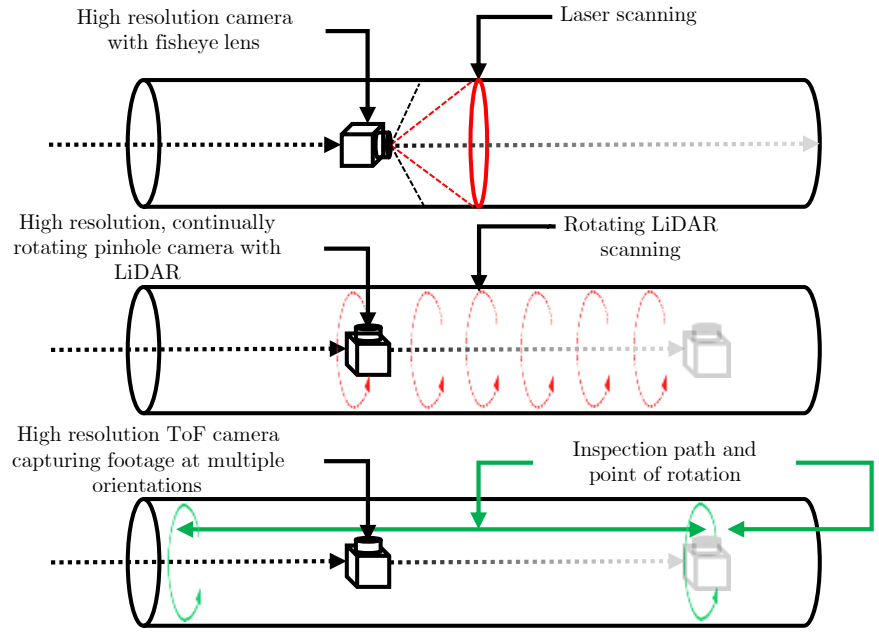
Active, non-contact inspection techniques are the most predominant method of inspection within nuclear, however, almost all systems require on an array of ultrasonic transducers to do this. Newer technologies with regards to surveying and inspection have become increasingly common in the field of NDE such as range-imaging approaches which incorporate the passive and active approach together to produce range-imagery.

Camera systems such as ToF camera systems would also be an excellent solution[3] due to its compactness, rapid processing time and its accuracy [23]. Furthermore, active visual sensor systems such as structured-light 3-D scanning [13], LiDAR [47] or laser scanning [46] could be used within a nuclear environment in order to produce highly accurate 3-D representations of a scene.

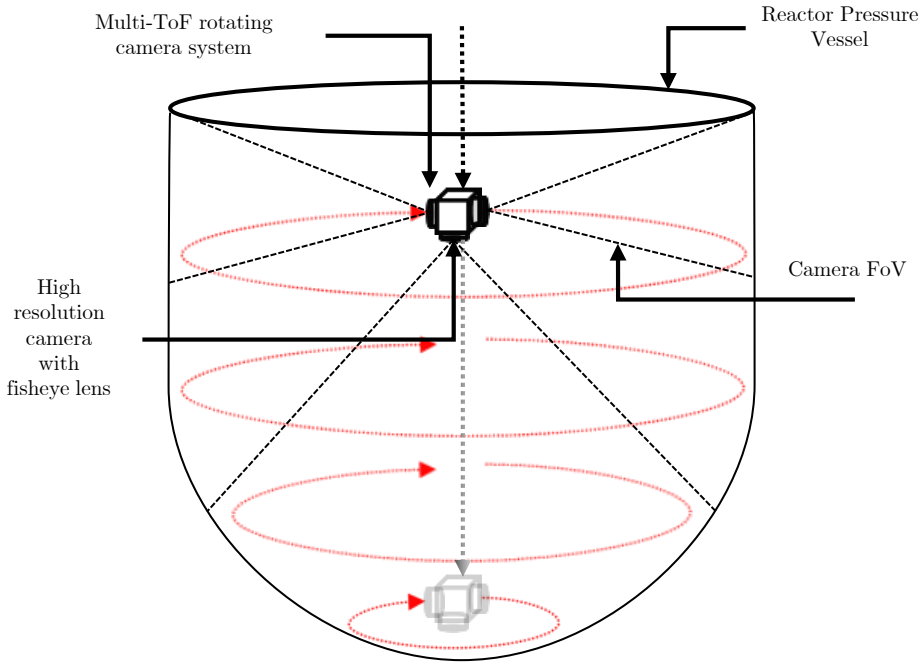
Evolution of the inspection protocol

Coinciding with the advancement of inspection hardware, the inspection protocol within the reactor can also be vastly improved by taking into consideration during the reactor design stage the accessibility of remote inspection devices, simplifying their geometry to make inspection much easier. The two most common sets of geometries that must be inspected within nuclear is a cylindrical tube or pipe for pressure [38,52], fuel channels [32,33], general ducts and piping [34,53] and the Reactor Pressure Vessel [54,55]. By simplifying the scene geometry that needs to be captured, the corresponding inspection can also be highly simplified. Examples of different, hypothetical inspection approaches using modern visual sensory techniques can be seen in Figure 2.10. Although there can be no panacea when it comes to inspection approaches as each reactor requires inspection tools to cater to the environment and application specific problems which each geometry provides, by utilising technologies such as ToF or LiDAR based inspection approaches, high resolution 2-D imagery and 3-D mapping can be obtained of even the most challenging reactor components.

[3]It is important to note that RGB-D cameras, although a ToF-type of camera, this particular type of visual sensor would be unusable in a nuclear environment due to its vulnerability to heat.



(a)



(b)

Figure 2.10: Different inspection protocols using most common structural geometries. a) Pipe-based geometries; b) RPV-based geometries.

2.6 Discussion

This chapter has introduced the fundamentals of AGR reactor operation and the processes used in order to determine and monitor the health of the graphite moderator. Visual inspection plays an important part in ensuring safe reactor operation and over the years, the approaches used in order to visualise the AGR fuel channel interiors have become increasingly advanced. From hand-drawn representations to full panoramic views through the use of image stitching [2,32]. With respect to other reactor designs, a variety of nondestructive visual inspection techniques have been deployed through industry to facilitate inspection engineers and reduce the critical outage path, with many methods being bespoke to the specific reactor design and the component under investigation. Generally, image-based inspection is not deployed throughout the nuclear industry, with many engineers relying on active sensory approaches in order to investigate the reactor design, and this is a direct result of most reactor designs being water-based. With potential components being entrenched in water, applying image-based inspection to underwater components could produce additional challenges that could be alleviated by simply deploying more suitable hardware, such as ultrasonic inspection tools [4,37]. The use of advanced image processing techniques deployed on source TV imagery is extremely limited within literature with visual inspection being deployed in order to facilitate the inspection [6,38] rather than be a key element of data capture itself, and this is in part due to the underlying sophistication of being able to extract additional information from imagery often being prohibitive. Due to the conditions within the AGR channel being harsh and highly irradiated, active sensory equipment is highly likely to fail. With additional facilitatory equipment being added to the current inspection process being cost prohibitive, the challenge was presented to produce 3-D reconstructions using 2-D inspection imagery alone.

In the introductory work of [33,56], a proposition was made to extract 3-D information from 2-D RVI footage acquired in the AGR fuel channels. The use of pivot videos highlighted the value of being able to perceive structural components from multiple perspectives, and this naturally evolved into a concept involving multiple view geom-

etry called Structure-from-Motion (SfM). To obtain a 3-D visualisation of the channel from strictly 2-D inspection imagery is a difficult challenge which is aggravated further as the inspection tools and the associated inspection protocol are not designed with 3-D scene reconstruction in mind. In this chapter, the inherent hardware and theoretical limitations with regards to deriving the 3-D structural geometry of the AGR fuel channel are introduced, such as challenging image quality or the constrained environment in which the inspection imagery is acquired. The chapter ends with giving potential pathways to future nuclear reactor inspection capabilities that could be adopted not only in constrained, boroscopic environments but more open-ended, yet still difficult inspection geometries. The following chapter gives an introduction to the key steps performed within 3-D scene reconstruction to the reader, as this will prove useful in contextualising the technical image processing challenges.

Chapter 3

Image-based 3-D Scene Reconstruction

3.1 An Introduction to Structure-from-Motion

Geometric vision is a subdomain within computer vision which is focused on the derivation of 3-D geometry, given images of the scene from multiple points of view which are spatially distinct in position and orientation. This problem is encapsulated under the mathematical formulation of SfM and can be expressed as the inverse of the image formation process where by exploiting the intrinsic and extrinsic parameters of the camera model, and the distinct locations and the inter/intra-properties of the images relative to the target structure, a 3-D reconstruction with the corresponding camera view-point locations and pose can be ascertained. Further information on the topic can be found in books by Hartley et al [57], Ma et al [58] and Szeliski et al [59]. This chapter is prefaced with a small introduction to SfM and how it works on a fundamental level, before performing a literature review on the historical evolution of SfM and application-specific approaches which approximate the thesis problem space, before introducing the concept of image-based 3-D scene reconstruction to footage acquired for our own work.

Image-based 3-D scene reconstruction can be performed using two distinct variants; SfM [21] and SLAM [22] which are performed offline and online respectively. However, the underlying principles of both approaches are the same. The first and most important stage of both approaches is to determine the locations within the assessed images where they share a correspondence to determine potential scene overlap. By determining the correspondence between image pairs, relationships between assessed imagery can be produced which illustrates the visibility of scene features across multiple images. This

is often termed the visibility matrix or the scene graph which comprises of geometrically verified image pairs [21]. The next stage is to estimate the pose of the camera relative to the captured scene and to reconstruct the scene structure as a set of 3-D points. These stages are often termed in literature as the "Correspondence searching" and the "Reconstruction". There now follows a brief explanation of these stages.

3.1.1 Correspondence searching

Correspondence searching comprises of 3 main stages as illustrated in Figure 3.1, namely Feature detection, feature matching and geometric verification.

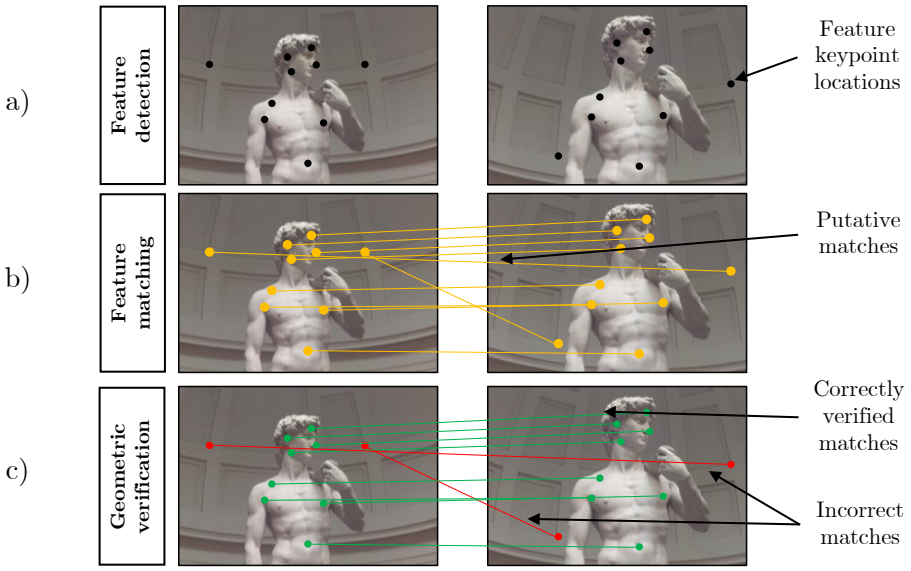


Figure 3.1: Visual breakdown of the correspondence searching process; a) Identified feature locations using arbitrary feature detector; b) Matching procedure used to determine unique matches based on the detected feature descriptors; c) Geometrical verification of the feature matches using epipolar geometry

Feature detection

Correspondence searching involves taking in a unordered or contiguously captured image data and extracting features from the images. The extracted features sets $\mathcal{F}_i = (\mathbf{x}_j, \mathbf{f}_j) \mid j = 1...N_{F_i}$ where $\mathbf{x}_j \in \mathbb{R}^2$ is the feature location and $\mathbf{f}_j$ is the descriptor which represents the feature point with properties which describe the intrinsic char-

acteristics of the feature and its surrounding neighbourhood. One of the fundamental requirements for successful correspondence searching is that the detected features must be recognised regardless of scale, orientation or illumination across a series of images. The Scale Invariant Feature Transform (SIFT) [60] and its associated derivatives [61] are regarded as the best feature detection methods for describing an image using a abstract set of salient points. An example of this is shown in Figure 3.1a).

Feature matching

After features are detected within every image, the subsequent stage is to effectively match the derived image feature-sets together in order to determine unique 1-to-1 correspondences. The features can be matched effectively using metrics such as the squared Euclidean distance [32, 60] to determine a set of putative matches. In more challenging environments, images may require more extensive matching regimes [62,63], or the use of dense matching procedures [64, 65] which utilise every pixel in the image as a potential feature.

Geometric Verification

Once the matching procedure is complete, the matches must be geometrically verified under the principle that the features matched between one image and another follow a consistent geometric transformation. Using the principles of epipolar geometry [57], the geometric transformation between the feature matches can be estimated using an Essential matrix $\mathbf{E}$ [66] for calibrated cameras or the Fundamental matrix $\mathbf{F}$ [67] for uncalibrated cameras. If the transformation described by $\mathbf{E}$ or $\mathbf{F}$ sufficiently maps enough feature matches, the feature matches are said to be geometrically verified with the outliers being removed from the feature set. For a robust estimation, estimators such as RANSAC [68] is used to help prune outliers from the set.

3.1.2 Reconstruction

Reconstruction generally has 4 key quasi-parallel stages as shown in Figure 3.2, namely, Model initialisation, pose estimation, triangulation and Bundle Adjustment (BA).

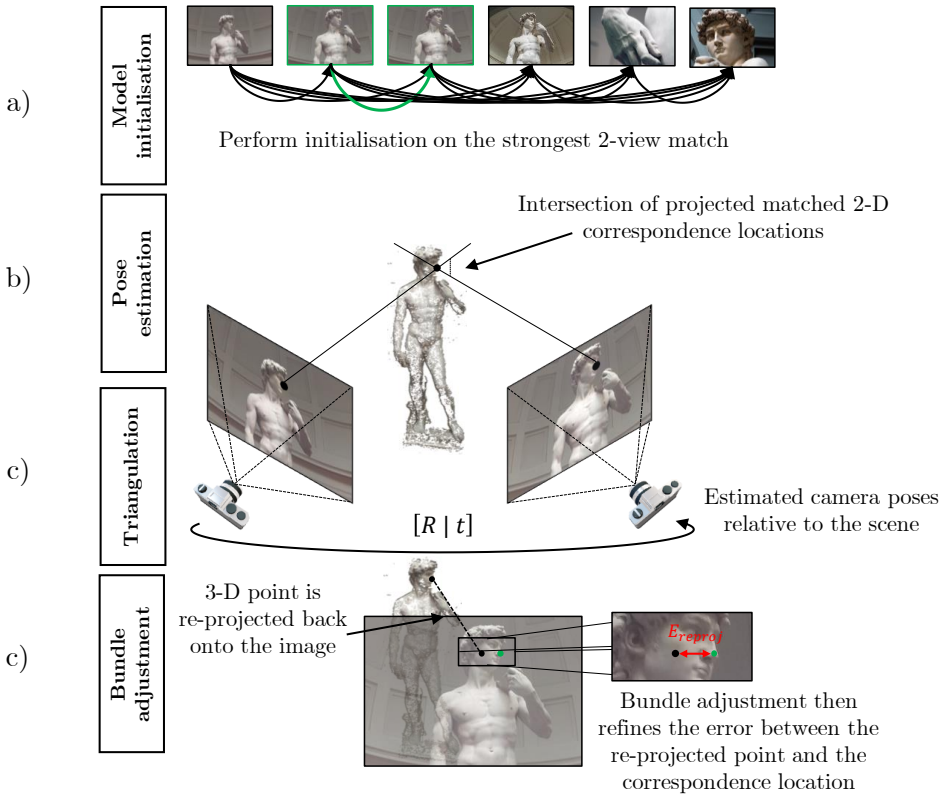


Figure 3.2: Visual breakdown of the reconstruction process; a) Model initialisation using the strongest 2-view match; b/c) Estimation of the camera and triangulation of 3-D points ; d) Refinement of the re-projection error using Bundle adjustment

Model initialisation

Model initialisation, especially in incremental frameworks [25], require an initial location in order to seed the reconstruction [19]. This process operates by selecting a pair of images [19], or an area within a scene graph that is highly dense [21] to initialise the model. Then new images are continually registered to the model, and the corresponding 3-D points and camera locations are triangulated and estimated respectively.

Pose estimation and triangulation

Pose estimation and triangulation are two quasi-parallel processes that operate in tandem when new images are being registered to the model. Pose estimation is per-

formed to resolve the cameras extrinsic parameters $\mathbf{P}_c = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]$ where $\mathbf{R}$ is the rotation matrix and $\mathbf{t}$ is the translation, representing the rigid-body motion of the camera through the scene [57, 58]. By solving the Perspective-n-Point (PnP) [69] problem, 2-D image correspondence locations $\mathbf{x}$ are mapped to 3-D triangulated points $\mathbf{X}_k \in \mathbb{R}^3 \mid k = 1...N_X$, the rigid-body motion of the camera can be estimated, in addition the intrinsic parameter matrix $\mathbf{K}$ for the purposes of auto-calibration [70]. This is illustrated in Figure 3.2b) and c).

Bundle adjustment

The procedure performed by BA is to effectively perform a joint refinement of the camera parameters and the 3-D triangulated points. This procedure works by re-projecting the triangulated 3-D point $\mathbf{X}$ back onto the 2-D image using the estimated camera parameters and 3-D point location, and minimising the error between the reprojected 2-D location between the detected 2-D feature correspondence location $\mathbf{x}$ in order to refine the generated model. The error minimised is termed in literature as the reprojection error E_{reproj} expressed below:

$$E_{reproj} = \sum_j \rho \left(||\pi(\mathbf{P}_c, \mathbf{X}_k) - \mathbf{x}||_2^2 \right) \tag{3.1}$$

where ρ_j is the loss function and π is the function that projects scene points to image space [21]. The process operates on both a localised level and global level to continually add and refine the reconstruction and camera pose estimates.

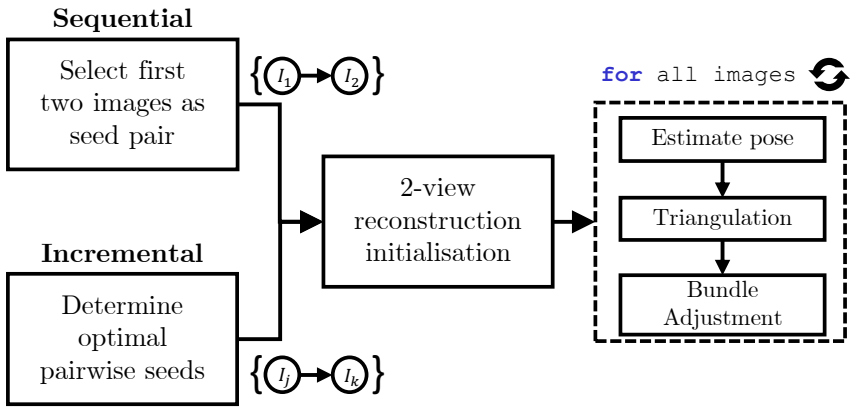
3.2 Contemporary SfM frameworks

Contemporary SfM frameworks originate on the original work performed by Koenderink, Luong and Faugeras [67, 71, 72] which allowed the determination of Euclidean reconstructions of scenes given sets a set of correspondences. The work of Pollefeys et al [73] and Schaffalitzky et al [74] introduced the general framework for performing image-based 3-D reconstructions which many contemporary frameworks are based on today. The procedures developed allowed for unordered and ordered feature matching,

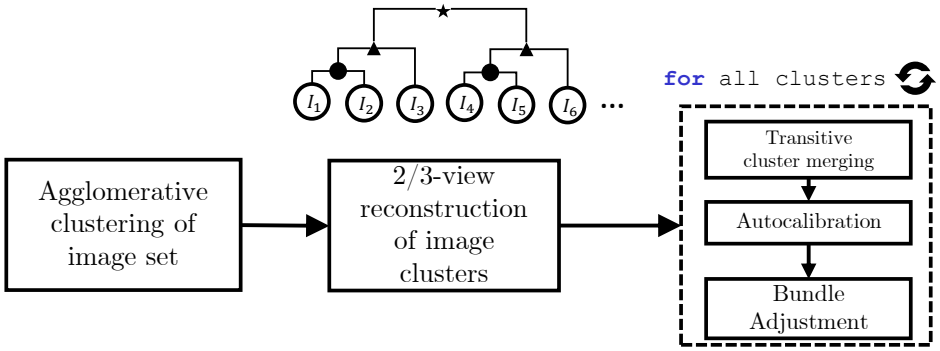
integration of auto-calibration [70, 75], the use of feature descriptors [73] and the use of BA which was adopted from the field of photogrammetry [76] to refine the final reconstructions. Over the years, SfM reconstruction frameworks have diverged from simple sequential methods to 3 reconstruction archetypes - incremental, hierarchical and global SfM frameworks as shown in Figure 3.3. All the reconstruction methods contain the same underlying fundamental mechanisms; determination of geometrically verified feature correspondences between images using consensus methods such as RANdom Sampling and Consensus (RANSAC) [68], image registration [77], triangulation [78, 79] and BA [76, 80]. The latent dissimilarity between these frameworks, however, originates in how the images are effectively processed and reconstructed, with each framework having fundamental strengths and weaknesses dependent on the source data properties and volume.

3.2.1 Sequential and Incremental SfM

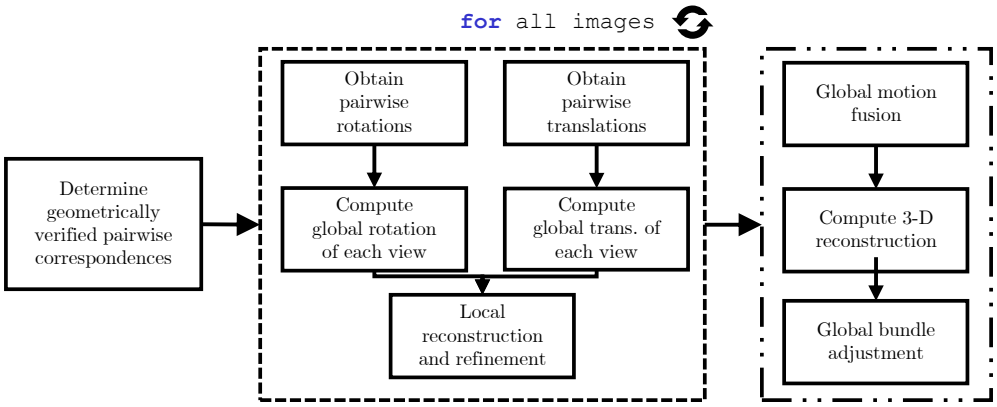
Sequential and incremental methods were the original approaches since data-sets were often small and observed highly salient objects or scenes at that point in time. In the last 15 years, sequential methods have evolved from off-line implementation frameworks such as SfM to on-line SLAM approaches introduced first by Davison [22, 81, 82] which exploit the contiguous nature of the image capture process to simultaneously localise the camera and reconstruct the scene. However, both SfM and SLAM were extremely limited at the time, with robust feature correspondences being challenging and the BA process being very computationally expensive when large amounts of images were introduced. With the introduction of SIFT feature descriptor and sparse BA, which exploits the sparsity of the block structure reflecting the "loose coupling" [83] of camera parameters and 3-D points [79, 80, 83], computation of large datasets was feasible. This gave way to the first ever large scale incremental reconstruction system which was developed by Snavely et al [19], where the input data was derived from internet-based images pertaining to specific areas or objects acquired from a range of view-points, camera models and environmental conditions. Due to the diversity of image quality and available information, the produced reconstruction was not guaranteed to be metric,



(a)

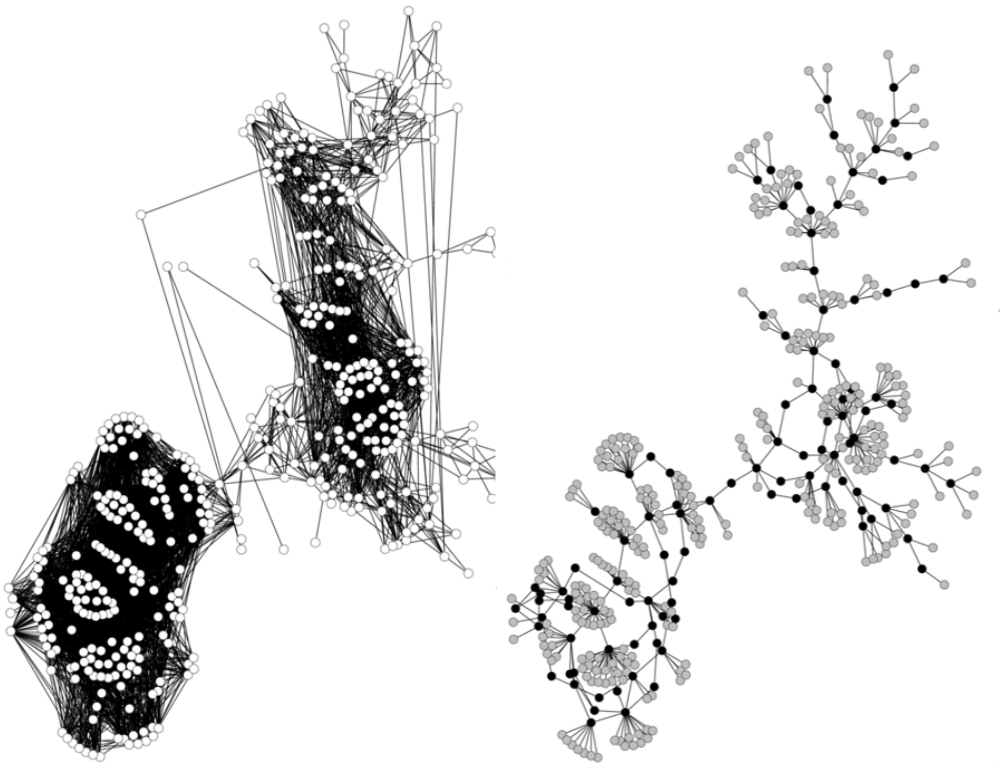


(b)



(c)

Figure 3.3: Visualisation of the different SfM strategies implemented within literature; a) Incremental, b) Hierarchical, c) Global



(a) (b)

Figure 3.4: Skeletal graphs to improve efficiency of incremental reconstruction methods as observed in [84, 85].

nor was the point-cloud density if the image quality was insufficient. As demonstrated in [19], the incremental system performance also becomes progressively slower as the image dataset size increases in size with computational complexity $O(n^4)$, where n is the number of images.

However, [19] laid bare the computational complexity issues of not only incremental methods, but general-purpose image-based 3-D reconstruction in terms of scalability. To improve computation time, especially of unordered data-sets, redundancy within the current SfM approach must be suppressed. The first notable approaches to reduce the computational complexity of the reconstruction at the matching stage was through the use of a optimal sub-set of images and their respective matches termed a skeletal graph [84, 85]. This approach is illustrated in Figure 3.4. Agarwal et al [86, 87] improves upon the work of Snavely et al [19, 84, 85], implementing a whole image similarity mechanism

to cluster and match similar images as a sparse matching graph then increase the match graph density via transitive matching (see Figure 4.4). The approach by Agarwal et al also introduced the use of Multi-View Stereo (MVS); a technique which estimates depth information from multiple views and merges the 3D points into a singular model based on the underlying derived relationships of the sparse 3-D model [88]. This work was then directly broadened upon by the research of [89] by exploiting geo-location information (if available) and the use of a vocabulary tree for iconic areas of the surveyed city for the matching process in addition to Graphics Processing Units (GPU) accelerated appearance-based clustering to further reduce the computational load. The presented techniques had some fundamental issues; with multiple data sources, global thresholds become problematic when dealing with a large variety of images and their corresponding feature spaces, BA refinement of the model (either locally or globally) was inflating the time-complexity of the reconstruction and there is no considerations to images with large degrees of homogeneity which could contain valuable structural information.

Moulon et al [90] attempted to tackle the issue of thresholding during RANSAC model estimation and camera pose estimation through the use of a *a-contrario* methodological approach to adaptively select thresholds for both correspondence searching and geometric verification, but presents a fundamental weakness where the method requires a preinitialised hypothesis of expected inlier correspondences to obtain a geometrically verified consensus. With regards to incremental reconstruction, Wu et al [25] introduced the popular closed-source "VisualSFM", an incremental SfM framework which offered several novelties which vastly improved pre-existing incremental methods. In addition to a preemptive matching strategy similar to skeletal graphs [84] to reduce the feature space, the BA procedure exploited the Preconditioned Conjugate Gradient (PCG) to reduce the time complexity of incremental reconstruction from $O(n^4) \rightarrow O(n^2)$. Finally Wu introduces "Retriangulation" which aims to reconstruct previously failed feature matches and decrease the threshold for reprojection errors in order to densify the point cloud and reduce drift. However, by introducing weak matches, there is a chance of noisy or erroneous 3-D matches being introduced into the system, sacrificing accuracy for reconstruction completeness. In application problem spaces as described

in this thesis where accuracy of the reconstruction is crucial, retriangulation will not be considered.

Schönberger et al [21,91] improve upon the steps introduced by Wu et al [25] with the development of COLMAP, an incremental SfM framework which is currently considered as the state-of-the-art when it comes to incremental SfM frameworks. The method introduces a scene graph which takes into consideration the geometric verifiability and augments the scene graph with geometric relations to improve the initialisation and triangulation stages. To alleviate reconstruction drift which is also symptomatic of poor image incorporation into the reconstruction process, a next-best view selection strategy is implemented based on the spatial observation of triangulated feature points from each camera across multiple scales and selects the best image based on their cardinality and the distribution. This strategy also operates alongside a mechanism that is used to grouping cameras with high scene overlap into a singular camera, reducing the number of processed images. To improve triangulation accuracy, RANSAC is introduced to remove outliers based on the triangulation angle and assumption of positive depths. Like VisualSFM [25], retriangulation is also utilised, however, it is reimagined using an iterative refinement scheme which operates through filtering observations in addition to performing retriangulation before and after to improve reconstruction completeness.

3.2.2 Hierarchical SfM

Hierarchical approaches were introduced in the work by Nistér [92] where the reconstruction methodology was to effectively partition the processing into subsets of images where reconstruction would be performed. The approach incorporates the use of a series of sequentially obtained trifocal tensors in order to improve feature tracking and geometric verification and therefore the resulting reconstruction. Although the approach is not strictly hierarchical SfM with regards to the literature, by partitioning the reconstruction process into smaller subsets, the prominence of a key issue, incremental drift, is effectively suppressed and consequently results in a more accurate reconstruction. Farenzena et al [93] is the first notable example in literature of a true hierarchical framework and operates in a manner illustrated in Figure 3.3b). The method introduces

a bottom-up agglomerative clustering algorithm to hierarchically organise the images into subsets based on keypoint matches. Reconstruction and BA is then performed on each image cluster in the dendrogram (see Figure 3.5) with each cluster being merged to produce the final reconstruction. Like Bundler [19], the approach provided has time complexity of $O(n^4)$, however this hierarchical method can be parallelised to speed up the reconstruction process. Gherardi et al [94] provide an addendum onto [93] by further improving computation through the introduction of a balancing mechanism to the dendrogram in addition to an autocalibration procedure to improve reconstruction accuracy.

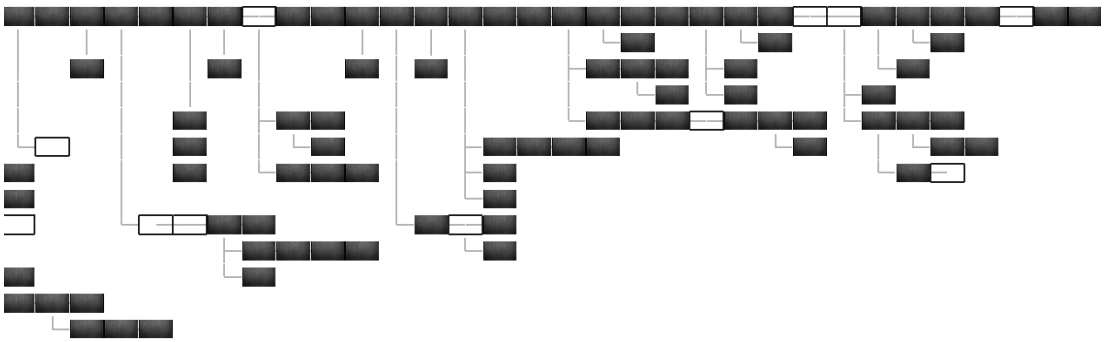


Figure 3.5: Hierarchical dendrogram of identified, sequential matches of RVI footage obtained from an AGR fuel channel. The matching process was restricted to sequential matching of image sub-sets.

HyperSfM by Ni et al [95] approaches the hierarchical reconstruction strategy in a different manner proposed by [93,94] in which it implements "hierarchical optimisation at the variable level", as opposed to the image level. This approach utilises optimised hypergraphs to hierarchically partition the image data in an more efficient manner than the use of the dendrogram. This produces a much more effective hierarchical clustering approach and furthermore, due to the use of compressed edges in the hypergraph which merges images that observe the same 3-D points, it vastly speeds up the BA, either local or global making it more efficient.

Toldo et al [96] acts as a later continuation of the work from [93, 94], with the modifications performed based on the correspondence searching stage in order to reduce the number of processed images. Similar to previous methods which aim to reduce

redundant processing of images, [96] utilises an epipolar graph which represents image overlap to identify similar images via the "broad phase" and keypoint clustering via the "narrow phase". This approach allows for the number of assessed images to be vastly reduced before being clustered and hierarchically reconstructed.

3.2.3 Global SfM

The thought process of hierarchical SfM shares the same motivations as the global SfM strategy, namely to reduce incremental drift. Hierarchical SfM effectively compartmentalises the error by partitioning the reconstruction process into smaller subsets, however, each smaller subset is reconstructed in a manner similar to incremental SfM approaches. Global SfM takes a fundamentally different approach in that, it considers the entire viewgraph only once rather than incrementally processing images one at a time or considering small clusters [90]. Global methods can be segregated into factorisation-based approaches [97] and motion-averaging based approaches [90, 98, 99]. Motion-averaging techniques have become the dominant approach due to factorisation-based approaches being susceptible to poor feature correspondences.

The global SfM approach aims to solve all the camera poses contained within the viewgraph by first creating an initial viewgraph between 2-view correspondences before computing the global camera rotation and translations across each camera using the relative pose information between images. The concept of rotation and translation averaging originate from the work of Govindu [100, 101], highlighting the deficiencies introduced by noisy outliers in relative rotation and the relative translation only known up to a certain scale factor [102]. For further reading on this topic, Hartley and Wilson et al [103, 104] give a good review of rotation averaging and Cui et al [101, 102, 105] for translation averaging respectively.

3.2.4 Applicable reconstruction pipelines

The history of 3-D reconstruction and an understanding of the current state-of-the-art generic SfM frameworks is important to convey the different strategies and incumbent methodologies used to deal with some of the general image-based 3-D reconstruction

problems. However, when the application space is considered, many of the proposed techniques either work on highly curated image datasets which are captured often in ideal conditions. Therefore it is prudent to identify application based 3-D reconstruction frameworks that demonstrate similarity to the application domain aforementioned in this thesis. Since there is no application-specific SfM within AGR fuel channels available to our knowledge, techniques which are deployed within pipelines, tunnels and sewer structures form the basis of the review.

The work by Kolesnik and Baratoff in 2000 [106] was the first notable kind of reconstruction applied within a constrained environment. The work aimed to improve visual surveying of the circular structures of underground sewers. A robot was developed with the primary purpose of autonomous surveying and was equipped with sensors, light sources and a camera. The framework functioned by first extracting circles through the use of the Hough transform and applied robust conic fitting and distance measurements to form a 3-D model of the evaluated sections of the sewer channels. This process would be repeated as the robot traversed on the frontoparallel plane. This technique may produce a 3-D model however it is limited in its representation of the 3-D structure as it only extracts 3-D points and not necessarily the image-based characteristics of said points. In literature, there is no notable techniques after the work produced by Kolesnik [106] for approximately 8 years and this is due to generic SfM pipelines making very little progress within this period.

Kannala et al [15] was the first application of SfM to the interior of a constrained object, in this case sewer pipes, building upon his previous work to reconstruct sewer pipes using a calibrated, omnidirectional camera. By using such a visual setup, it provides a greater FoV, however the video resolution from the robot was 320×240 within a poorly textured environment. Kannala utilises the Harris feature detector [107] and geometrically constrains the matching process by exploiting the constant translation of the robot to provide robust matches, allowing for lower thresholds for more edges to be detected without introducing spurious noise into the feature matching process. For the reconstruction approach, Kannala implements a hierarchical SfM approach similar to [92] which obtains image triplets and performs the subsequent processes on

subsets of the input images to allow distribution of the reconstruction error and camera pose drift in a manner similar to work by Farenzena et al and Gherardi et al [93, 94]. Afterwards a piecewise cylindrical model is implemented to allow the development of a non-rigid cylindrical structure to capture the bending of the sewer pipes and geometric constraints are applied for a surface fitting procedure. The technique proposed by Kannala et al [15] is a very effective method of constrained reconstruction however the use of the Harris feature detector restricts the systems ability to accurately pick up feature points and instead may be picking up spurious, environment noise.

Esquivel et al [16] takes a similar approach to Kannala et al [15] and deploys the use of a robot fitted with an omnidirectional camera and sensors which is suspended into manholes with the aim of reconstructing vertical sewer shafts. Esquivel [16] deploys a different mechanism for SfM where the problem is segregated into a boot-strapping and feature tracking stage using the Kanade Lukas Tomasi (KLT) tracking algorithm and estimates the essential matrix $\mathbf{E}$ from cylinder-mapped images. The camera is calibrated and $\mathbf{E}$ is continually estimated to reduce the drift as the system traverses the channel. Esquivels application differs as the footage doesn't have pure translation like [15] but is vertically suspended via a wire which leaves it open to unwanted rotation caused by the environment. This is offset with the use of a rotation sensor and an Inertial Measurement Unit (IMU) which tracks the device as it is lowered into the shaft and this knowledge is integrated into the SfM problem allowing for a more accurate reconstruction.

El Kahi et al [12] goes one step further by introducing pipe inspection without the use of constant translation or the use of an IMU to help geometrically constrain multiple stages of the generic SfM framework. Furthermore, El Kahi implements the system with a calibrated monocular pinhole camera with the KLT tracker similar to [16] which allows the tracking of features close to the camera in non-descript environments. The paper also instigates a new method which deploys the use of a binary occupancy map so that defects inside the pipe can be detected and quantified with centimetre accuracy. El Kahi et al [12] also highlight an important issue in accordance with non-uniform illumination; the difference in lighting causes triangulation to produce

erroneous results as the system assumes differing depths which then causes the final point cloud reconstruction to bend. El Kahi solves this by enlarging the window size of the descriptor and is a important point to note for our own application. She also highlights different issues such as the application of KLT tracker (*i.e. an optical flow algorithm*) which is only valid for use in a channel that has observable and detectable texture therefore techniques deployed by [12, 15, 16] won't necessarily work with all pipelines.

Seki et al [9] proposes a new approach which is currently the only available application of SfM to the nuclear domain (to the writers knowledge) with the aim of reconstructing the interior of the damaged Fukushima Daiichi nuclear power station in real-time; it would be more appropriate to say that this is a visual SLAM approach but has been included in this literature review for sake of completeness. The visual sensor is a radiation shielded, monocular camera system suspended from a crane, controlled remotely. Due to the radiation shielding, the FoV of the camera is limited to 30° and with the suspension from a crane, multiple takes of footage is required due to the sub-optimal recording conditions due to wind and other environmental conditions. The algorithm is based on the Parallel Tracking and Mapping (PTAM) visual SLAM algorithm by Klein and Murray [108]. The underlying feature descriptors are modified through the combination of Shi-Tomasi corners and the ORB feature descriptor [109] at multiple scales to allow for detection, tracking, pose estimation and mapping of the environment reliably. Real time MVS is implemented using the technique developed by [110] and two novel 3-D descriptors - Co-occurrence Histogram of Angle and Distance (CHAD) and Fast Point Feature Histogram (FPFH) - to allow for point cloud registration with structural occlusions.

Hansen et al [13] operates similar to the work of [15] through the utilisation of a robot with an omnidirectional camera retrofitted with a structured lighting system which passes through a pipeline consisting of straight pipes and junctions. The camera system is calibrated using a normal checker-board pattern in addition to the use of ground truth data about the pipe to calculate accurate estimates of the cameras intrinsic parameters. For feature tracking, a region-based Harris detector is used for feature

detection, similar to [15], and a key-frame based approach is undertaken in order to increase the baseline separation between matches and suppress the issue of ill-posed small-baseline matching as seen in [98] and improves pose estimation and triangulation. For unmatched features, Hansen et al [13] deploys the use of the cosine similarity metric in addition to Zero Mean Normalised Cross Correlation (ZNCC) guided by the epipolar constraint provided by **E** which drastically improved the number of feature correspondences. A structured lightning system is also integrated into the system which determines metric pose and mapping estimates which overcomes the issue of scale drift and ambiguity.

Summan et al [14, 111] implements SfM in combination with a laser metrology system and an IMU in order to produce a dense 3-D reconstruction of the interior of steel pipes. Similar to our own application, the interior of the pipe has very low textural content which causes issues when deriving feature points. To combat this, Summan take into consideration the spatial distribution of SIFT feature points within the pipe, and in combination with IMU data and prior knowledge of the pipe geometry, an accurate 3-D reconstruction can be obtained.

3.2.5 Critical literature review

This section aims to provide a overarching critique of the reviewed methods within academic literature and the various strategies deployed which could prove crucial for the application domain. Within general SfM literature, global SfM reconstruction methods are identified and objectively proven to provide the most accurate results. However, the global SfM strategy is demonstrably avoided in methods evaluating pipelines and other challenging constrained industrial structures due to their repetitive and pseudo-cyclic nature [112]. Therefore, many application-specific techniques operate in a sequential or incremental fashion to build 3-D models [9,13,111], however, these approaches often rely on passive sensory data or *a-priori* information to augment the reconstruction process [9, 13, 15], as producing 3-D reconstructions purely on image data often causes deformations of the estimated observed structure [12]. The main reason identified within the literature is that the strength of correspondence between the image data-sets

is of vital importance as it provides the fundamental structure to proceeding techniques used within 3-D reconstruction such as camera pose and depth estimation. To solve this core issue, techniques have taken to clustering similar images together as a pre-process to improve matching [86,87], augmentation of geometric relations to the scene graph to improve geometric verification and triangulation [21, 91], quasi-adaptive thresholding [90] or the use of re-triangulation to incorporate weaker matches [21, 25] to densify the resultant point cloud. Although these techniques are effectively general purpose solutions and are very effective for generalised image datasets that consist of typical scenes rich of unique textural components, the effectiveness of these techniques do not translate well into specialised applications as documented by this thesis. One of the major drawbacks with regards to many of the proposed techniques that aim to improve reconstruction accuracy is that the importance of a robust correspondence mechanism is consistently ignored with the computational efficiency of the 3-D reconstruction method and the density of the 3-D point cloud being considered more important. However, in applications such as described in this thesis, reconstruction accuracy is critically important and this stems from how robust the correspondence searching algorithm implemented is.

There are several issues with regards to constrained structure representation and these issues don't necessarily pertain to the 3-D reconstruction method but of the capture method. Other complications have been identified within literature, such as the non-uniformity of inspection tool lighting or the structures possible light reflectivity can cause considerable issues in the later stages of triangulation as documented by [12], who notes that the resulting 3-D point clouds often bend and suffer from structurally misrepresentative curvature due to the changes in lightning being interpreted as a function of depth. Another noticeable theme is that many of these application-based papers rely on the use of dated correspondence techniques such as KLT or Harris edge detector with only the latest paper from Summan et al [46] incorporating the standard SIFT feature extraction paradigm. Within this application space, to the writers knowledge, there is no exploitation of dense techniques such as the Dual TV-$\mathcal{L}^1$ optical flow algorithm [113] in order to help supplement many of the problems documented by the

papers in this application space, which those heavily relying on generic feature matching. In the following section, an investigation is performed into applying image-based 3-D reconstruction techniques reviewed within the literature review on the AGR RVI footage.

3.3 Image-based 3-D reconstruction of AGR RVI footage

The preliminary work by West et al. [33] introduced the concept of taking regions of interest within a reactor and producing 3-D representations by exploiting the basic principles of multiple view geometry [57]. The two approaches introduced were the use of anaglyphic images and pivot videos where an illusion of panning around a specific defect is employed to view the feature point from different angles.

3.3.1 Anaglyphic imagery

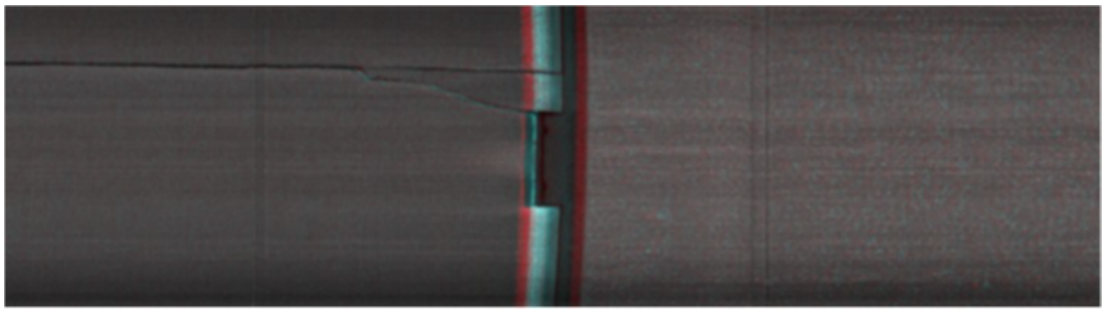


Figure 3.6: Red/Cyan anaglyphic image of a cracked region of interest within the AGR fuel channel taken from West et al. [33]. Note the colour filter within the crack denoting depth computation of defects being possible.

Anaglyphic imagery is a passive 3-D imaging method which deploys the underlying concept of stereopsis [51], encoding each monocular viewpoint in a different colour to give the illusion of depth when viewed through accompanying colours [43]. To demonstrate the concept, an anaglyph taken from West et al [33] can be seen in Figure 3.6. With an anaglyph of a region of interest, an engineer could look at a crack from two separate viewpoints through exploitation of binocular disparity, potentially improving the visualisation. This idea then evolved from a two-view concept to an N-view concept

through the use of "pivot videos".

3.3.2 Pivot footage

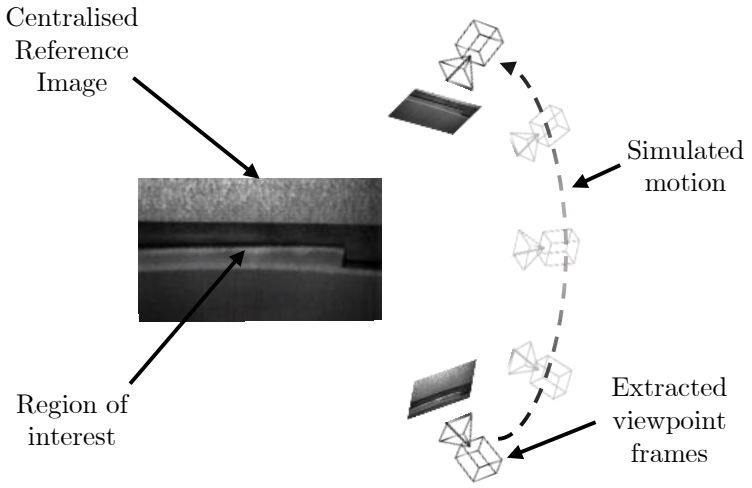


Figure 3.7: Pivot video generation process to view the AGR brick interface from multiple viewpoints.

Pivot videos [33,56] centralise a feature or region of interest within the image frame and by selecting images from various overlapping view-points, giving the illusion of camera pivoting (the more common term in cinematography is "panning") around an object, allowing for a more comprehensive inspection. This process is visualised in Figure 3.7. With this approach, reactor operators were interested in inspecting specific defects - in one particular case as described in [33] where a fragment of a cracked brick was to be investigated to see if the fragment had recessed into the crack or protruded into the channel as it may damage or interfere with the inspection equipment. By exploiting the underlying concept of multiple-view geometry, this gave way to the next step of theory which is to directly extract 3-D information from the RVI footage.

3.3.3 Image-based 3-D scene reconstruction

The concept of employing image-based 3-D scene reconstruction was introduced in [33, 56] where it was suggested, based on the application of anaglyphic imagery and

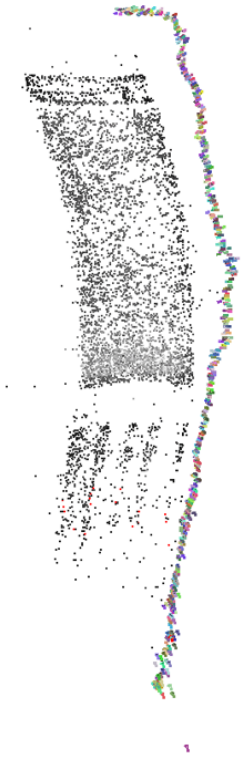


Figure 3.8: First 3-D point cloud produced using AGR footage using VisualSFM as shown in [56]. Note how the point cloud and the associated camera pose estimation is inaccurate.

pivot videos that the extraction of 3-D information from the footage was possible. In [56], a 3-D representation given the 2-D RVI footage was produced through the use of the closed-source VisualSFM software [25] as seen in Figure 3.8 and was the catalyst for future research in this area.

3.4 Image-based 3-D scene reconstruction with AGR in-core RVI footage

3.4.1 Framework Evaluation

As described in Chapter 2.4, performing 3-D visualisation of AGR fuel channels using the current hardware is speculated theoretically to be a significant challenge due to

Author	Name	Framework availability	Framework type	Initialised	Reconstruction
Microsoft [114]	Photosynth	Proprietary	Unknown	No	Failed
Agisoft [115]	PhotoScan	Proprietary	Unknown	Yes	Partial
Autodesk [116]	123D Catch	Proprietary	Unknown	No	Failed
BentleySystems [117]	Acute3D	Proprietary	Unknown	No	Failed
Snavely et al [19, 84]	Bundler	Open-source	Incremental SfM	No	Failed
Changchang Wu [25]	VisualSFM	Proprietary	Incremental SfM	Yes	Full
Moulon et al [118]	OpenMVG	Open-source	Global SfM	No	Failed
Fuhrmann et al [18]	MVE	Proprietary	Incremental SfM	No	Failed
Schönberger et al [21]	COLMAP	Open-source	Incremental SfM	Yes	Full
Montemerlo et al [119, 120]	FAST-SLAM	Open-source	SLAM	No	Failed
Civera et al [121]	EKF-SLAM	Open-source	SLAM	No	Failed
Klein et al [108]	PTAM	Open-source	SLAM	No	Failed
Engel et al [122]	LSD-SLAM	Open-source	SLAM	No	Failed
Mur-Artal et al [123, 124]	ORB-SLAM	Open-source	SLAM	No	Failed

Table 3.1: List of open-source and proprietary image-based 3-D reconstruction frameworks tested using AGR RVI footage

the hardware being sub-optimal for 3-D visualisation. Throughout this chapter, the concept of image-based 3-D reconstruction has been described in detail, both from a technical and an historical perspective. Within this section, an evaluation of different available open-source or proprietary image-based 3-D reconstruction software using AGR RVI footage is performed to highlight the technical challenges posed by using footage captured in a constrained, homogeneous environments. This test incorporates the testing of different SfM methodologies (Incremental and global) in addition to SLAM frameworks. The testing of image-based 3-D reconstruction techniques using the AGR RVI footage were segregated into proprietary and open-source entities with the frameworks evaluated listed in Table 3.1.

To evaluate the performance of each framework and provoke discussion, two specific sets of AGR RVI footage obtained within AGR reactors is utilised:

Brick 1 (B1) footage Footage containing only brick layer 1. This brick layer has a large amount of observable characteristics and as a result, features that can be used in the channel to help feature detection. It also contains a small featureless region. The video is decomposed into 681 images.

Brick 1&2 (B12) footage Footage that contains both layer 1 and layer 2 of a AGR

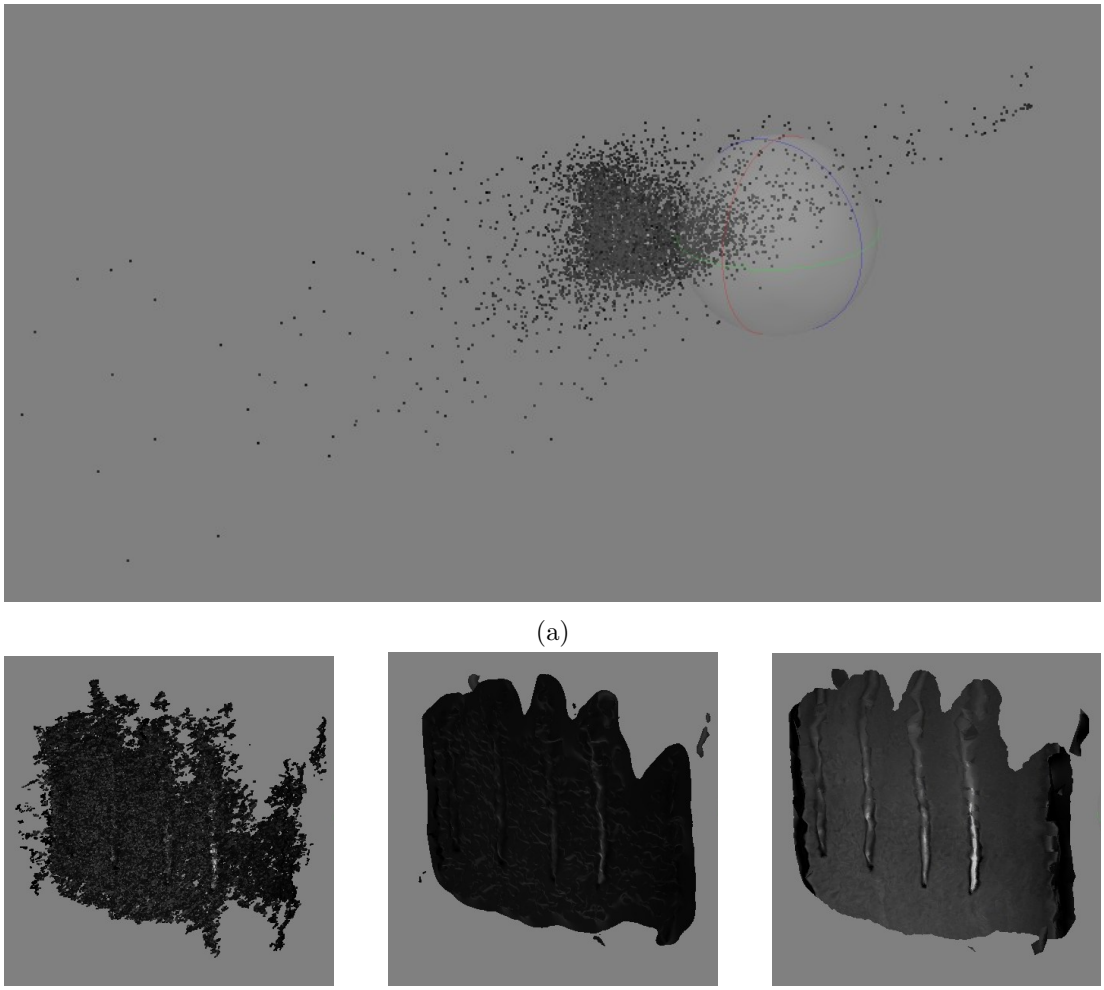
fuel channel. The first layer contains lots of features however the second layer is almost featureless and will challenge image processing and reconstruction algorithms. The video is decomposed into 1511 images.

3.4.2 Proprietary frameworks

The first portion of testing was dedicated to evaluating proprietary software solutions. These solutions are all closed source and the underlying algorithms to produce the 3-D representations are unknown. However, since the solutions often come in the form of easy to use interface, this was inherently the quickest form of framework to test. Photosynth, Autodesk 123D Catch and Acute3D were unable to produce any meaningful representation. Agisoft PhotoScan was the only tested proprietary framework that was able to produce any tangible representation of the input footage with the underlying statistics of the model being obscured due to the assessed software being a trial-based solution. The results of which are seen in Figure 3.9 using B1 footage and Figure 3.10 using B12 footage respectively.

PhotoScan when evaluating B1 footage resulted in an sparse point cloud as seen in Figure 3.9a) representing only 7% (6,875/98,169 points successfully triangulated) of the entire footage which was highly deformed and containing numerous noisy points that do not adhere to the AGR fuel channel structure. During dense reconstruction, more points are produced using a dense matching scheme resulting in a dense 3-D point cloud as seen in Figure 3.9b) containing 279,509 3-D points in a matching process that took approximately 2 hours and 20 minutes. PhotoScan also has the ability to produce a surface reconstruction and a textured model which can be seen in Figures 3.9c) and 3.9d) respectively. The B12 footage results in a different reconstruction being produced with many of the sparse points somewhat adhering to the AGR fuel channel geometry with the resulting 3-D representations being shown in Figure 3.10. It is clear from both sets of results that the model initialisation procedure used within PhotoScan is based on the density of overlap, and would explain why both reconstructions initialise and produce reconstructions of different parts of the assessed inspection footage, despite

both sets of footage sharing its own intrinsic overlap.



(a)

(b) (c) (d)

Figure 3.9: 3-D Representations produced by the Agisoft Photoscan Professional software using the B1 footage. a) The sparse point cloud; b) Dense point cloud; c) Surface reconstruction; d) Textured model.

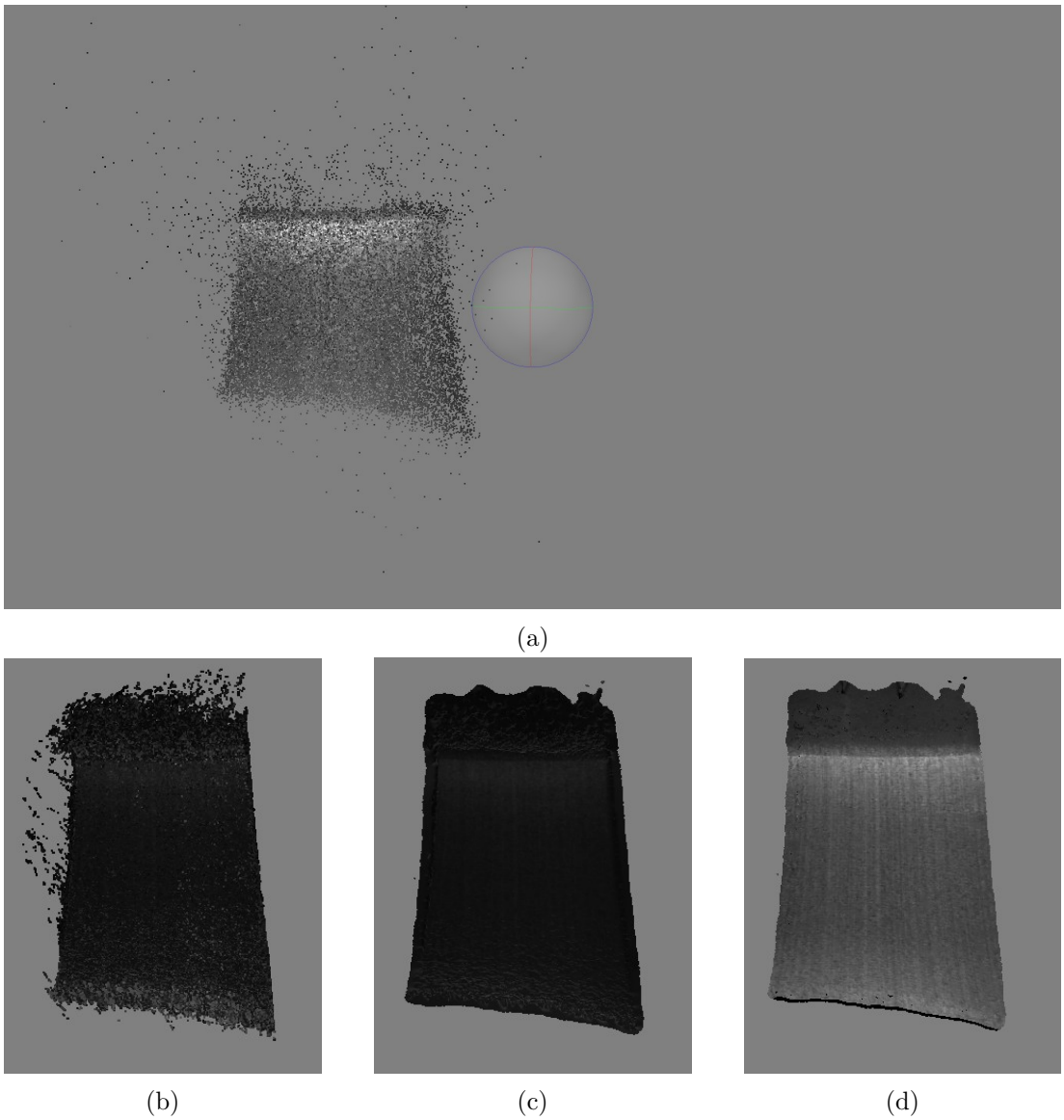


Figure 3.10: 3-D Representations produced by the Agisoft Photoscan Professional software using the B12 footage. a) The sparse point cloud; b) Dense point cloud; c) Surface reconstruction; d) Textured model.

3.4.3 Open-source frameworks

Open-source reconstruction frameworks are a better method of evaluation since they often incorporate state-of-the-art algorithmic approaches and also provide or utilise easily accessible, open source code. The best frameworks, either from a historical or current stand-point were tested include Bundler, VisualSFM, OpenMVG, MVE and COLMAP

due to their performance on generalised benchmark data [18,19,21,25,84,118]. Like the commercial entities however, nearly all open-source SfM frameworks struggled to produce any usable reconstructions with both incremental reconstruction methods VisualSFM and COLMAP being the only frameworks producing tangible results. Therefore, these frameworks will be the focus of discussion.

The first evaluated framework is VisualSFM [25] using the B1 footage and as seen in Figure 3.11a) and 3.11b) produced a near "complete" reconstruction with over 90% (610/681 images were used during reconstruction) of the visible scene being recovered which produced 7176 3-D points. The dense reconstruction unlike PhotoScan has a simple implementation of MVS and as a result, the dense reconstruction is patchy and demonstrably less dense than the sparse reconstruction. Performing a reconstruction using the B12 footage however yields very interesting results from a theoretical standpoint and backs up the scenario presented in Chapter 4.2.1 and can be in Figure 3.11c) and 3.11d). The results produced are very similar to that of the B1 footage and this is not because of the framework failing to further produce a reconstruction once it hits the brick interface and goes into the second layer, but instead is a case where that Layer 2 is being mistakenly identified as the Layer 1 of the brick. This can be demonstrated below in Figure 3.12 with VisualSFMs ability to view the visibility matrix in the form of a heat-map:

The visibility matrix provides an interesting insight into when pseudo-cyclic or repetitive scene elements are introduced into the reconstruction phase. The sequential matching procedure was implemented during this test, however, it is clear from the visibility matrix that the sequential matching procedure is not properly implemented. Since VisualSFM is closed-source, no code-based investigations can be carried out. Therefore, it provides an interesting case scenario when performing exhaustive matching and the issues it causes when trying to reconstruct environments that are repetitive in nature.

COLMAP [21], like VisualSFM, is an incremental SfM framework designed for generic reconstruction environments. Like VisualSFM, it contains a suite of functionality allowing for general SfM in addition to dense and mesh reconstruction. Both sets

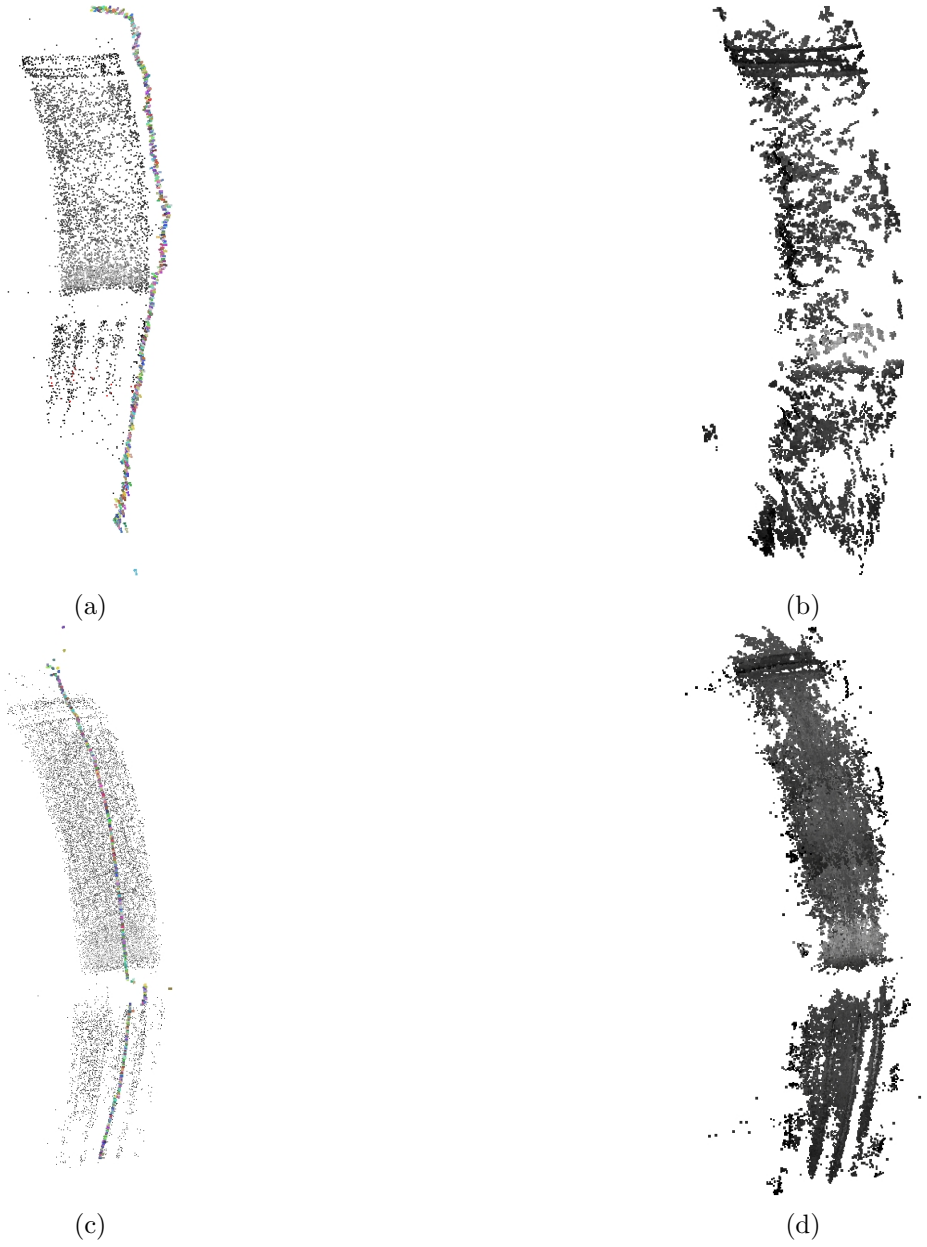
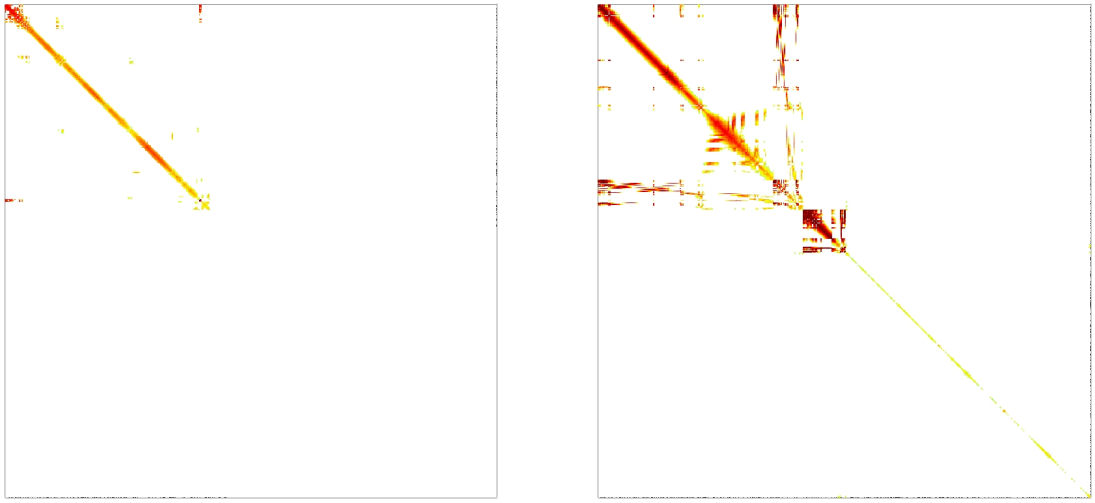


Figure 3.11: 3-D Representations produced by VisualSFM using the B1 (a) & b)) and B12 (c) & d)) footage. a) The sparse point cloud; b) Dense point cloud of the B1 footage; c) The sparse point cloud; d) Dense point cloud of the B12 footage.

of footage were fed into COLMAP with B1 footage producing excellent results despite only using 60% of the images (400/681 registered images), producing a reconstruction with 12183 3-D points. However, B12 produces 4 separate models with 139/1512 images and 10971 3-D points, 140/1512 images and 10788 3-D points, 399/1512 images



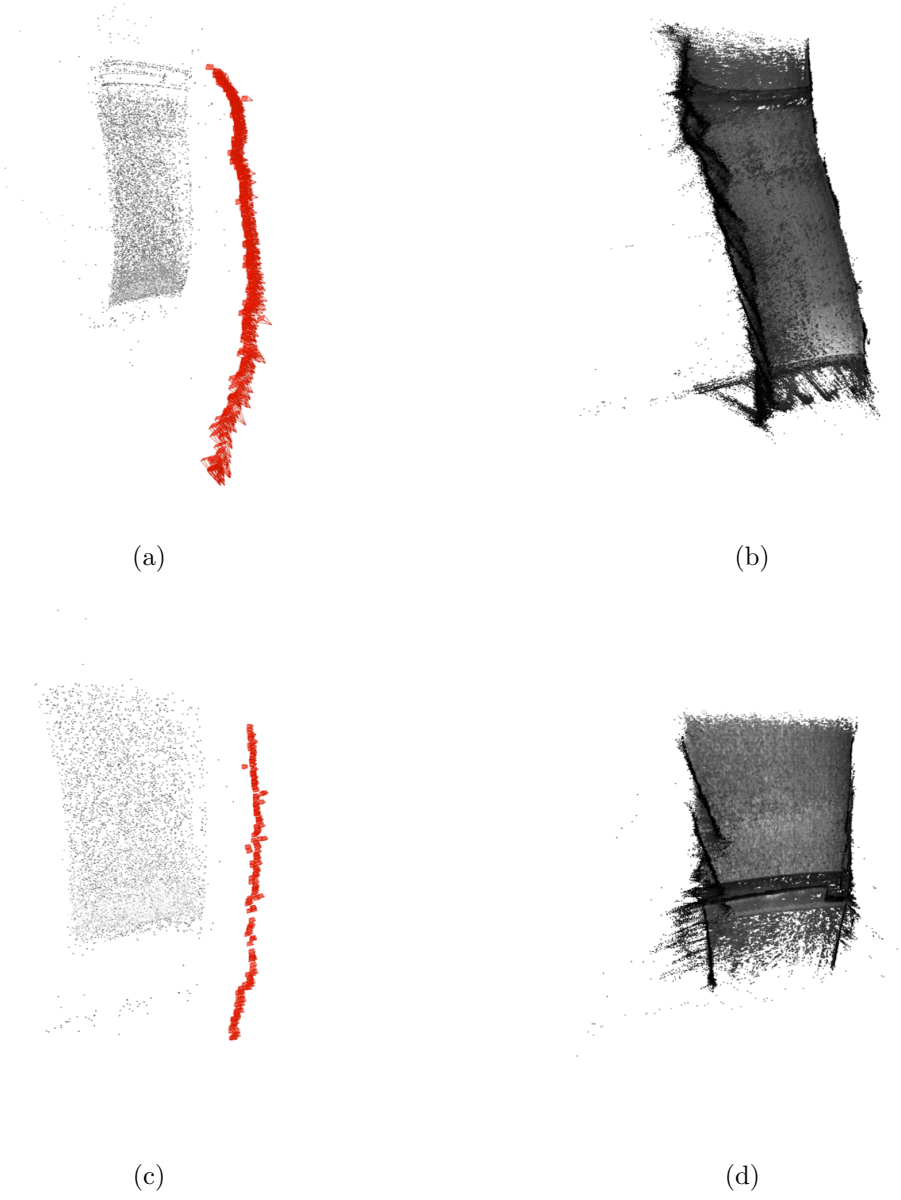
(a) (b)

Figure 3.12: Visibility matrices produced by VisualSFM using a) B1 footage and b) B12 footage. Note how b) overlaps on a) instead of continuing the diagonalisation of the visibility matrix.

with 12337 3-D points and 141/1512 images with 6892 3-D points respectively, with 55% scene coverage (839/1512 registered images) with 40,988 3-D points overall. These models, however, cannot be combined and as a result, each 3-D model is self-contained in its own model. The dense 3-D models as viewed in Figure 3.13b) and d) introduce far more points with 1628281 and 564222 3-D points for B1 and B12 footage[1] respectively. To investigate this further, the results of both sets of footage can be observed in Figure 3.13 and their respective visibility matrices in Figure 3.14.

As can be observed in Figure 3.14b), the visibility matrix obtained from B12 footage in Figure 3.14a) is a sub-matrix pertaining to the same images within the B1 footage, with the model being produced containing 399/1512 images with 12337 3-D points in the B12 footage but in the same area with regards to the B1 footage using 400/681 images, producing a reconstruction with 12183 3-D points. Across the diagonal however, there is a substantial gap and this pertains to a challenging feature area observed in the mid-brick layer between the first layer and the second brick layer and as a result, no features are reconstructed in this area. The second sub-matrix comprises of the other three

[1]The dense reconstruction of the 4^{th} model is shown and processed since the $1^{st}/2^{nd}$ models are effectively the same reconstruction

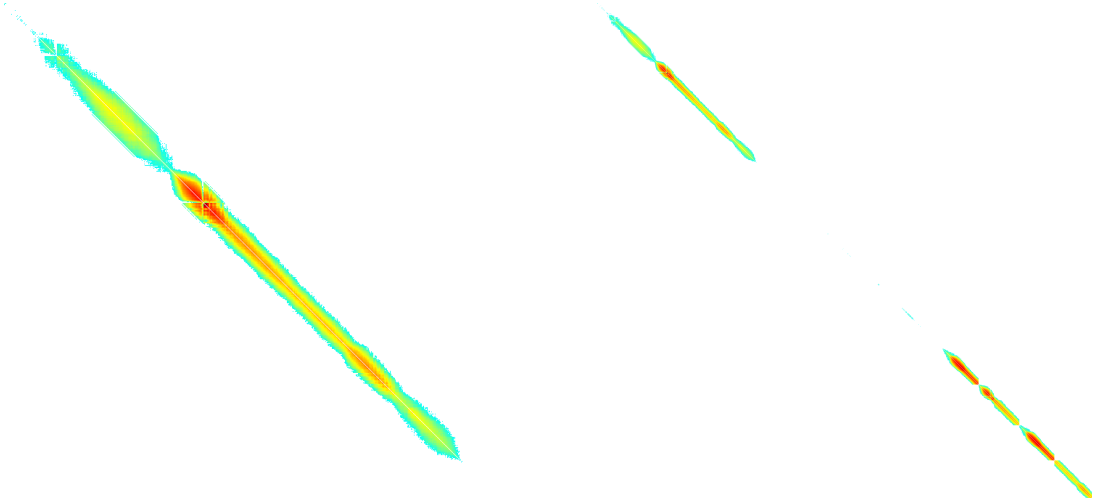


(a) (b)

(c) (d)

Figure 3.13: 3-D Representations produced by COLMAP using the B1 (a) & b)) and B12 (c) & d)) footage. a) The sparse point cloud; b) Dense point cloud of the B1 footage; c) The sparse point cloud; d) Dense point cloud of the B12 footage.

models that simply visualise the same brick interface repeatedly in the three separate models. This points to the difficulty of being able to obtain reliable reconstructions of challenging feature areas and highlights the value of the work produced in Chapter 4. Furthermore, it highlights the value of taking the incremental approach to performing



(a) (b)

Figure 3.14: Visibility matrices produced by COLMAP using a) B1 footage and b) B12 footage.

reconstruction over other forms of image-based 3-D scene reconstruction.

3.4.4 Academic SLAM frameworks

Another approach to image-based reconstruction tested was the use of open-source SLAM frameworks with particular emphasis on the two state-of-the-art methods - LSD-SLAM [122] and ORB-SLAM [123] due to their uniquely different approaches. LSD-SLAM operates in a purely direct manner (i.e. Every pixel in the image) and ORB-SLAM operates in an indirect manner (i.e. Using descriptive feature sets from the image) to obtain camera pose and scene geometry in real-time. Each SLAM framework was tested in Ubuntu 14.04 however both frameworks either simply failed to initialise or would struggle to find matches. The primary reason for this is the approach sensitivity to factors such as camera calibration, camera noise and environments with non-uniform lighting due to the incremental nature of the map building process. Since camera calibration parameters are unavailable in addition to the feature-less environment and the changes in lighting within B1 and B12 footage, no evaluated SLAM framework produced any meaningful representation of the AGR channel as previously mentioned in Chapter 3.4.1.

3.5 Discussion

This chapter opened the discussion on image-based 3-D scene reconstruction by describing the basics in a visual manner, before performing a review of the historical evolution of SfM, contemporary SfM frameworks and application specific methods. A critical review of the techniques in Section 3.2.5 demonstrated that many general purpose reconstruction approaches are generally unsuitable, with the application-based approaches discussed highlighted numerous issues that occur when applying image-based 3-D reconstruction algorithms to challenging imagery. The hardware and theoretical image processing limitations pose a considerable challenge and therefore, an investigation is carried out in Section 3.3 - 3.4 by applying state-of-the-art reconstruction algorithms to AGR RVI footage. In Table 3.1, a list of both open-source and proprietary methods were tested, with Agisoft PhotoScan [115], VisualSFM by Wu [25] and COLMAP by [21] being the only methods capable of producing reconstructions from the two subsets of imagery used for evaluation.

Through experimentation, VisualSFM and COLMAP produced the strongest results, with both frameworks being the incremental SfM archetype. Both frameworks, although producing somewhat partial reconstructions, demonstrated their ability to cope with challenging imagery. One of the major points of discussion however is that the point-cloud models suffered from deformation, or failed to reconstruct areas that were inherently feature-poor. On further investigation, the feature matching approaches used by both reconstruction frameworks appeared to struggle in mid-brick layers, with non-uniform illuminated images or changes in lighting between the imagery. This correlates with observations made in application-based papers posited by El Kahi et al [12] and Esquivel et al [16] where changes in lighting or areas of poor texture causing the 3-D point cloud model to drift and causes the channel reconstructions to warp. Due to the generality of both VisualSFM and COLMAP, both methods fail to address a key issue of intra-image issues introduced by challenging environments where the imagery captured is inherently sub-optimal, which then causes ambiguity at the matching stage, then deformation of the model at the reconstruction stage.

To alleviate or improve the reconstruction quality, many frameworks include MVS which produces photo-realistic reconstructions using dense correspondence algorithms such as PatchMatch or DeepMatching [64, 65] which are guided by the underlying relationships derived on the sparse 3-D model. This logic, however, is fundamentally flawed because if the underlying sparse model suffers from deformation, the resultant 3-D dense reconstruction effectively compounds the error of the derived sparse 3-D point cloud. The fundamental weakness of these algorithms in repetitive or feature-poor environments is the overreliance in what will be termed in this thesis as "anchor points". Anchor points are low frequency, unique textural intra-image components which are easily distinguishable across a multitude of viewpoints and often, the spatial distribution of the strongest matches will reside predominantly in this area and will help proceeding methods, particularly geometric verification of image matches. As observed within the application domain techniques, every single surveyed method available in literature exploit the use of sequential (incremental) reconstruction techniques as matching across small base-lines hugely simplifies the matching problem [13, 15, 46]. This technique is also exploited in SLAM and VO methods [22,122,125] for constrained or urban environments. Therefore, there is a drive to produce a bespoke 2-D correspondence algorithm which can effectively negate the deficiencies caused by lack of texture and non-uniform illumination by exploiting the contiguous nature of the image capture process. This forms the primary motivation for the work produced in Chapter 4, where a novel semi-dense matching scheme is introduced to suppress the issue of working with sub-optimal imagery.

Chapter 4

Structure-from-Motion using deficient feature spaces

Summary of chapter contributions

Two scenarios are presented with regards to performing image-based 3-D scene reconstruction within challenging environments, indicating issues with ambiguity from the correspondence searching stage when using feature-poor or repetitive imagery. With erroneous or ambiguous matches, it results in the resulting 3-D point cloud being misrepresentative of the target scene. This is discussed thoroughly at both the image and feature representation level, indicating potential sources and solutions to improving feature matching in challenging scenes. The chapter introduces a novel correspondence searching algorithm called Motion-based Iterative Feature Matching (MIFM), which incorporates both direct (Dual TV-L^1) and indirect (SIFT) feature representations to obtain dense motion and salient image features. An intra-image regional strength determination mechanism is used to quantify the matching quality within regions, which allows for suppression of regions that will produce unstable or incorrect matches. This method is designed to enhance correspondence searching in contiguously captured imagery, and allows for improved matching quality and density when using imagery that is feature-poor or contains repetitive structures.

4.1 A Problem Definition

The derivation of 3-D geometric structures from 2-D image datasets is a task that is computationally complex. One of the fundamental challenges posed by applying 3-D

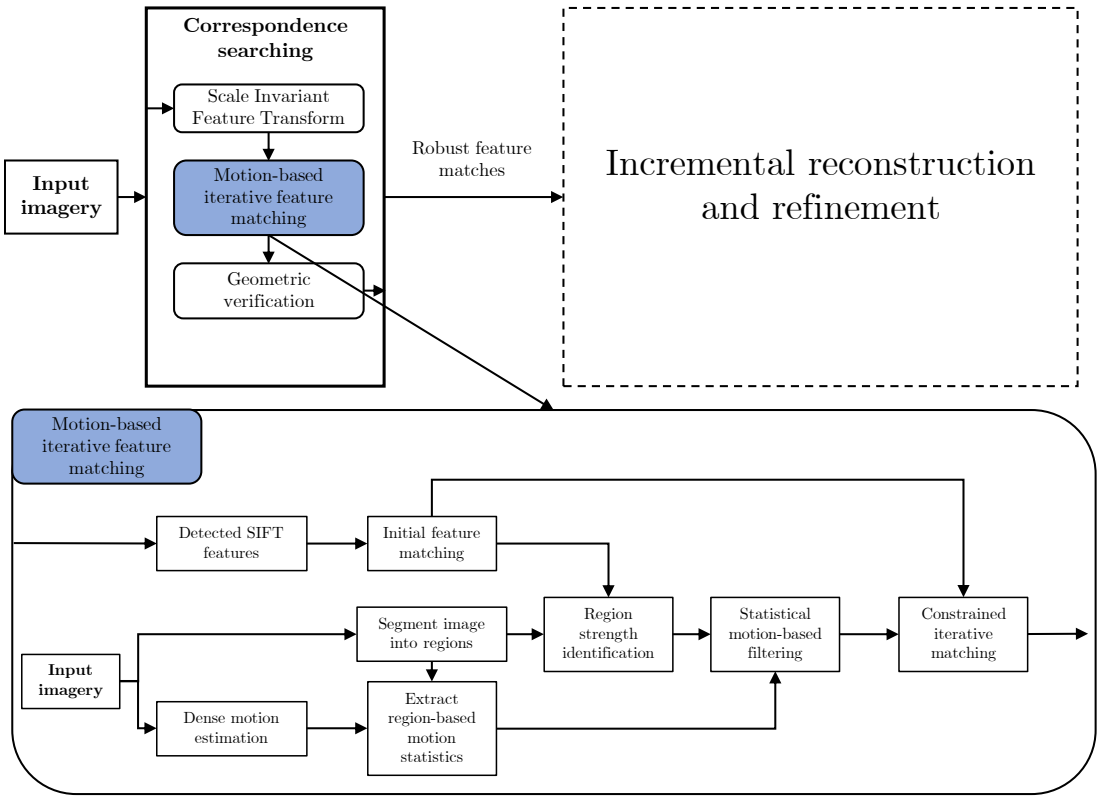


Figure 4.1: Proposed Motion-based iterative feature matching approach

reconstruction algorithms to challenging footage is the derivation of accurate, robust feature matches. When the derived feature matches are incorrect or ambiguous, the resultant 3-D reconstruction becomes distorted, resulting in a misrepresentation of the target scene. Consequently, it is important to look at the underlying issues which cause ambiguity; the feature representation and the feature matching process. These fundamental issues are not unique to Structure-from-Motion but are relevant many classical image processing techniques which require highly descriptive and feature-rich imagery such as motion estimation [113, 126], Visual Odometry (VO) [127, 128] and Simultaneous Localisation and Mapping (SLAM) [123].

The discussion in this chapter focuses on correspondence searching, a key component of a SfM framework in order to derive accurate matches. The chapter begins by contextualising the issues caused by performing image-based 3-D reconstruction using

deficient feature spaces, and the resulting ambiguities caused by feature homogeneity and matching ambiguity using both general and application-specific examples. The purpose of this is to understand why the correspondence searching procedure struggles using RVI imagery at a deep, fundamental level and what can be done to suppress or even circumvent the issues caused. In Chapter 4.3, a novel approach to feature matching visualised in Figure 4.1 is introduced by combining the strengths of both sparse and dense feature matching to determine regions of stability within the image which can be used to guide the correspondence search procedure across the image to acquire a robust set of feature matches. This chapter will discuss the methods proposed to deal with such challenges which are characterised by an encompassing term called a *deficient feature space*.

4.2 Deficient feature spaces

4.2.1 Feature Homogeneity

Feature homogeneity is one of the challenges when determining the structural composition of a scene or object and is a direct result of the environment in which the images have been captured. Homogeneity of the features occurs when the assessed image dataset has repetitive structures or suffers from textural uniformity, resulting in feature descriptors such as SIFT or ORB [60,109] generating analogous feature vectors. Consequently, ambiguity can be introduced into the system. Therefore it is important to define what characteristics constitute a good feature for image-based 3-D reconstruction with a list presented below:

Saliency *A determination of distinctness of a feature within an image.*
If the processed image contains a large amount of unique and salient points or regions, image-based 3-D reconstruction techniques will operate optimally and be able to determine structure and motion with high degrees of accuracy. Conversely, if an image contains a low degree of saliency, techniques such as Structure-from-Motion will struggle to produce an accurate representation of the scene due to the textural

uniformity of the image. Visual or feature saliency can also be interpreted in the frequency domain [129], with salient points being high-frequency components whereas areas of image homogeneity will be regarded as low-frequency. Furthermore, the occurrence of repetitive image structures can also be determined by observing the amplitude spectrum.

Repeatability *The ability in which the same feature can be found despite undergoing geometric and photometric transformations between images.*

Repeatability is especially important with image datasets which are unordered and/or obtained from multiple imaging devices as the image features can be highly diverse due to the variability in view-point, illumination, environmental conditions, image quality or its appearance [21, 74, 86]. With contiguously captured image datasets such as videos, repeatability can be problematic if surfaces or objects captured in the images have non-Lambertian (i.e. Lambertian reflectance indicates the light diffusing equally in all directions, non-Lambertian reflectance indicates the light diffusing varies in different directions) properties [130], specularities (e.g. light reflecting off a window), non-uniform illumination [131] or inconsistent camera parameters [73].

Locality *A feature that occupies a small area of the image and is observable despite clutter and occlusion.*

If a feature or object within the observed scene is distinguished and salient whilst taking up very little area in the image, it becomes important and effectively 'anchors' any correspondence mechanisms which processing the feature from multiple view-points. Locality is important when evaluating non-rigid scenes (e.g. producing a 3-D reconstruction of a street containing many pedestrians) [87,132,133] or even semantic segmentation of images [134, 135] of environments which are particularly challenging

Efficiency *A feature or group of features that are a compact, salient, abstract representation of an image whilst being computationally inexpensive to compute and or process.*

Due to computational limitations when reconstructing scenes with a large number of input images, the representation of the features, either indirect[1] or direct, have to be taken into account. Indirect feature descriptors such as Scale Invariant Feature Transform (SIFT) [60] or Speeded-Up Robust Features (SURF) [136] reduce an image into a abstract set of salient points. Binary feature descriptors such as Binary Robust Invariant Scalable Keypoints (BRISK) [137] or Oriented FAST and rotated BRIEF (ORB) [109], reduce computational complexity dramatically when processing images [138]. Conversely, direct methods sample all pixels within each acquired image and requires no feature extraction phase [122, 139] and operates purely on the image intensity values, meaning no image data is discarded and all the content of the image can be used [122]. Hybridised methods often termed "semi-direct or semi-dense" have also been proposed as a method of exploiting the computational efficiency of feature-based methods whilst obtaining the descriptive power of direct approaches [122, 127]

For optimal and computationally efficient performance of an image-based 3-D reconstruction technique, the input image dataset has to contain features which exhibit all of the aforementioned properties. To illustrate the first three characteristics, Figure 4.2 shows a dense reconstruction of an feature-rich object surrounded in a featureless environment which contains a direct contrast between a black drawer and a white wall, a microcosm of the challenges faced while reproducing a scene from images. The first characteristic, saliency, is exemplified with the statue containing a large number of features which is juxtaposed with the black drawer and white wall which are both uniform in colour and the stand of the trophy which all contain very little usable and

[1] Indirect refers to a sparse, feature-based descriptor whereas direct refers to dense representations of each pixel in a scene [128]

inconsistent features. Repeatability is also exhibited as the resulting reconstruction in Figure 4.2f) highlights the comprehensive reconstruction and the ability for the object to be found whilst subjected to photometric and geometric transformations. To test the locality and repeatability properties, the object was subject to different lighting

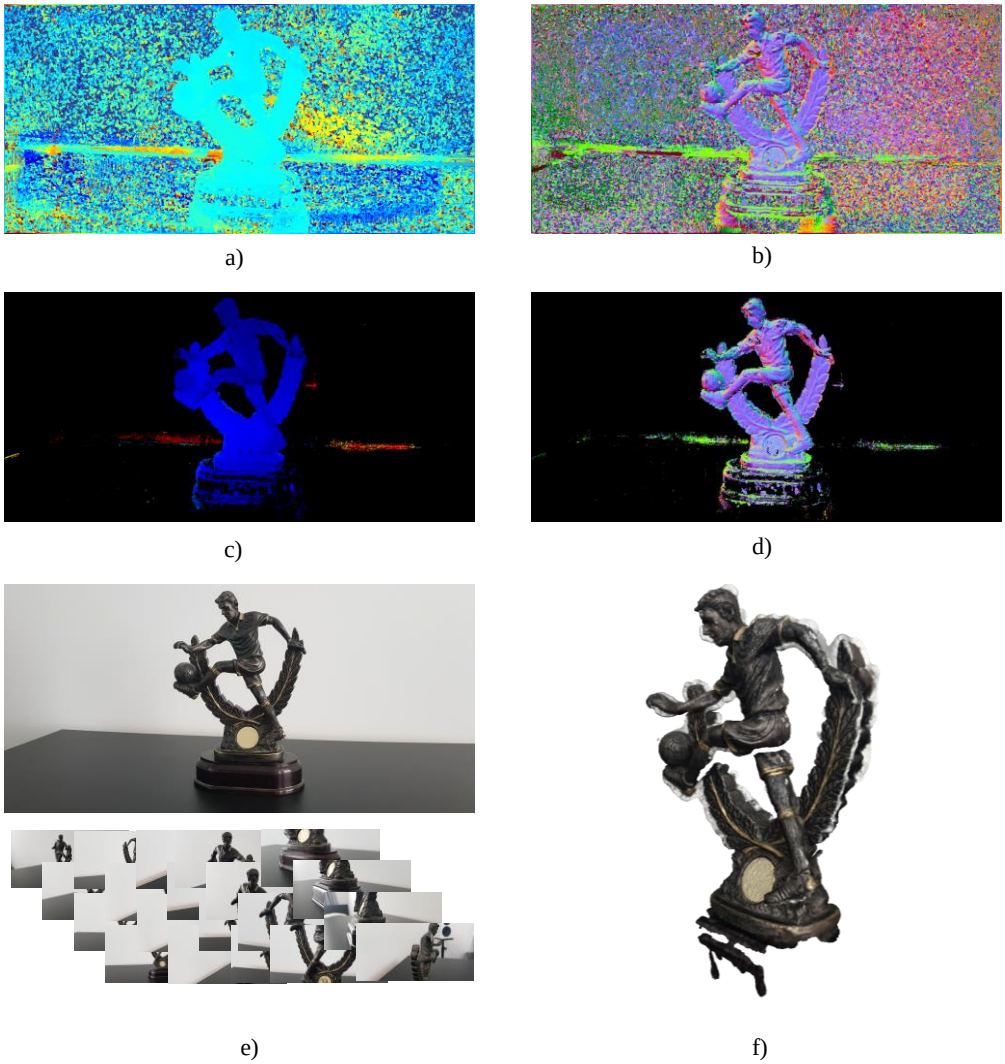


Figure 4.2: a) Photometric depth map of the trophy; b) Photometric normal map of the trophy; c) Geometric depth map of the trophy; d) Geometric normal map of the trophy; e) Original image and dataset used to produce a 3-D reconstruction of the trophy; f) A 3-D Poisson surface reconstruction of the trophy.

conditions in addition to the trophy stand containing a specular reflection in all images from an independent light-source. As expected, the drawer, wall and the trophy stand regions are weakly supported and contain very little to no tracked features in these regions which is verified by examining the geometric depth and normal maps in Figure 4.2c) and d). This example demonstrates a fundamental problem when dealing with image regions which exhibit intra/inter-image feature homogeneity and consequently, result in weakly-supported or missing regions when examining the resultant point-cloud or surface reconstructions.

Due to the locality and saliency of the object in the scene observed in Figure 4.2, this effectively 'anchors' the produced 3-D reconstruction. With regards to academic literature, post-processing of the dense 3-D reconstruction is performed in order to interpolate or fill in weakly supported regions that originate from homogeneous or featureless areas of the scene [140,141], consequently relying on an salient object or feature region to anchor the correspondence searching reconstruction processes to produce a dense reconstruction of the scene. Herein lies a significant challenge that is still to be addressed - "What if the target scene contains repetitive structures and/or if the target scene has no salient objects in which to anchor the rest of the reconstruction upon?" To investigate this problem, two different scenarios are proposed. The first scenario contains repetitive structures in the form of a number of objects which are the same and the second scenario contains a scene with no salient objects or features.

Scenario 1 - Repetitive Structures

This scenario introduces the use of multiple objects which are structurally identical and placed in a simple, repeated pattern on the same desk surface observed in Figure 4.2, captured from multiple viewpoints with a image of the target scene illustrated in Figure 4.3a). The structures used within this environment exhibit the first two characteristics; saliency and repeatability and the geometric depth and normal maps derived from the image of the target scene demonstrate these characteristics in Figure 4.3b) and c). The resultant reconstruction of the target scene observed in Figure 4.3d) however is indicative of a reconstruction with repetitive or cyclic elements within the environment

and this originates from the following:

Concordance In a normal target scene, it is unlikely to contain repetitive structures which have no dissimilarity with regard to the structure or the immediate surrounding area, making this scenario unique with regards to image-based 3-D reconstruction. When evaluating a scene, local feature descriptors such as SIFT will incorporate neighbouring pixels into each individual features key-point descriptor to improve feature observability whilst making it robust to photometric and geometric changes [60,61]. However, if the feature/structure and the neighbouring area observed in the image has poor locality, saliency and repeatability

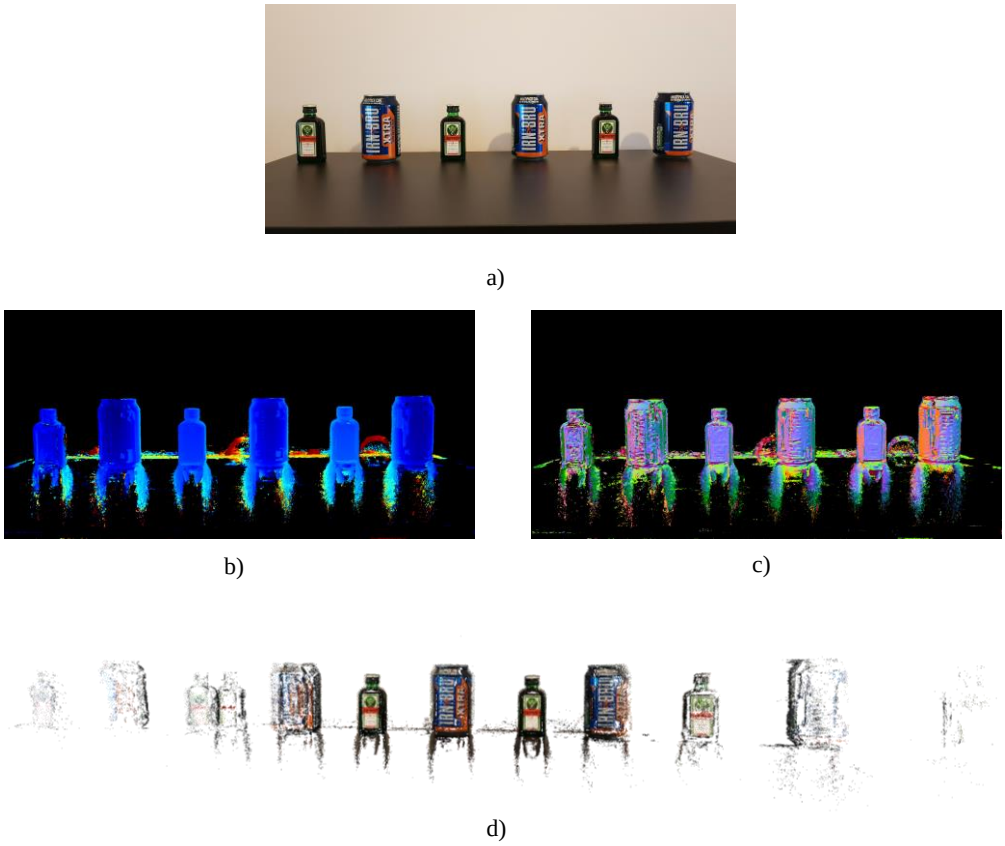


Figure 4.3: a) Image of the target scene; b) Geometric depth map of the scene; c) Geometric normal map of the scene; d) A 3-D Poisson surface reconstruction of the scene.

characteristics, the generated feature descriptors regardless of viewpoint will demonstrate high similarity, especially within repetitive or cyclic environments.

Ambiguity The primary issue with repetitive or cyclic structures observed within an image dataset occurs when the system is trying to search for correspondences [112]. If an image consists of a constrained viewpoint of the scene (e.g. an image contains a view of a single structure within the scene), determining what structure that the image observes without having ground-truth information or integrating knowledge into the matching process results in feature ambiguity. This issue will be discussed in greater length later in this chapter.

Scenes with high degrees of concordance are proportional to the resulting ambiguity produced from the correspondence searching mechanism. The dense reconstruction proposed by this scenario exhibits a repetitive yet salient scene, however, the repetitive nature of the scene directly results in deformation or misrepresentation of the reconstructed target scene as observed in Figure 4.3d), consequently reconstructing apparitions that do not appear within the original scene as seen in Figure 4.3a).

Scenario 2 - Repetitive Indiscriminate Structures

This scenario introduces the use of indiscriminate structures where the features observed within the image demonstrate concordance due to the observed scene having no discernible objects which are salient. Furthermore, the image intensities and textural content of the image demonstrating homogeneity. This is the most challenging scenario since there is no visually discernible depth information available, an example of this is shown in Figure 4.4 where images are taken primarily of the wall observed in the previous examples of Figure 4.2 and 4.3. In this example, due to the lack of features, no 3-D representation can be determined from the three images using any available image-based 3-D reconstruction framework. The uniformity of the images can be observed in Figure 4.4b) where surface plots using image intensities demonstrate textural

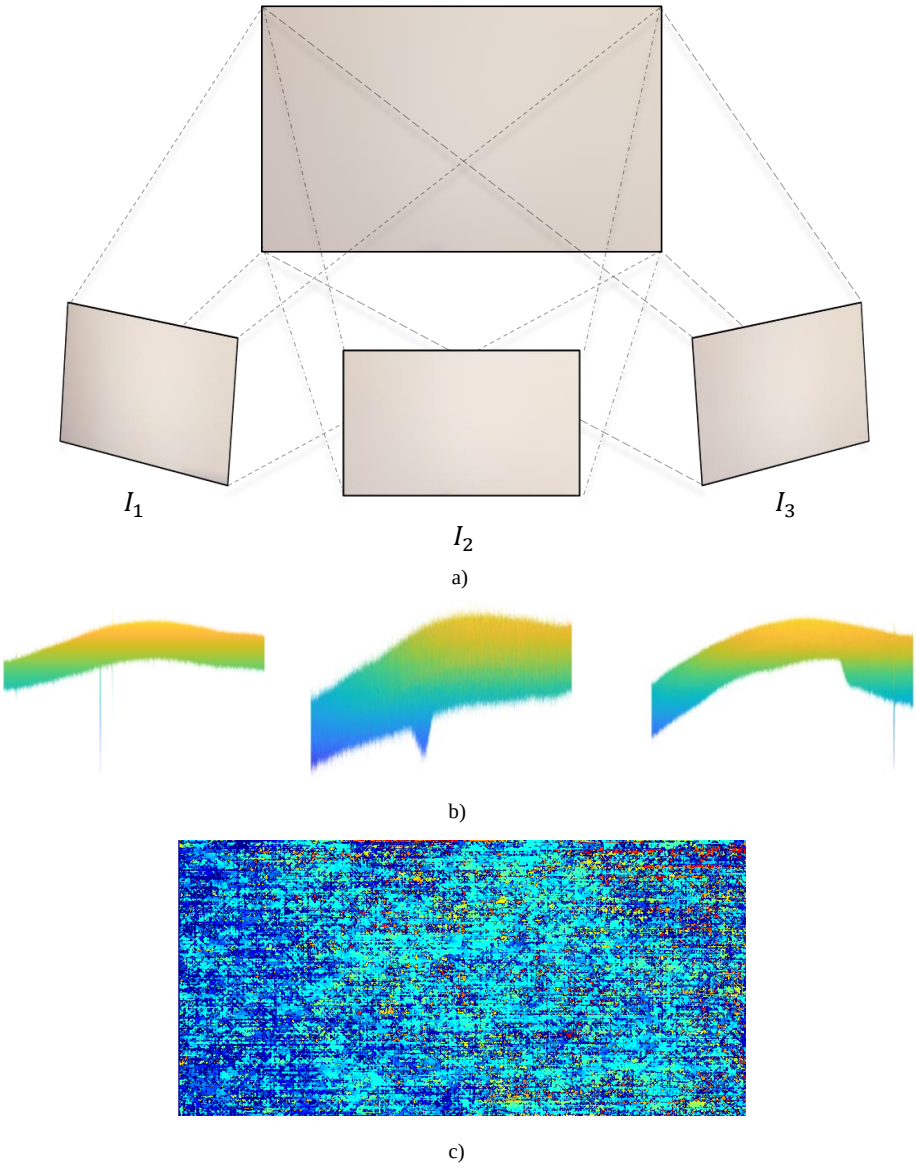


Figure 4.4: a) Ground-truth movement between the three images; b) Surface plots of the respective intensity values for each image; c) Disparity map between I_1 and I_3 (Cold colours represent little disparity $\longrightarrow$ Hot colours represent high disparity)

uniformity and by determining the stereo disparity between I_1 and I_3 using a block matching approach that utilises the Sum of Absolute Differences (SAD) [142] further demonstrates the homogeneous nature of the scene. To derive or retrieve 3-D geometry from a scene with very limited heterogeneity, techniques such as Shape-from-Shading (SfS) or photometric stereo can be used given one image or multiple of a scene respec-

tively, exploiting pixel intensity, knowledge about the position of light-sources within the scene or Lambertian reflectance properties to derive depth [143, 144].

4.2.2 Feature Ambiguity

Ambiguity is described in the previous section as a fundamental problem when dealing with scenes with repetitive structures, with the underlying problem instigated in the feature extraction phase when searching for a candidate set of matches. The origin of ambiguity within an image-based 3-D reconstruction pipeline resides within the data association stage of Structure-from-Motion (SfM) and it is termed in literature as the correspondence searching stage. The indeterminateness of the search for correspondences between detected features depends on two stages of the matching process a) the matching of images and b) the matching of the detected candidate feature sets corresponding to each image. Although these stages are intrinsically linked with image matching being an abstraction of feature matching, for the purpose of this thesis, these stages will be decoupled and treated as separate mechanistic entities.

Image matching ambiguity

There are various image matching paradigms available within literature with the suitability being dependent on the structure and possible meta data of the processed input image dataset $\mathcal{I} = \{I_i \mid i = 1, ..., N_i\}$. Typical methods such as exhaustive, sequential, transitive and spatial matching as demonstrated by Figure 4.5 are amongst the most common techniques used due to their generality and applicability across a variety of datasets, and occur in many general purpose reconstruction pipelines [21, 95]. Exhaustive matching of an image dataset is arguably the most comprehensive matching methodology to determine scene overlap in input images. Given the detected image feature sets $\mathcal{F}_i = \{(\mathbf{x}_j, \mathbf{f}_j) \mid j = 1, ..., N_{F_i}\}$ where $\mathbf{f}_j$ is the appearance descriptor and $\mathbf{x}_j = [x, y] \in \mathbb{R}^2$ is the detected feature locations, an exhaustive brute-force approach has the computational complexity of $O(N_i^2 N_{F_i}^2)$. Although this approach is suitable for datasets which are unordered, as the number of images N increases, the feasibility of such exhaustive approaches becomes untenable. Techniques to compress datasets

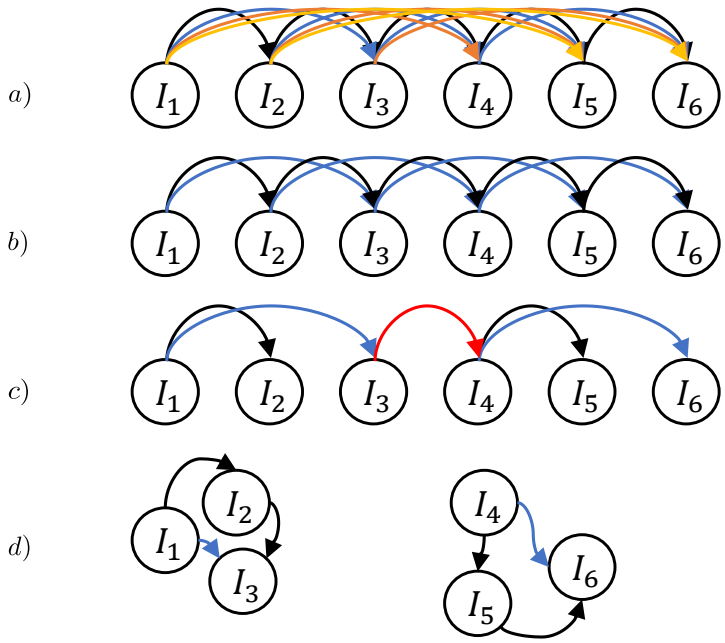


Figure 4.5: Visualisation of the typical image matching methodologies used within literature. a) Exhaustive matching; b) Sequential matching; c) Transitive matching; d) Spatial matching.

by exploiting image characteristics such as texture, metadata and colour have been deployed to remove redundant images from unordered data-sets whilst maintaining the intricacy of the target scene have been proposed in literature [20, 84]. Nonetheless, these techniques can introduce ambiguity if the source images environment demonstrates consistent characteristics or in particular, if the environment is cyclic [112], making exhaustive matching sub-optimal for challenging feature spaces.

Contiguously captured or spatially coherent image datasets take a different approach, where the matching procedures inherently exploiting scene overlap of the neighbouring images in the data-set[2] to reduce the search space. Neighbour-based image matching paradigms are particularly common within VO [128, 145] and SLAM [108, 122] which operate on contiguous video sources, often exploiting a filtering approach [121, 146] or a keyframe-based non-linear optimisation framework [125, 147] to match and enforce geometric consistency between matched images. Sequential match-

[2]This is under the assumption that the image data-set is ordered in the manner of capture

ing using an offline image-based 3-D reconstruction approach such as SfM can utilise simple sequential or spatial matching to restrict the search space to a specific window of images as demonstrated in Figure 4.5b) and d), reducing potential scene coverage but improving computational efficiency from $O(N_i^2 N_{F_i}^2) \longrightarrow O((N_{\text{SEQ}} N_i) N_{F_i}^2)$ where N_{SEQ} is the number of neighbouring matches to be made. Methods such as transitive matching can then be used as a form of extending image matches outside these identified neighbourhoods as visualised in Figure 4.5c) by matching image sub-sets to find further possible image matches. For example, $\mathcal{M}_{I_1} = \{I_2, I_3\}, \mathcal{M}_{I_3} = \{I_4\}$ are two matched image sub-sets and $\exists\ I_n \in \mathcal{M}_{I_1}$ where I_n is a shared element of both subsets, then the union of both sub-sets $\mathcal{M}_{I_1} \bigcup M_{I_3} \implies \mathcal{M}_{I_1} = \{I_2, I_3, I_4\}$[3].

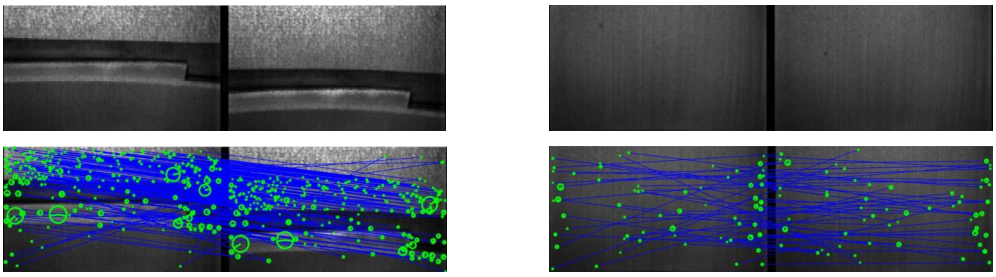
4.2.3 Feature matching ambiguity

In Section 4.2, three scenarios were proposed utilising a image-based 3-D reconstruction framework and examined in accordance with ideal feature characteristics. In this section, it will examine the feature matching paradigms between candidate feature sets or pixel intensities for each image and demonstrate possible inceptions of ambiguity. Feature matching ambiguity arises in different forms based on the underlying method of feature extraction. There are two types of feature descriptors available in academic literature; hand-crafted or learned local feature descriptors (termed in literature as indirect), direct methods which operate on pixel intensities or a hybridisation of both. The following sub-sections will discuss the difficulties pertaining to each form of feature representation in context with the target application for this thesis, namely the AGR fuel channels.

Local feature descriptor techniques (Indirect)

Chapter 3.2.1 introduced the types of different local feature descriptors and the variety of different approaches in literature regarding hand-crafted [60, 148] or learned feature representation [149, 150] and matching within an ideal, heterogeneous environment. In texturally repetitive, cyclic environments, however, many matching procedures of

[3] If A = B and B = C then transitive matching assumes A = C



a) *b)*

Figure 4.6: SIFT key-point matching using the squared Euclidean distance (Equation 4.1) and Lowes ratio [60] using a) Salient image pair acquired at a brick interface and b) Feature-poor image pair acquired mid-brick layer.

the underlying features either fail to disambiguate mismatches or fail to produce a set of robust matches. So, how does this problem arise? To demonstrate the issue, the descriptor space of a hand-crafted feature detector and a learned descriptor will be compared.

SIFT operates by first detecting "blobs", an informal name given to local extrema values in the Difference of Gaussians (DoG) scale space [60], and subsequently extracts a circular image region containing a geometric frame $\mathcal{F}$ of three parameters: the key point location $\mathbf{x}_j$, key-point scale σ and orientation θ. These parameters effectively give a 3-D spatial histogram of the image gradients with the gradient at each pixel formed by $\mathbf{x}_j$ and θ. The histogram for each detected key-point consists of 128 bins ($8 \times 4 \times 4$), which is represented as a single 128-dimensional vector. To match detected feature-points, the squared Euclidean distance d is used between candidate feature vectors to produce a unique set of matches $\mathcal{M}$ as so:

$$d = \sqrt{\sum_{i=1}^{N_{F_i}} (\mathcal{F}_a - \mathcal{F}_b)^2} \tag{4.1}$$

$$\mathcal{M}_{ab} = \underset{\mathbf{f}_j \in \mathcal{F}}{\operatorname{argmin}}\, d(\mathcal{F}_a, \mathcal{F}_b) \tag{4.2}$$

With Equation 4.1, a set of putative matches is ascertained and this is demonstrated in Figure 4.6 using "salient" images at a brick interface and feature-poor images ac-

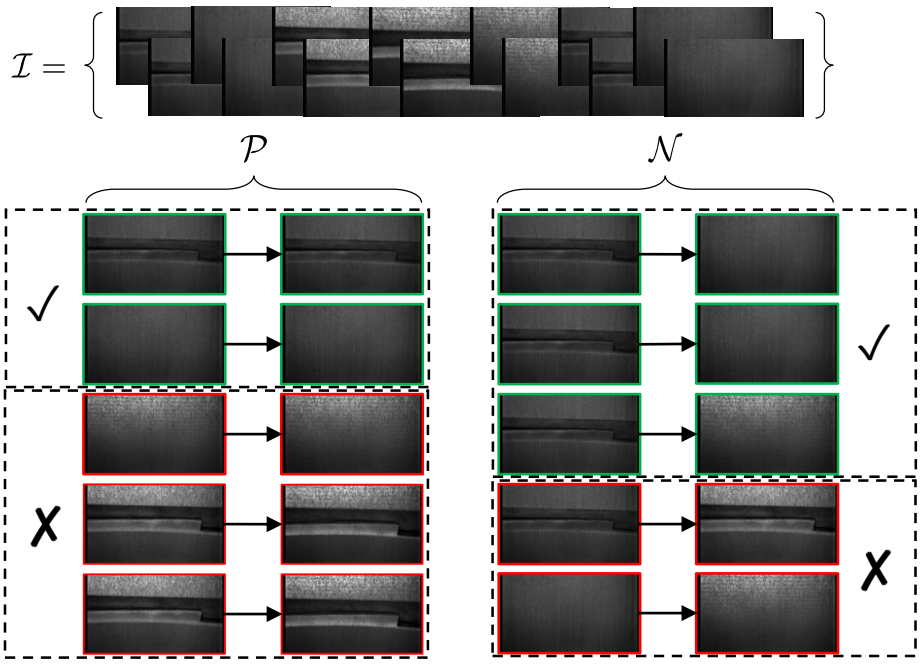


Figure 4.7: Learned feature descriptor approach through positive and negative pair creation through the training process using AGR RVI footage. The input vectors have been abstracted to images for visualisation purposes. Green represents a true-positive pair and red denotes false-positive and false-negative based identification of the image pair.

quired mid-brick. Both image pairs demonstrate a similar vertical translation for parity. As seen in Figure 4.6a), a large consensus of matches are geometrically correct (>75%) meanwhile the feature-poor image pair matching process as seen in Figure 4.6b) demonstrates high levels of ambiguity with only a small portion of matches (∼15%) being geometrically correct. With heterogeneous image pairs, salient objects within the image can be easily matched due to the spatial histogram properties of the SIFT descriptor for the correctly matched key-points being distinct therefore a unique match can be found. Conversely, image pairs which are either texturally uniform or have a high degree of global similarity struggle as a direct result of the descriptor being unable to detect a candidate set of stable key-points in either image which demonstrate saliency, locality and repeatability properties. Therefore, when matching these feature descriptors, the

determined correspondences consist largely of false positives and introduces ambiguity in subsequent stages of any reconstruction. It is important to note that this ambiguity is not exclusive to SIFT features or the use of squared Euclidean distance matching but is symptomatic to even more advanced hand-crafted local feature descriptors such as KDE and DSP-SIFT [151, 152] and state-of-the-art indirect matching procedures such as Gaussian Field Consensus (GFC) and Grid-based Motion Statistics (GMS) [63, 153].

Learned local feature descriptors operate differently to the likes of the ratio test by Lowe [60] where it operates on the disparity of the descriptor space, employing positive matching $\mathcal{P}$ and negative matching descriptor pairs $\mathcal{N}$ as part of a supervised learning approach so that a learned local representation of a feature or object in a scene can be ascertained. One such approach is the margin constraint [61, 154]:

$$d(\mathbf{p}_1, \mathbf{p}_2) + \tau < d(\mathbf{n}_1, \mathbf{n}_2) \forall (\mathbf{p}_1, \mathbf{p}_2) \in \mathcal{P}, (\mathbf{n}_1, \mathbf{n}_2) \in \mathcal{N} \tag{4.3}$$

where $\tau \in \mathbb{R}_{>0}$ is the margin between pairs and the distance metric d utilised is the ℓ_2-$norm$. The system can be trained using a series of input feature vectors in order to effectively determine a discriminative metric or descriptor representation which allows for optimal matching of feature pairs. An abstracted version of this process can be observed in Figure 4.7 where image sub-sets $\mathcal{P}$ and $\mathcal{N}$ pairs are demonstrated. For this particular example, we show how potential false-positive and false-negative pairs can be introduced if the images observe a similar object or feature characteristics, or subject to extreme radiometric changes. Techniques such as hinge-embedding [155] or contrastive loss [156] can be used to enforce a minimum size of margin between unrelated, negative pairs however, environments which result in high degrees of mis-matching and consequently struggle to determine a discriminative metric.

Photometric-based techniques (Direct)

Photometric-based matching techniques operate in a different principle to a feature-based approach. They exploit all the pixel intensity values rather than a candidate sub-set of features in order to derive matches between observed features within the images.

Due to the computational complexity of direct-based methods, direct methodologies are deployed primarily as a post-processing method such as Multi-View Stereo (MVS) [88, 157] to develop densely textured 3-D geometric models of the environment or scene with the underlying sparse point-cloud reconstruction acting as the foundation. However, it must be noted that online direct VO and SLAM methods have been proposed within literature [122,158] which operate on the pixel intensity disparity are not truly "direct" as they often rely on sub-sets of pixel intensities which demonstrate sufficiently large gradients, implying intra-image regions which demonstrate saliency or heterogeneity, consequently resulting in these techniques being more accurately termed as semi-direct.

One of the key approaches towards dense matching between a pair of images is through the utilisation of patch-based descriptors. The most simple of methods rely on a small image patch surrounding a feature to directly represent it then the use of

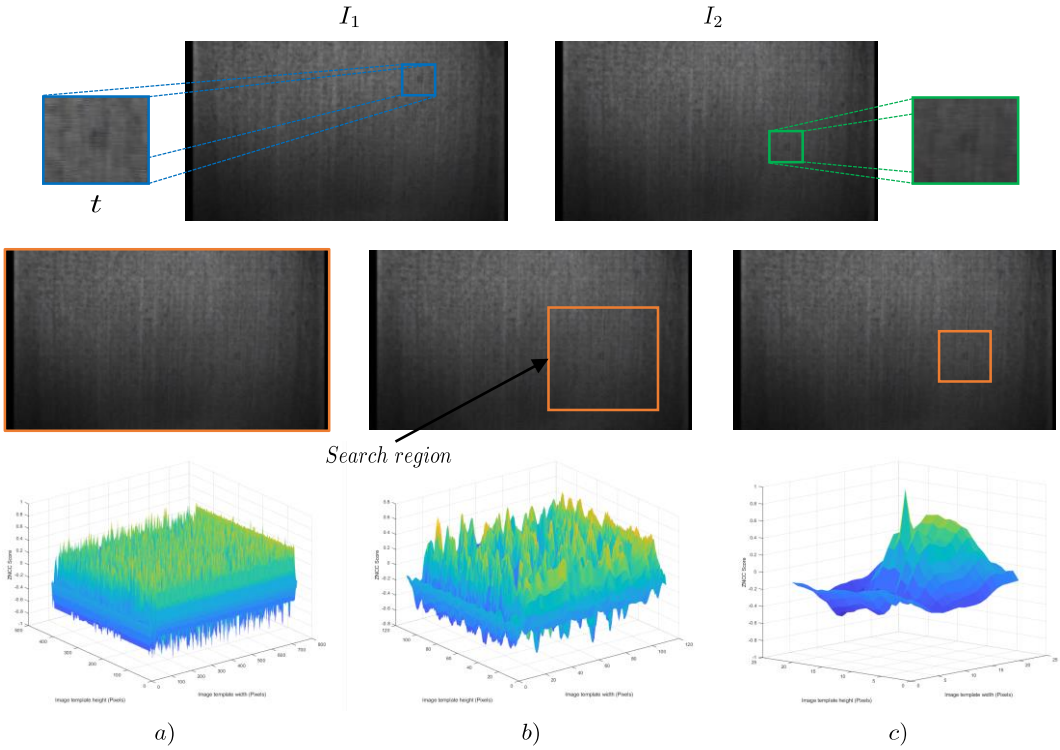


Figure 4.8: Use of ZNCC to match extracted template t from I_1 to baseline shifted I_2. a) Full image search for match; b) Restricted image search based on epipolar constraint; c) Constrained radial search based on dense motion estimation.

SAD, Zero-mean Normalised Cross Correlation (ZNCC) or Mutual Information (MI) to compute a potential feature pairs similarity [159]. More advanced methodologies have been proposed which allows for dense matching approaches such as the DAISY descriptor, PatchMatch or DeepMatching [64, 65, 148]. For the purposes of this thesis, ZNCC and DeepMatching will be used in order to demonstrate the ambiguity which can occur even in dense matching schemes when evaluating challenging environments on a local feature and global image scale.

ZNCC and SAD are regarded within photometric stereo applications to be the best patch-based descriptor however SAD is very sensitive to radiometric change, making it unsuitable for AGR RVI footage [45, 160]. For this example, we extract a single detected SIFT feature location $\mathbf{x}_j$ and apply an 11×11 moving-average filter to suppress illumination changes. ZNCC is mathematically expressed as:

$$ZNCC(u,v) = \frac{\sum_{x,y}\left[I\left(x,y\right) - \bar{I}_{u,v}\right]\left[t\left(x-u,y-v\right) - \bar{t}\right]}{\left\{\sum_{x,y}\left[I\left(x,y\right) - \bar{I}_{u,v}\right]^2 \sum_{x,y}\left[t\left(x-u,y-v\right) - \bar{t}\right]^2\right\}^{0.5}} \in [-1,1] \quad (4.4)$$

where t is the template that will be searched in I and $\bar{I}_{u,v}$ is the mean of surrounding image region of the feature assessed by the template. Three different modes were tested and visualised in Figure 4.8. These were a) a full image search, b) a restricted image search based on the epipolar constraint and c) a constrained radial search performed by using ground-truth motion [45]. As seen in Figure 4.8a), the full image-search demonstrates the intra-image similarity of the mid-brick layer with the existence of multiple maxima and no discernible, unique match. This is even extended to Figure 4.8b) with a region-based constraint by exploiting the epipolar constraint with no unique match. The constrained radial match which is derived from approximate ground truth motion [2] as seen in Figure 4.8c) is the only search method that produces a singular maxima, highlighting the difficulty of feature matching even through direct means. Even by constricting the search region, the computational complexity of applying a moving average filter and ZNCC even whilst exploiting Graphics Processing Unit (GPU) acceleration, takes several seconds for a single feature, making it

unsuitable for a large candidate feature set.

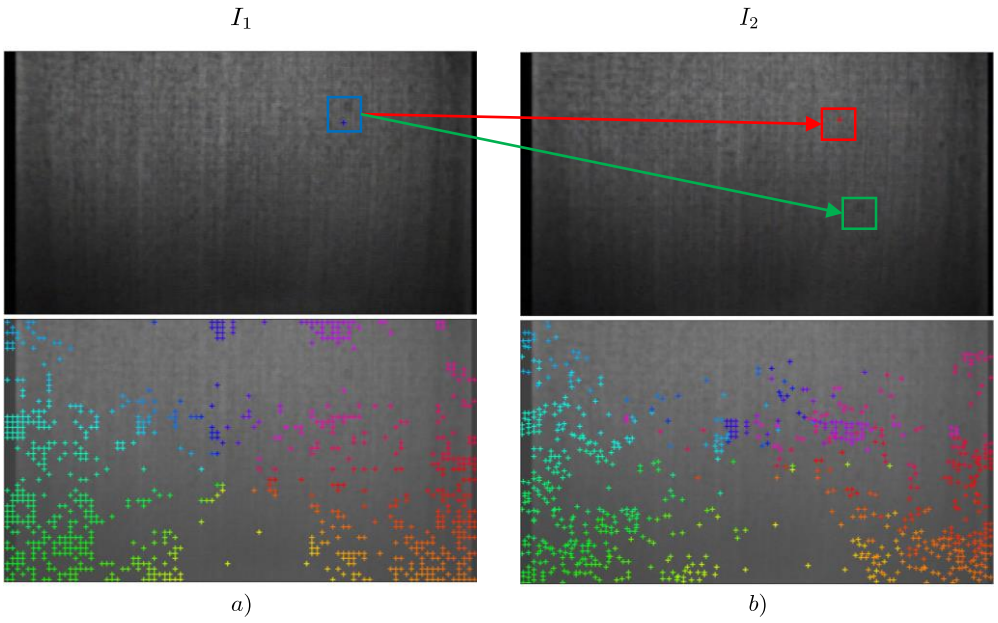


Figure 4.9: Use of DeepMatching [65] between AGR image stills. The top row of images is an highlighted individual match in the region extracted for Figure 4.8. The red cross is the identified match in I_2 from the blue cross in I_1 (boxed with the corresponding colour to improve visibility), with the green box indicating the true match region. The bottom row has the matches associated in colour; Note - The visualisation method used in this figure corresponds directly to results published in [65].

On the other hand, dense correspondence searching techniques have become increasingly common in order to determine a large set of robust feature matches between images, especially within challenging environments. Dense feature descriptors such as DAISY [148] and dense matching schemes [65, 148, 161] can be used not only to determine dense correspondences between images but have underlying applications in optical flow as the dense correspondence field can be used as a method of initialisation for variational energy minimisation schemes [65, 162]. For the purpose of this thesis, DeepMatching has been selected for its self-stated ability of performing well in areas of repetitive texture, relying on a hierarchical, multi-layer correlational architecture which aggregates correlation maps from multi-size image patches to do so [65, 163]. To test the framework, we used the same image pair from Figure 4.8 that was used to

test ZNCC as the input to the DeepMatching algorithm, the result produced can be observed in Figure 4.9.

To evaluate the effectiveness of the matching, since the camera motion is known via the previous example (the camera is translating upwards in a vertical manner), this can then be used as a filter to determine correct correspondences. The results subvert the expected ability of the system, with approximately 10% of the dense correspondences correctly aligning with the ground-truth translation. One particular thing to note about the results produced by DeepMatch in Figure 4.9 is the spatial distribution of the detected matches. Since a large amount of matches have been found at the bottom of I_1, these matches cannot be possibly observed in I_2 as the "corresponding" features are out of view, further highlighting its inability to match accurately within feature-poor or repetitive environments which are the basis for the target application of this thesis.

As discussed in this chapter so far, generalised correspondence searching techniques and the encompassing image-based 3-D reconstruction pipelines are often insufficient at solving the resulting degeneracies when evaluating images that observe feature-poor environments. In this thesis, the particular application of this work is to reproduce 3-D models of AGR fuel channels which contains large sub-sets of the captured imagery that observe repetitive and challenging environments. The following sections propose novel technical contributions in order to alleviate the correspondence searching process in constrained environments, targeting the underlying spatial distribution of detected features to produce robust candidate feature matches.

4.3 Motion-based iterative feature matching

In this section, a novel matching scheme is introduced called the Motion-based Iterative Feature Matching (MIFM) is formulated to deal with the identified problems in Sections 2.4, 3.4 and 4.2, namely deficient feature spaces which challenge conventional feature matching approaches. The approach demonstrates novelty by combining SIFT feature detection and dense optical flow approach Dual TV-L^1, a novel intra-image

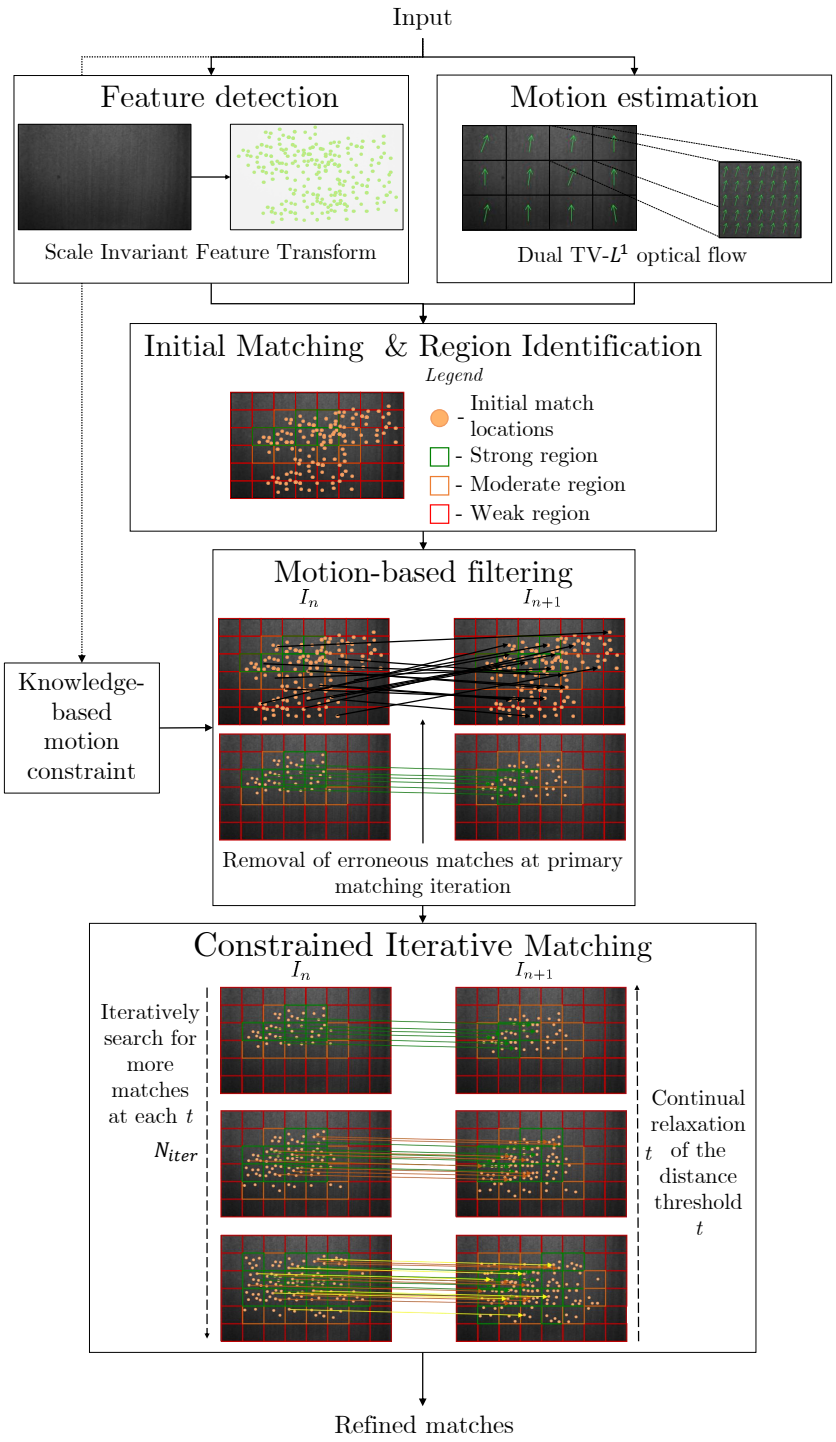


Figure 4.10: High-level overview of the Motion-based Iterative Feature Matching approach

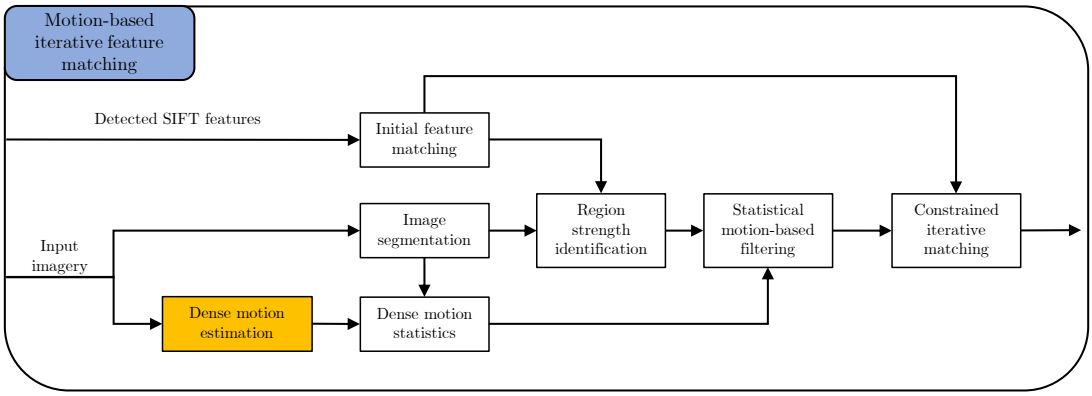


Figure 4.11: Highlighted regions (in orange) of the framework that incorporate motion estimation discussed in Section 4.3.1

regional strength determination mechanism to suppress regions that produce unstable matches and utilises a constrained iterative matching scheme to improve match density which is guided by the stronger regions of the image. A visual high-level overview of the approach is provided in Figures 4.1 and 4.10, code-based implementation of the approach explained in Appendix A.3 and the following sub-sections detail the novel functionality of the MIFM approach.

4.3.1 Feature detection and motion estimation

The first step of the algorithm is to process all the images using the SIFT algorithm, the commonly used benchmark algorithm, into a set of features which represent identified DoG scale-space maxima and their surrounding neighbourhood in the form of a descriptor [60]. The motivation for choosing this specific method is due to its excellent radiometric and geometric invariant properties and the performance attained in descriptor surveys by Miksik et al [138] and Schonberger et al [61] for its performance for image-based 3-D reconstruction.

Concurrently, a direct approach is taken by determining the optical flow between pairwise images which allows the estimated displacement for every pixel value within the image. To perform optical flow, the Wedel et al implementation of the Dual Total-Variation L^1-*norm* (Dual TV-L^1) algorithm [113] was selected as it takes a different

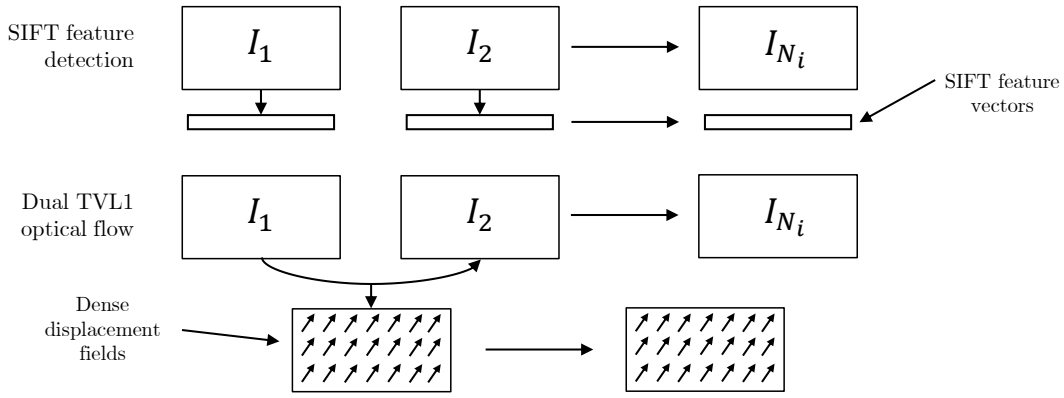


Figure 4.12: Use of both the indirect and direct approaches to produce descriptor feature vectors and a dense displacement field between assessed imagery

approach from many direct approaches with regards to non-uniform illumination which operate on image gradients [164] and instead uses a structure-texture decomposition model which retains fine image details, and denoises the image under the assumption of oscillatory noise being present [165] so that:

$$\min_{I_S} \int_{\Omega} \left\{ |\bigtriangledown I_S| + \frac{1}{2\theta} \left(I_S - I \right)^2 \right\} dx \tag{4.5}$$

$$I_T \left(\alpha, x \right) = I \left(x \right) - \alpha I_S \left(x \right) \tag{4.6}$$

where I_S and $I_T \left(\alpha, x \right)$ are the decomposed structure and texture elements respectively and α is the blending factor which is introduced as I_T is more vulnerable to sensor noise and sampling artifacts [113]. Unlike other dense optical flow methods such as Horn-Schunck (HS) [164], the Total Variational (TV) approach [165] allows for outliers to be preserved whereas the HS method tends to oversmooth the motion field and obscure regional information. To test the effectiveness of Dual TV-L^1 optical flow, it was evaluated on a small sub-set of AGR RVI footage. Due to the lack of ground-truth motion, a 2-D panoramic image produced using ASIST software [2] is used to determine the approximate misalignment between the start and stop image. In the evaluated area where Dual TV-L^1 was applied, there is a vertical shift of roughly 2600

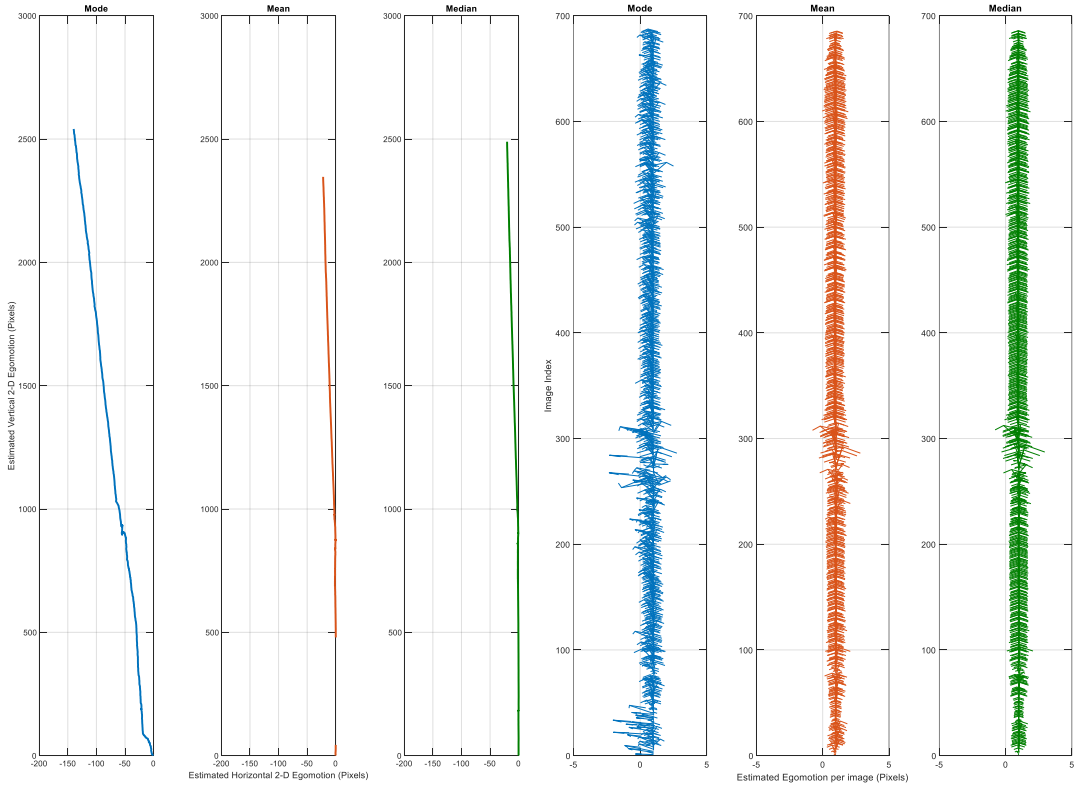


Figure 4.13: Dual TV-L^1 algorithm applied to small subset of footage with camera moving up the channel. a) Cumulative estimated 2-D modal, mean and median egomotion of the camera; b) Estimated modal, mean and median motion on an individual image basis as a quiver plot.

pixels and a horizontal shift of approximately -30 pixels relative to the starting frame. To phase out potential outliers in the displacement field, the arithmetic mean $\bar{w}_I$, median $\tilde{w}_I$ and mode $\texttt{Mo}(w_I)$ of the estimated flow field $w = (u, v)$ is determined. The cumulative modal, mean and median egomotion ($C_i = \sum_{i=1}^{n} \texttt{Mo}(w_{I_i})$, $C_i = \sum_{i=1}^{n} \bar{w}_{I_i}$ and $C_i = \sum_{i=1}^{n} \tilde{w}_{I_i}$) is shown in Figure 4.13a) and $\bar{w}_I$, $\tilde{w}_I$ and $\texttt{Mo}(w_I)$ of each image is shown as a quiver plot in Figure 4.13b).

As demonstrated in a similar example in [45], one of the papers published from this work, the mean cumulative horizontal displacement and the modal vertical displacement ascertained from the derived motion are the most accurate, with 22 pixels and

2541 pixels of horizontal and vertical displacement respectively. Furthermore, Figure 4.13b) is an excellent demonstration of the ego-motion of the camera at each image and accurately depicts areas of irregular motion which could be particularly problematic if pure knowledge constraints are used. After all the images have been processed into a set of features and displacement fields using SIFT and Dual TV-L^1 respectively, the next stage in the pipeline is to derive matches using this information.

4.3.2 Initial image matching and regional strength identification

Once the initial feature detections and the motion displacement fields have been derived, the next steps are to perform an initial matching stage, discretise the image into a set of regions using segmentation and extract regional motion statistics pertaining to each image region (see Figure 4.14). Each of these stages is now described in detail.

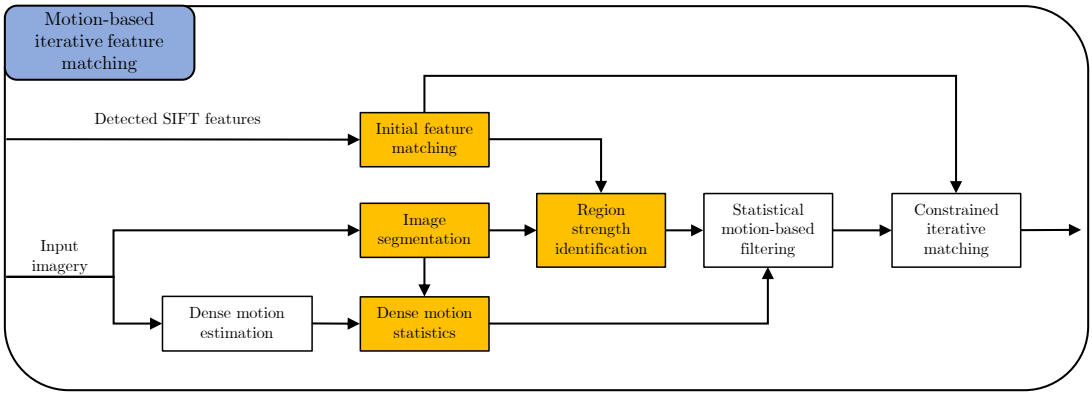


Figure 4.14: Highlighted regions (in orange) of the framework that incorporate the initial matching, image segmentation and regional motion statistics extraction discussed in Section 4.3.2

Initial feature matching

The first stage of this technique is to perform an initial bi-directional feature matching procedure using the squared Euclidean distance shown in Equation 4.1 which aims to deriving a unique match for each feature point, adhering the uniqueness constraint (i.e. 1-to-1 matches) for the purposes of geometric verification. Lowes ratio [60] is used to

effectively determine the unique matches using a threshold t which ensures that the derived match maintains a distance from its second strongest match. In the initial matching stage, t is set high at a value of 2.0. If the number of detected $|N_{matches}|$ exceeds a hard coded high threshold of required matches, then the rest of the process is simply ignored. This is indicative of a brick interface region, or a general feature-rich constrained environment where no further refinement is required. If applying this to challenging imagery however, several unwanted use cases can occur as shown in Figure 4.15.

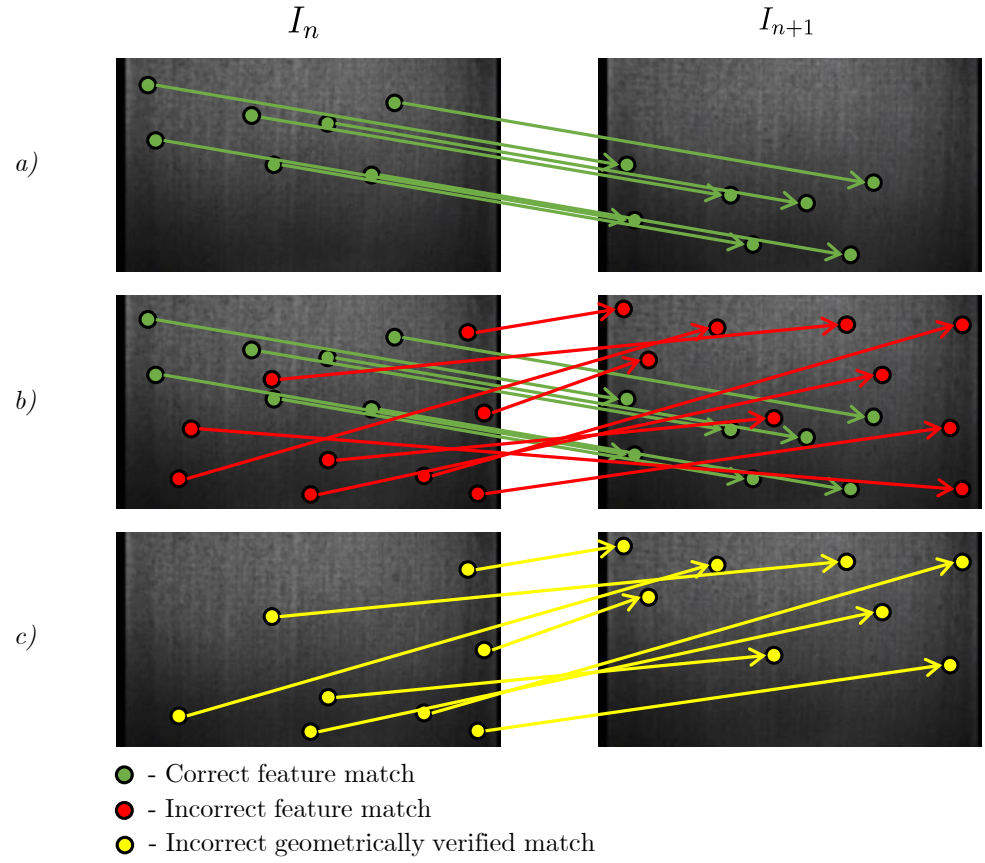


Figure 4.15: Possible issues in the initial matching stage. a) Insufficient (yet correct) matches to estimate essential matrix ($N_{matches} \geq 8$); b) Sufficient matches however there is many incorrect matches. This is the most common problem with matching using AGR footage as seen in Figure 4.6b); c) Direct result of geometric verification of b), as incorrect matches outnumber correct matches resulting in incorrectly, yet geometrically verified consensus of matches.

Insufficent matches The matches determined are correct however there are too few for determination of the Fundamental matrix which requires at least 8 feature matches.

Incorrect matches The resultant matches contains minimal number of true positive matches with a large number of mismatches. This is indicative of an image which contains textural uniformity or repetitive structures.

Incorrect consensus The incorrect matches overwhelm numerically the correct matches resulting in an false consensus and subsequently, an erroneous Fundamental matrix.

The matching procedure has 4 different potential states that can be enacted after this initial matching procedure:

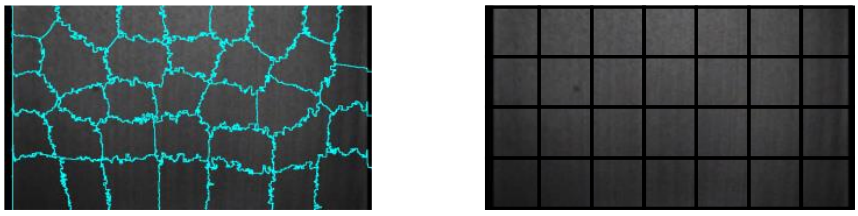
State 1 The number of matches is insufficient and DMS being enabled within the system means that the dense optical flow is combined with the feature matching procedure to identify regions of strength based of feature match density and number of correct matches. Matches are then filtered based on the derived motion and iteratively lower the threshold whilst constraining the matching procedure using the derived motion from strong regions.

State 2 The second state operates very differently and is only entered if knowledge-based constraints about the motion between frames is enabled, and performs threshold relaxation using the knowledge-based constraints.

State 3 The third state is entered if no knowledge-based constraints are provided, and the dense motion estimation is disabled. This procedure derives region strength from the match density and the harmonic mean of the Euclidean distance for each feature match within the region instead, and then performs relaxation of the threshold.

State 4 The final state is reserved for when the initial matching phase was sufficient and there is no requirement to obtain further matches.

Image segmentation



a) *b)*

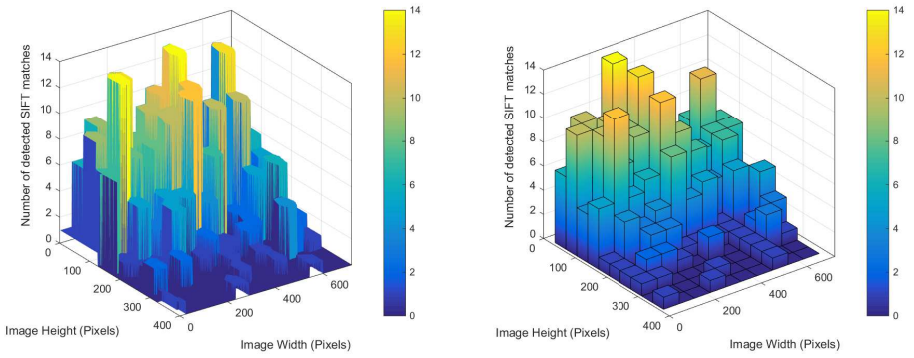
Figure 4.16: Segmentation of the image using a) SLIC superpixels [134]; b) Spatial discretisation of image into grid-based bins.

Segmentation is performed by quantising the image into a number of bins $\mathcal{K}$ using simple spatial bucketing [14] or through the use of Simple Linear Iterative Clustering (SLIC). Spatial bucketing is performed by breaking the image up using a grid into a number of cells that are of equal size, whereas SLIC is a gradient-ascent-based algorithm which discretises the image into regions called a "superpixel" based on colour image gradients, where the grid interval is $S = \sqrt{\frac{N_p}{\mathcal{K}}}$ where N_p is the number of pixels in the image and $\mathcal{K}$ refers to the number of initial superpixel cluster centres [134]. The segmentation process of choice can be determined based on user choice. Both these methods are visualised in Figure 4.16a)[4] and Figure 4.16b) respectively.

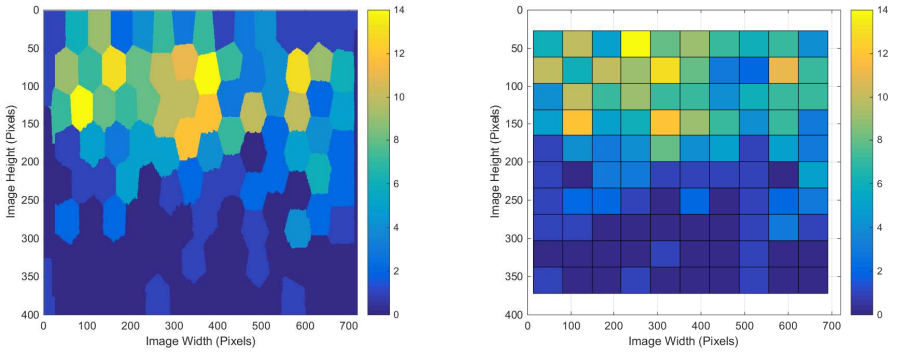
Regional strength identification

To derive the regional strength of the image initially, the feature match locations are spatially binned into their associated grid region or superpixel cluster $\mathcal{K}$. If the matching system is in state 1, the regional strength is derived via the density of the initial matching alone - an example of the density-binning can be observed in Figure 4.17. If the matching system is in state 2, translation-based geometric constraints established using *a-priori* knowledge about the rigid body motion is used to disambiguate candidate matches between image pairs. For example, in the AGR in-core inspection footage,

[4]The implementation in the figure generates $\mathcal{K} = 25$ superpixels. More super-pixels can be generated however this is unnecessary.



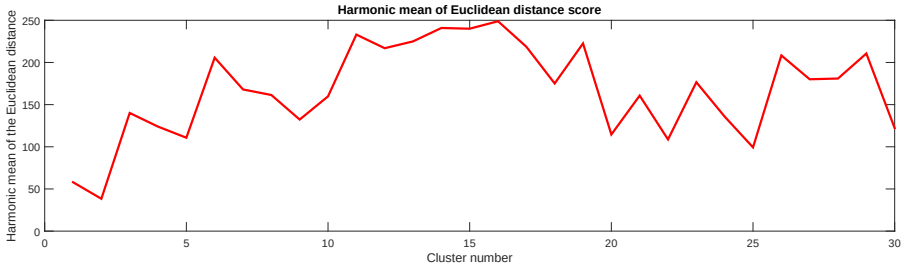
(a)



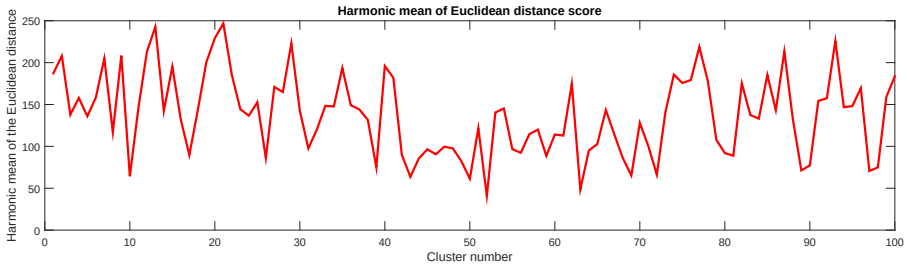
(b)

Figure 4.17: Region strength identification using initial SIFT feature match locations (Left column - SLIC superpixel segmentation, Right column - Spatial segmentation). a) Bivariate histogram of both methods of segmentation b) 2-D histogram of both methods of segmentation.

the NICIE/CBIU inspection camera moves approximately vertically ±5 pixels between each image decomposed at the camera frame rate of 25 frames per second (FPS) [2]. Therefore, a simple translation filter is implemented to ensure the feature locations $\mathbf{x}_j$ for each identified match fall within a boundary constrained by the horizontal (α) and vertical motion (β) so that for each corresponding point, $\exists!\ \mathbf{x}_j = [x \pm \alpha, y \pm \beta]$ after which the resulting match density is used to determine regional strength. If the matching system is within state 3, the match density or the harmonic mean of the Euclidean distance is incorporated in order to determine the strengths of the image regions. This is expressed using the following equation:



(a)



(b)

Figure 4.18: Harmonic mean of the Euclidean distance scores obtained from the SIFT feature matches for each cluster. a) Representative of a feature-poor or challenging image; b) Representative of a feature-rich image. The lower $H_{\mathcal{K}_i}$ is, the better.

$$H_{\mathcal{K}_i} = \frac{N_{\mathcal{K}_{matches}}}{\sum_{i=1}^{N_{\mathcal{K}_{matches}}} \frac{1}{d_i}} \tag{4.7}$$

By utilising the harmonic mean of each clusters Euclidean distance, it biases towards to the closer distances which gives an indication to the strength of each region. An example of the Harmonic mean being calculated across each assessed cluster is shown in Figure 4.18. The regions are then sorted into categorisations of strength based on hard thresholds $\{\mathcal{M}_{strong}, \mathcal{M}_{moderate}, \mathcal{M}_{weak}\} \subset \mathcal{M}$ on the match density or the Harmonic mean metric values, where $\mathcal{M}_{strong}$ pertains to region of high density or low $H_{\mathcal{K}_i}$ scores, $\mathcal{M}_{moderate}$ is indicative of moderate match density and $\mathcal{M}_{weak}$ regions are areas with little to no matches and high $H_{\mathcal{K}_i}$ scores.

Dense Motion Statistics

the next stage is to extract descriptive statistics from the derived displacement fields produced by the Dual TV-L^1 in regard to every region of the image. As previously

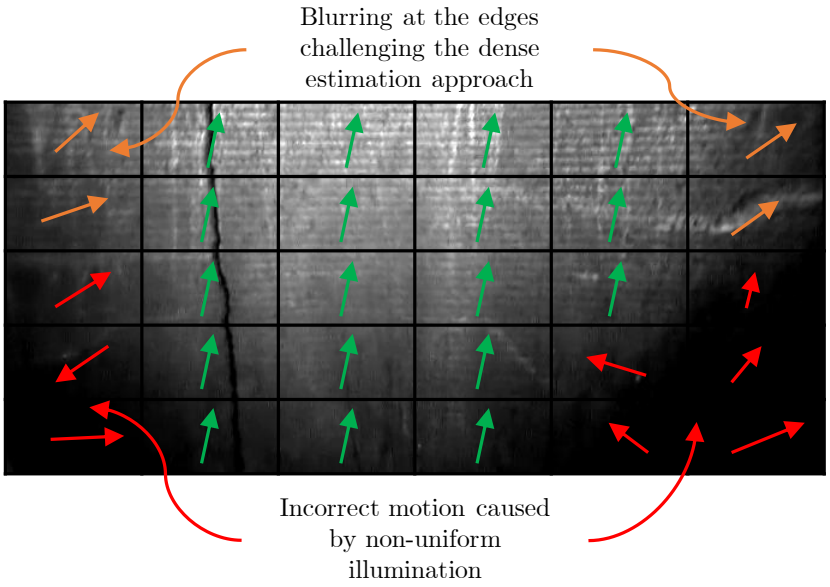


Figure 4.19: Synthetic example portraying the effect of non-uniform illumination or lens distortion can affect the derived optical flow between the assessed image pair

described in Section 4.3.1, the mean of horizontal displacement and the mode of vertical displacement is applied to every determined region of the image. The reason for this is that the motion displacement field calculated is highly likely to be inaccurate in areas with non-illumination or lens distortion as the underlying image values are being distorted by the hardware, an example of which is shown in Figure 4.19. In the subsequent stage, the region strength derived from the feature matching can be used in combination to extract motion regions in areas of saliency to improve the global image matching accuracy.

4.3.3 Motion-based filtering and constrained iterative matching

Statistical Motion-based filtering

With regional strengths ascertained, the next stage is to extract the DMS pertaining to the strong regions identified at the feature matching stage. This is performed by applying a binary mask R that can be applied to the displacement field w expressed below:

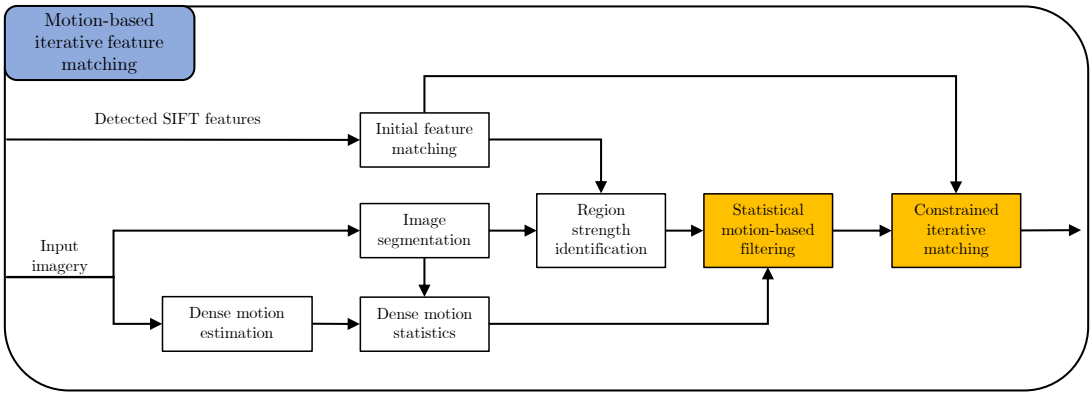
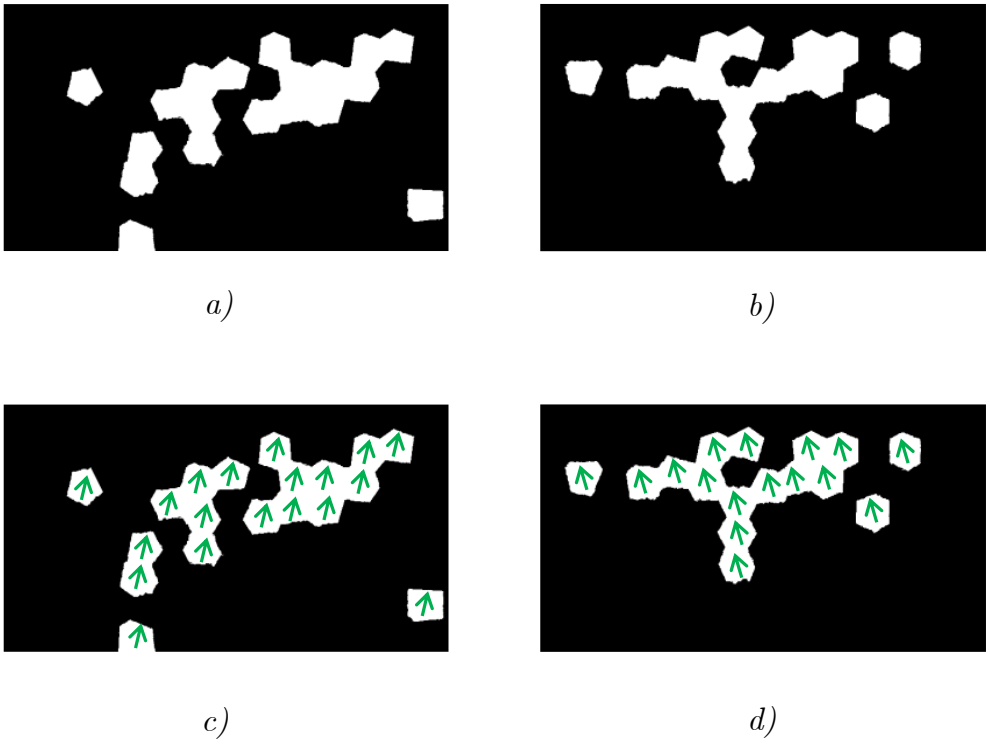


Figure 4.20: Highlighted regions (in orange) of the framework that incorporate the motion-based filtering and constrained iterative matching scheme, image segmentation and regional motion statistics discussed in Section 4.3.3

$$w_{acc} = (R \circ w) = (R)_{i,j}(w)_{i,j} \tag{4.8}$$

where $\circ$ is the Hadamard product and the w_{acc} is the resultant motion displacement field corresponding to heterogeneous clusters or spatial regions within the image. The horizontal and vertical displacement values of each region is captured and this motion value acts as a 2-D geometric constraint similar to the aforementioned translation filter. The matching regions are then re-quantified by first determining the regions with the highest and lowest match density. This is performed by identifying the k largest and smallest elements of $\mathcal{M}$ where $k = \frac{r|\mathcal{K}|}{100}$ and r is the retention factor with regards to the upper and lower percentile of matches whose region density that is evaluated respectively. With the geometric constraint or the translation filter applied, the regions which have the k highest density of rejected matched is also determined. The strong region is determined as the complement between the highest accepted and rejected match regions , the moderate being the intersection between the highest accepted and rejected match regions and the weak regions being the k regions with the lowest number of matches in parent set $\mathcal{M}$. $\mathcal{M}_{strong}$ and $\mathcal{M}_{moderate}$ are defined below as:

$$\mathcal{M}_{strong} = \mathcal{M}_{acc} \setminus \mathcal{M}_{rej} \tag{4.9}$$



a) *b)*

c) *d)*

Figure 4.21: Binary mask generation to determine salient SLIC clusters. a/b) Original generated binary masks pertaining to $\mathcal{M}_{strong}$ regions. c/d) Visualisation of the Dual TV-L^1 motion estimation contained within these regions.

$$\mathcal{M}_{moderate} = \mathcal{M}_{acc} \cap \mathcal{M}_{rej} \tag{4.10}$$

Once region strength has been determined, each region definition is expressed as a binary mask R that can be applied to the displacement field w. With each remaining pixel p containing a displacement vector (u, v), the mean of the remaining displacement field w_{acc} horizontal and vertical components are determined, of which the maximum components derived from the descriptive statistical measures are used as the estimated motion between the baseline-shifted images. With the estimated motion determined, a secondary filter (primary if no translation filter has been applied) is implemented with the estimated motion to further restrict the matches.

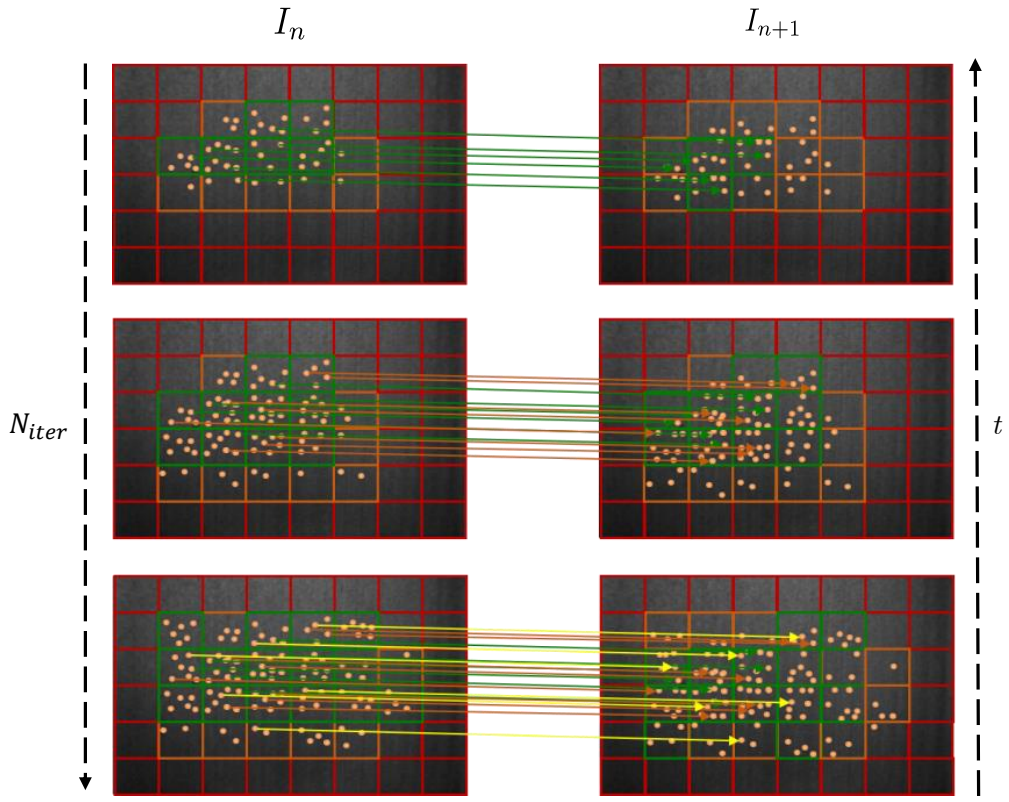


Figure 4.22: Visualisation of the constrained iterative matching approach using spatial binning. Demonstrates the relaxation of the matching threshold and with this combined with motion-filtering, can produce more matches even with challenging images. Green, orange and red cells indicate a strong, moderate and weak region respectively.

4.3.4 Constrained iterative matching

The next stage is to perform a constrained iterative process of matching SIFT features until $|N_{matches}|$ is sufficiently large, where the constraint in place is either the knowledge-based constraint or the derived motion constraint w_{acc}. This is done by continually relaxing the match threshold t by a step size μ whilst applying the constraint to filter the matches. If the number of matches is insufficient, then the procedure is repeated with a decremented threshold value until a pre-specified maximum number of iterations N_{iter} or until a minimum number of matches have been attained. This pro-

cess is shown in Figure 4.22, where it undergoes evaluation at three different thresholds to obtain enough matches to be stored.

One of the fundamental problems when the matching criterion required for matching feature descriptors is relaxed, the number of matches increase but at the expense of accuracy. By implementing the estimated motion derived from the salient regions of the image into the matching process, the outcome is a set of pruned matches that adhere to the acquired, quasi-dense motion (or even including the matches obtained from the initial translation filter criterion) and the removal of ambiguous matches that could produce an incorrect consensus as seen in Figure 4.15.

4.4 Discussion

This chapter provides an in-depth look at the problem of deficient feature spaces within not only 3-D reconstruction but general image processing. The two core issues examined are feature homogeneity and feature ambiguity, which look at the representation of features and the matching of these representations. To contextualise the challenges of processing homogeneous features, two scenarios are presented; a scene that contains observable, repetitive structures and a scene which contains repetitive, indiscriminate feature structures. These two examples represent issues that can arise in challenging datasets, either on the whole, or in small sub-sets within a larger dataset which prove extremely challenging for generic matching schemes. This proves a significant challenge to applying image-based 3-D reconstruction to asset inspection where the industrial asset itself is a consistent or repetitive structure [12, 134], or within urban environments which contain repetitive facades [166]. Most importantly, as highlighted by Chapter 3.4, typical SfM frameworks which incorporate generic feature matching when applied to AGR RVI imagery results in a misrepresentative or very sparse point cloud. In order to alleviate this, most SfM frameworks rely on the Multi-view Stereo (MVS) approach as a form of post-processing to rectify gaps within the sparse point-cloud through the use of dense correspondence. However, if the SfM output contains significant error at the matching stage due to the challenging imagery, then the application of MVS

will simply compound the error of the SfM output, resulting a dense but misrepresentative reconstruction. Therefore, it is important to bring the focus onto deriving accurate correspondences first in order to complement the subsequent sparse and dense reconstruction stages.

To provide robustness when evaluating deficient feature spaces, a novel matching scheme that revolves around the combination of knowledge constraints, motion estimation, segmentation and indirect feature matching is proposed. The method operates by first detecting keypoints within the image using SIFT [60] to produce feature vectors with rotation, scale and illumination-invariant properties. In conjunction with this, dense optical flow using the Dual TV-l1 approach is performed to determine the approximate motion between frames. Once complete, an initial pairwise match is obtained using the squared Euclidean distance between the feature vectors to determine potential matches between the images. If the number of matches is poor, or the consensus in the matching procedure is inconsistent, a region strength identification is utilised to refine the matches. To perform this, the image is discretised into regions, either via spatial bucketing [46] or segmented using the SLIC algorithm [134], and a bivarate histogram is obtained to determine the density of each region with regards to SIFT feature locations and their associated distance between matches. This allows for strong and weakly supported regions of the evaluated images to be determined, therefore allowing for matches to be effectively pruned by either the densely estimated motion or the knowledge-based motion constraint based on the strength of the region. Consequently, this method often reduces the number of matches significantly and a threshold relaxation strategy is performed to introduce more matches whilst adhering to the region strength and determined motion - providing a refined set of matches. This approach allows for a strong set of candidate matches to be obtained for the subsequent reconstruction procedure.

The proposed MIFM approach has several advantages and disadvantages:

+ The proposed matching method can identify regions within the assessed images that produce erroneous matches. This allows for the algorithm to suppress unstable matches produced by underlying issues with the imagery such as specularities

(reflections), non-uniform illumination, lens distortion or camera sensor noise.

+ MIFM can be used to generate robust and accurate matches of feature-poor and repetitive imagery by exploiting small baseline shifts between the assessed imagery.

+ The method utilised should be globally applicable across multiple camera models such as pinhole, omnidirectional, fisheye and more.

− The MIFM algorithm is computationally expensive in its current form of implementation, taking around 1-2 seconds per image frame. However, this could be drastically improved by using GPU implementations of both the DTVL1 [113] and SIFT [25] algorithms, in addition to paralleling the regional strength identification algorithm.

− Ordered, contiguously captured imagery is required as a source input for the algorithm to operate using the motion estimation component.

In the following Chapters, the implementation of this method within MATLAB is demonstrated in Chapter 5 and Appendix A.3, in addition to evaluation of this approach using real-world AGR RVI footage within Chapter 6. To evaluate the effectiveness of the proposed matching scheme with challenging, feature poor environments, several input data-sets gathered from different AGR reactors and inspection tools are utilised. The reactor RVI footage employed is captured by three different inspection tools; CBIU, NICIE2 and the Core Restraint Viewing Winch (CRVW) footage which all produce variable quality. Additionally, laboratory footage gathered from a bespoke experimental apparatus which emulates a small sub-section of the AGR fuel channel will benchmark the effectiveness of the MIFM approach, in addition to evaluate the resultant reconstruction quality from the produced MIFM matches.

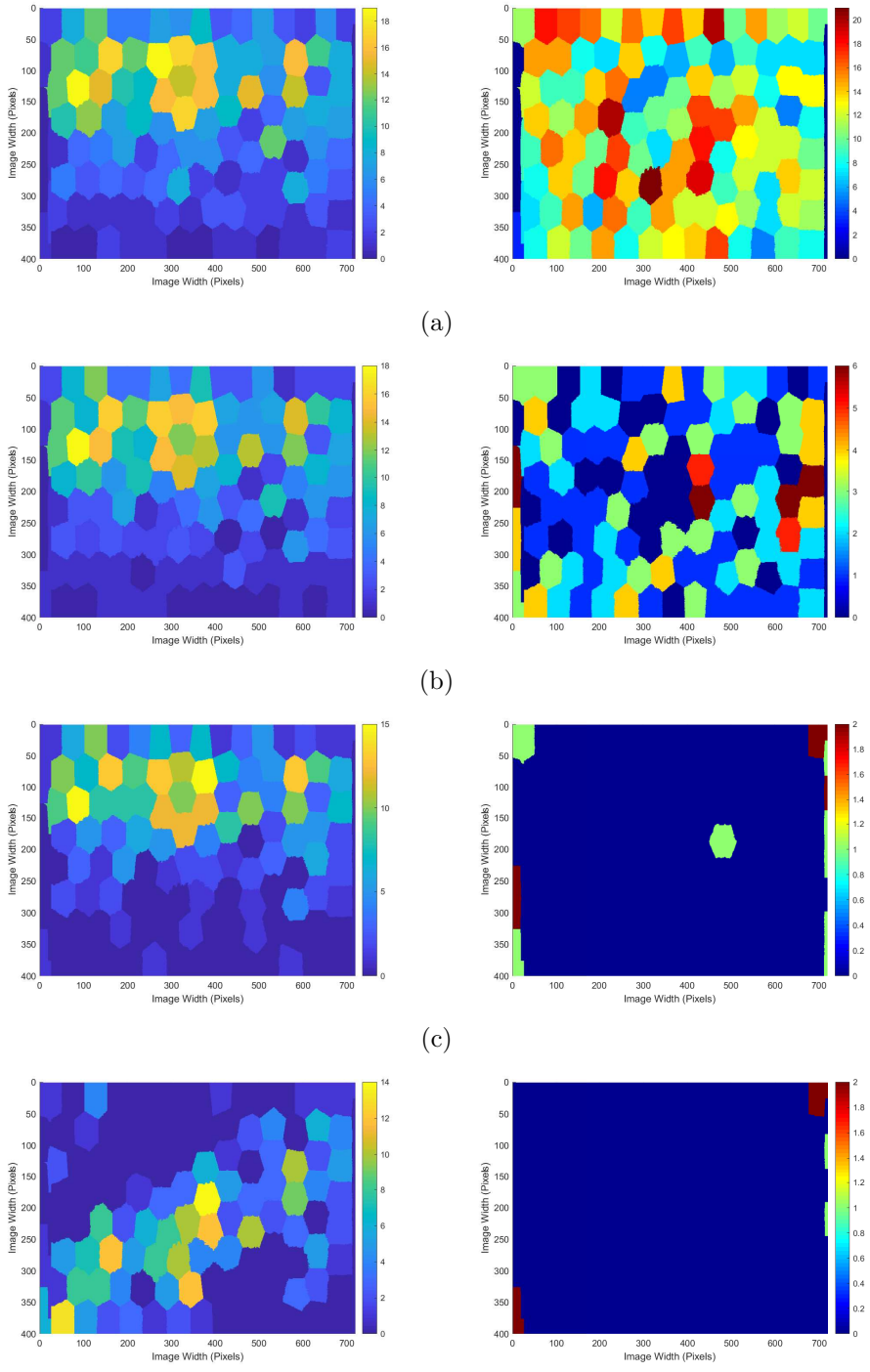


Figure 4.23: Region-based accepted (Left column) and rejected (Right column) match density at various principal matching thresholds using SLIC image segmentation. a) $t = 0.8$, b) $t = 1.2$, c) $t = 1.6$ and d) $t = 2.0$. Note the increased threshold leads to reduced accepted and rejected regional match density.

Chapter 5

Constrained Homogeneous Structure-from-Motion framework (CH-SfM)

Summary of chapter contributions

The chapter proposes the novel Constrained Homogeneous Structure-from-Motion (CH-SfM) framework, an incremental image-based 3-D reconstruction framework. The framework incorporates the novel feature matching scheme MIFM in Chapter 4, in addition to removing features available on most incremental SfM frameworks such as Retriangulation, with the aim of producing robust feature matches at the expense of density. Validation is performed using synthetic ground-truth data in order to ensure that the reconstruction framework operates as intended. To determine the capability of the CH-SfM framework, it is assessed using 3 distinct inputs which are categorised based on the afflicted noise; no noise, minor noise and considerable noise error. This assessment indicated the frameworks capability to recover original structure and estimated camera locations. However, it will also demonstrate the frameworks inability to operate when considerable noise sources are introduced. The CH-SfM framework also incorporates novel capabilities in being able to produce circumferential reconstructions.

5.1 Introduction

This chapter introduces the implementation of a image-based 3-D reconstruction framework that was designed to operate using contiguously captured images to produce a

representative geometrical structure of the observed scene. The designed framework is named the Constrained Homogeneous Structure-from-Motion (CH-SfM) framework as it has been designed to produce 3-D reconstructions of challenging and constrained environments where the captured image quality or the textural content of the scene is objectively deficient. The framework structure was developed primarily for the reconstruction of AGR RVI imagery and can be viewed in Figure 5.1. In Chapter 4, the introduction of the motion-based iterative feature matching approach is explained in great detail and the strength of this method is evaluated in Chapter 6 with respect to the methods robustness to challenging imagery. Within this chapter, the focus is instead of the incremental reconstruction element of the framework to ensure that reconstructions produced by the framework can be deemed representative of the target scene for general applications aswell as constrained imagery which is the main focus of the framework. In this chapter, the aim is to validate the reconstruction accuracy of the framework using a synthetically generated data source, with details on the operation of the framework and the key parameters provided in the associated Appendix A.

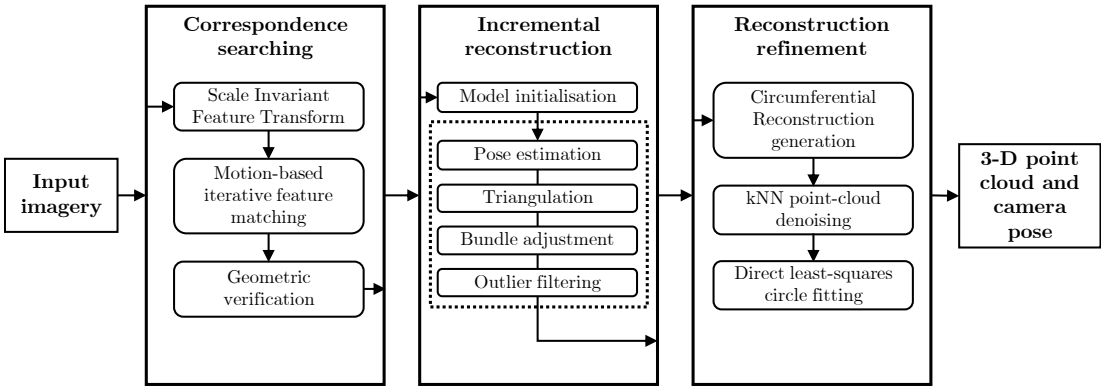


Figure 5.1: Structural overview of the CH-SfM framework

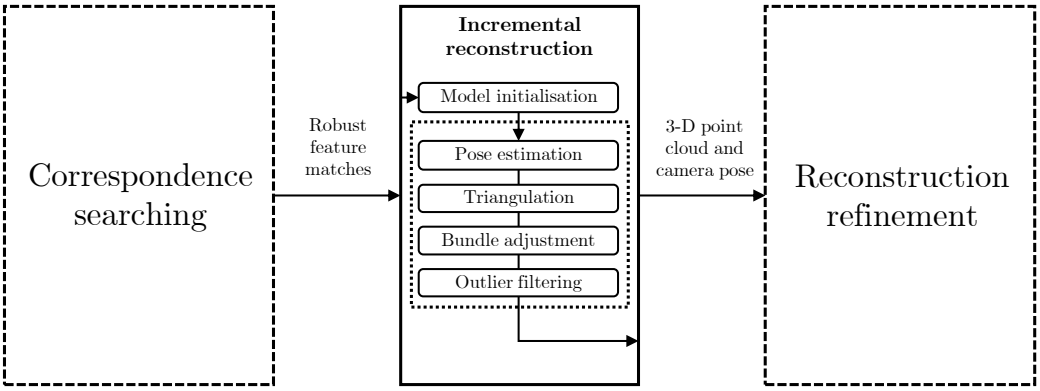


Figure 5.2: Flowchart indicating the part of the framework under validation

5.2 Reconstruction validation

5.2.1 Testing methodology

In order to evaluate the framework, the aim is to determine the effectiveness of the incremental reconstruction element of the framework. As previously mentioned, a synthetic scene is generated where a random camera path is generated, with the rotational and translational components of each respective camera being randomised to generate a path in which the camera travels. Alongside this, a 3-D point cloud is generated, with randomised locations constrained to ensure that there is a minimum number of features observed by the camera to ensure no issues with Perspective-n-Point (PnP) algorithms [69] and a maximum number of feature points to reduce redundant computation and to simulate a constrained environment. The key parameters used to generate the synthetic models are described in Table 5.1 below:

Parameter	**Parameter value**	**Parameter description**
$N_{Cframes}$	50	Number of camera poses registered for the model
$N_{featmin}$	10	Minimum number of features observed by each camera
$N_{featmax}$	50	Maximum number of features observed by each camera

Table 5.1: Parameters used to generate the synthetic model

Due to the scene being synthetic in nature, a synthetic camera calibration matrix $\mathbf{K}$ must be generated in order to not only compute 3-D locations of the camera and the 3-D points, but also ascertain the reprojection error E_{reproj} of each reconstruction. The camera calibration matrix is defined as below:

$$\mathbf{K} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{5.1}$$

The camera calibration matrix comprises of several entities; the focal length f (also known as the scale factor), s is the skew between the sensor and optical axes and the principal point $\mathbf{p} = [c_x, c_y]$ which denotes the optical centre of the camera expressed in pixel co-ordinates. To generate $\mathbf{p}$, an arbitrary image sensor with a resolution of 500×500 is used, indicating that $\mathbf{p} = [250, 250]$. The focal length of the camera is also determined, assuming square pixels as $[f_x, f_y]$, as $\frac{1}{500}$ for each focal length value respectively.

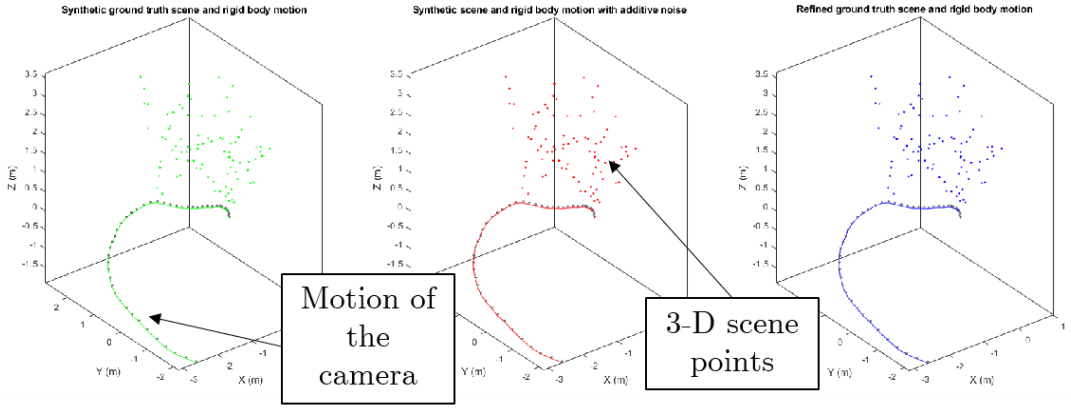
To determine the effectiveness of the reconstruction, noise must be introduced to the synthetic scene to make it a realistic representation of a real-world environment. The testing is split into three phases - no noise, minor additive noise and considerable additive normally distributed random noise using the `randn()` `MATLAB` function to all the main parameters such as the calibration matrix, rigid body motion (Rotation and translation) in addition to noise added to the detected 2-D correspondences and the associated triangulated 3-D scene point locations. To determine the error induced into the camera calibration, the skew parameter s is set to zero and $\mathbf{K}$ is changed to a 4-vector comprising of the key parameters $\mathbf{K} = [f_x, f_y, c_x, c_y]^T$, and the percentage difference between the ground-truth camera calibration matrix and the noise-based camera calibration matrix is determined. To determine the rotational E_{rot} and translational error E_{trans}, the $\mathbf{R}$ axis/angle representation is converted to a 3×3 rotation matrix using Rodrigues formula [59], and then the error is calculated based on the approach described in [167], which is used for the KITTI vision benchmark suite. The resultant rotational error is expressed as "*deg/m*" and the translation error is the mean

Euclidean distance error expressed as a percentage as shown in [167]. In Table 5.2, the error induced to every component is quantified:

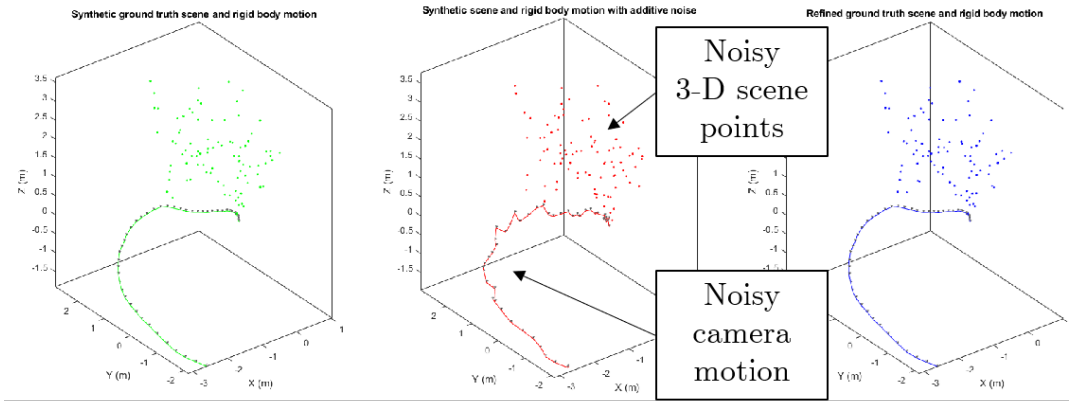
Parameter	Minor noise error	Considerable noise error	Parameter description
K	0.2%	2.1%	Camera calibration matrix
R	0.0676 deg/m		Rotational component of the camera poses
t	3.9%	8.4%	Translational component of the camera poses
X	0.1%	9.95%	Triangulated 3-D point locations
x	0.3%	1.1%	Derived 2-D feature correspondence locations

Table 5.2: Synthetically induced error introduced to the models for both the minor and considerable noise test cases.

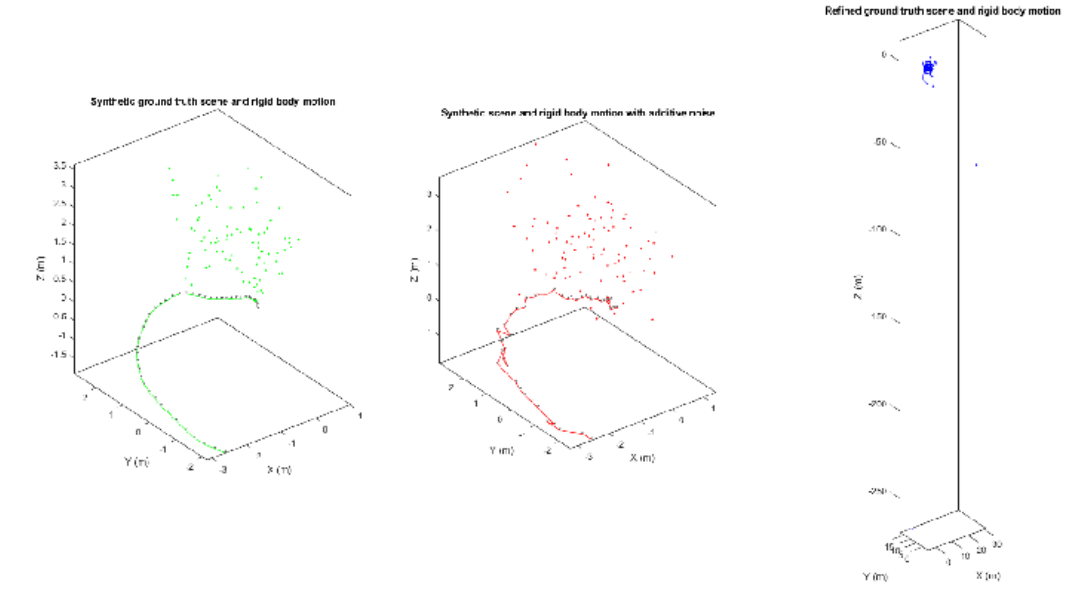
The minor error case is intended to simulate a well calibrated system with minimal error induced in the camera calibration process, and with derived 3-D points and the associated camera rigid body motion being relatively accurate. The considerable noise case is where the camera calibration procedure induces a larger error into the process, with the derived rigid body motion and the associated 3-D reconstruction points suffering as a result. Due to the constrained environment in which the system operates, the rotational error of both noise cases is kept consistent, and the 2-D feature correspondence locations also differ to a small degree to simulate a challenging feature space. The results of the reconstruction element of the system is presented with regards to the refined point-cloud reconstructions presented in Figure 5.3, the rotational errors in Figures 5.4 - 5.7 and the translational error in Figures 5.8 - 5.11.



(a)



(b)



(c)

Figure 5.3: Synthetic, noisy and refined point clouds with a) no noise; b) minor noise; c) considerable noise

5.2.2 Results and analysis

Rotation

The rotational component is assessed visually in its original axis/angle representation $\omega = \theta\hat{\mathbf{n}} \equiv (\omega_X, \omega_Y, \omega_Z)$ to determine the accuracy of each component of θ with the aligned ground-truth. In Figure 5.4 and 5.5, it can be observed that the reconstructed rotational component against the ground-truth rotation increases in error as expected as more error is induced into the system. Although both approaches were subjected to a same rotational error of $0.0676°/m$, the other closely related parameters clearly have an effect on the refined reconstructed camera pose throughout the scene. Due to the incremental nature of the framework, it can also be observed that the 5.4 and 5.5 suffers from error drift with the rotational error increasing regardless of noise case. With regards to the no noise case, the majority of the error can be linked to computational inaccuracies induced by MATLAB and its associated chosen variable representations. In Figure 5.6 and 5.7, the error bar graphs exhibit the mean and variance of the error throughout the assessed camera poses, with the associated error bars for both descriptive measures becoming more pronounced as the error introduced becomes more prominent.

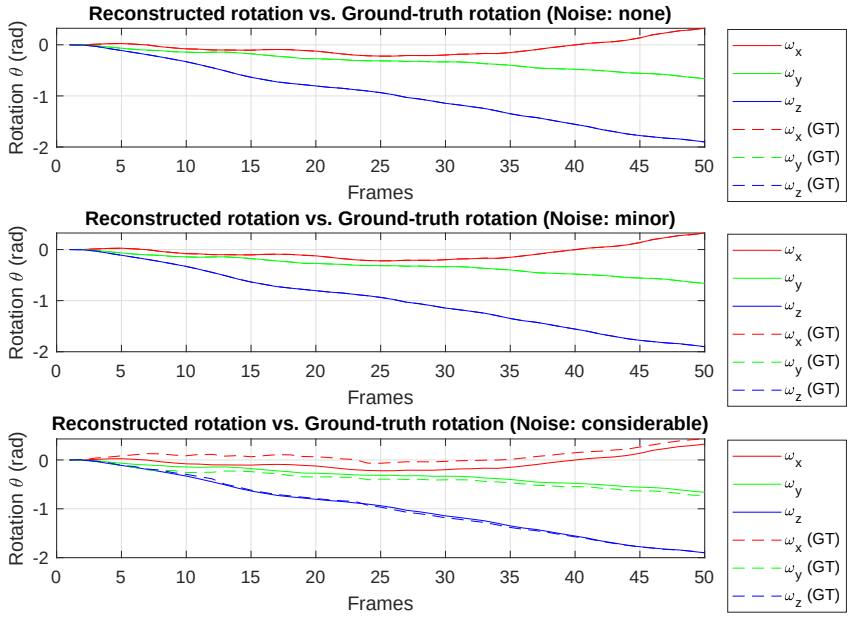


Figure 5.4: Reconstructed rotational component vs. ground truth rotational component

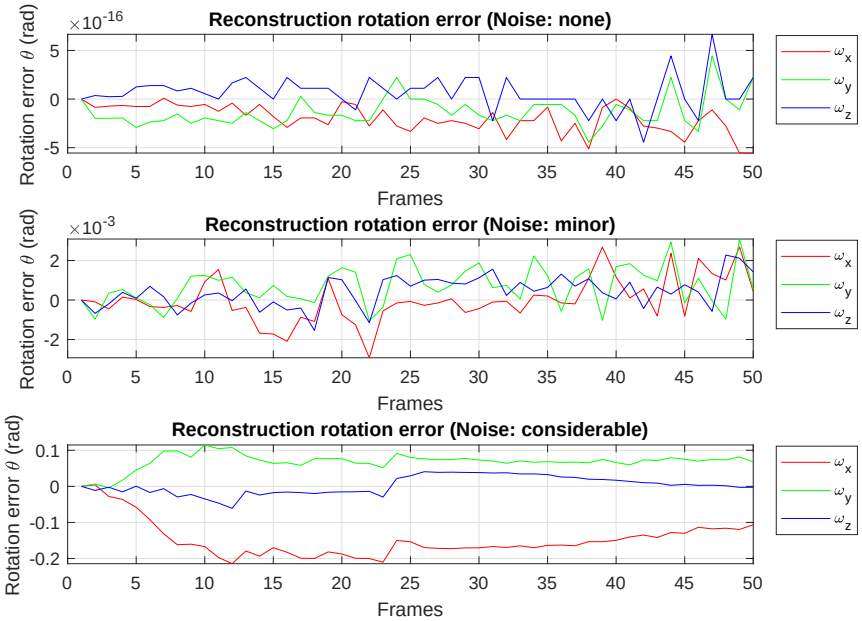


Figure 5.5: Reconstructed rotational error

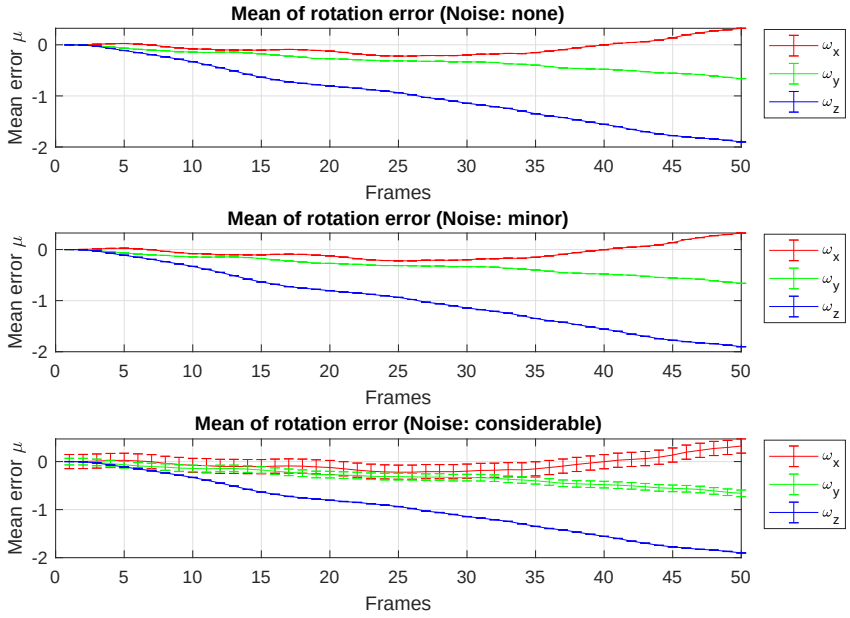


Figure 5.6: Average reconstructed rotational error

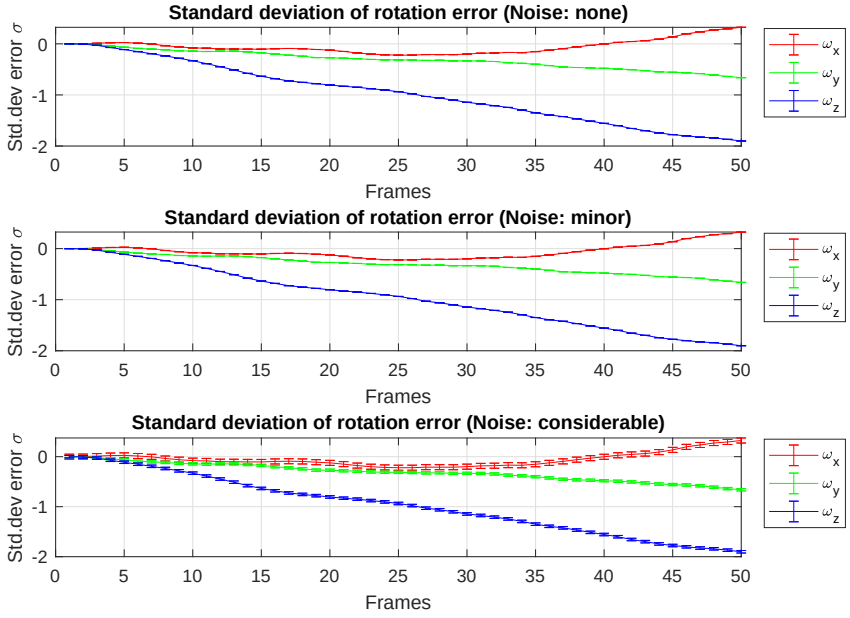


Figure 5.7: Standard deviation of the reconstructed rotational error

Translation

The translation component of the framework is extracted from the $\mathbf{P}_{3\times4}$ projection matrix used to express the camera pose and is simply expressed simply as a 3-vector $\mathbf{t} = (t_x, t_y, t_z)$. With regards to translation, both error cases differ with regards to the noise subjected, with the minor case exhibiting an average mean Euclidean distance of approximately 3.9% and the considerable noise case adding 9.95% into the error cases. It becomes clear in both Figure 5.8 and 5.9 that for the minor noise case, the refined error is actually very low in comparison to the deviation of the translation in the considerable case as expected - this is demonstrated in both Figure 5.10 and 5.11 where the mean and standard deviation of the error values where the minor error cases refined reconstruction aligns closely with the ground-truth whereas the considerable noise case deviates by approximately 0.5m in t_y and t_z components of the translation vector $\mathbf{t}$. It is clear that the error introduced to the 3-D scene locations in addition to the noise introduced to the translation invokes considerable error, and this is demonstrable in the point-cloud reconstructions shown in Figure 5.3.

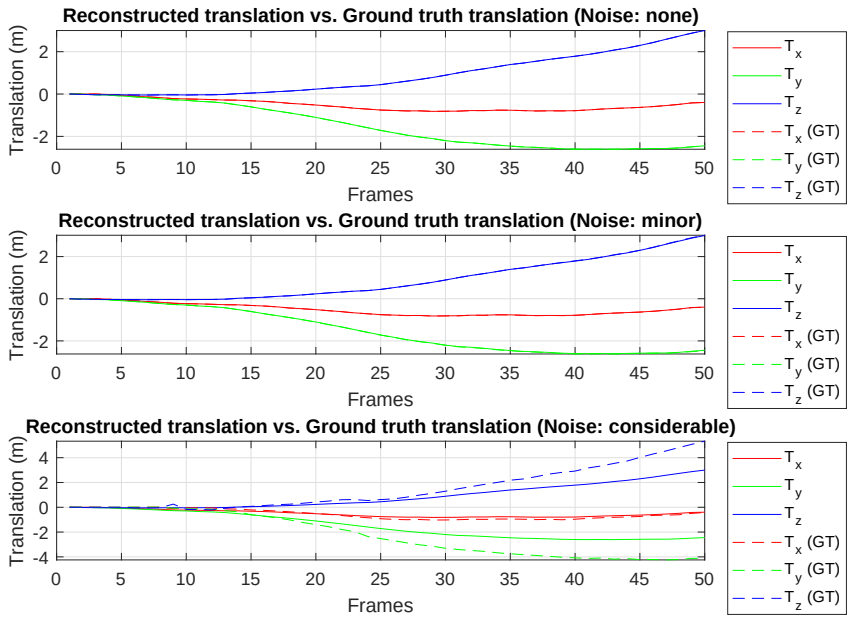


Figure 5.8: Reconstructed translational component vs. ground truth translational component

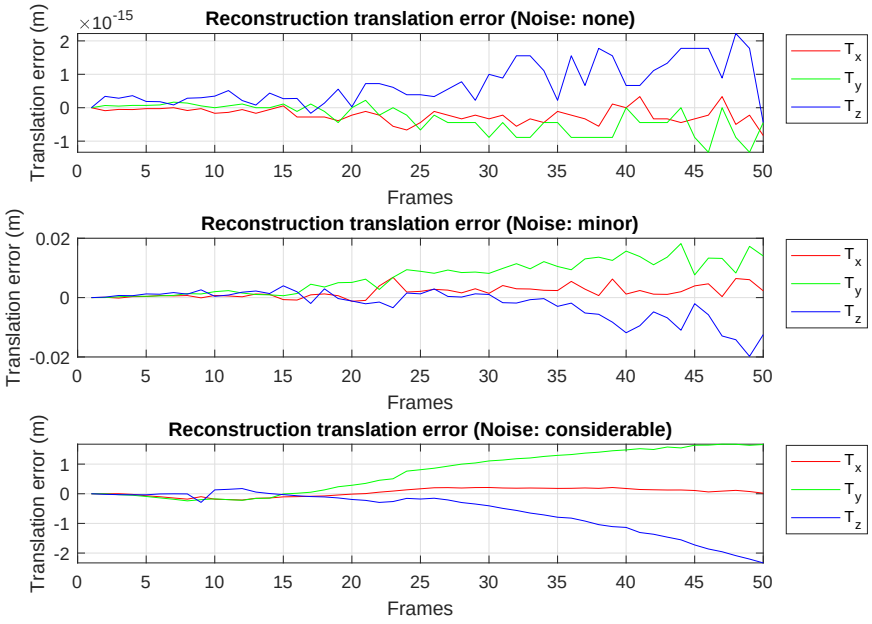


Figure 5.9: Reconstructed translational error

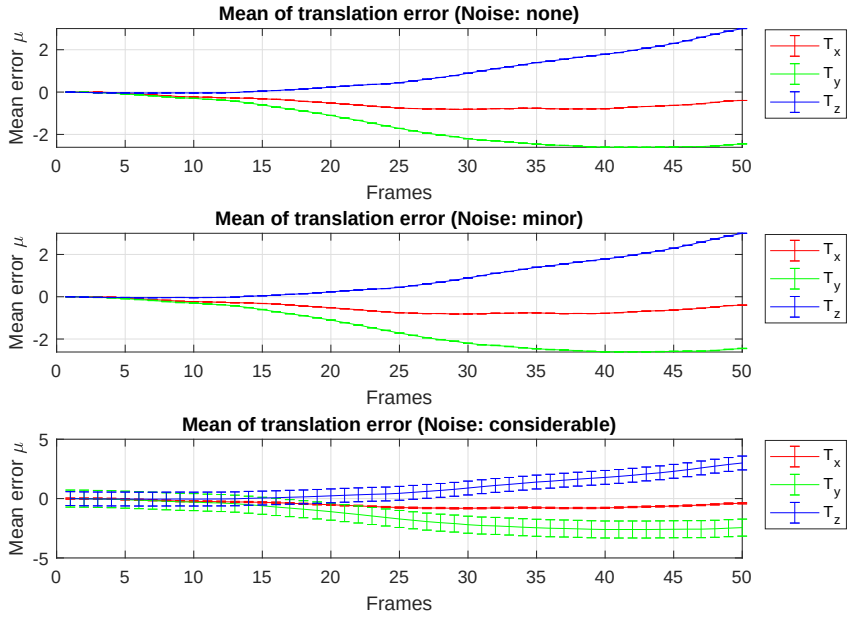


Figure 5.10: Average reconstructed translational error

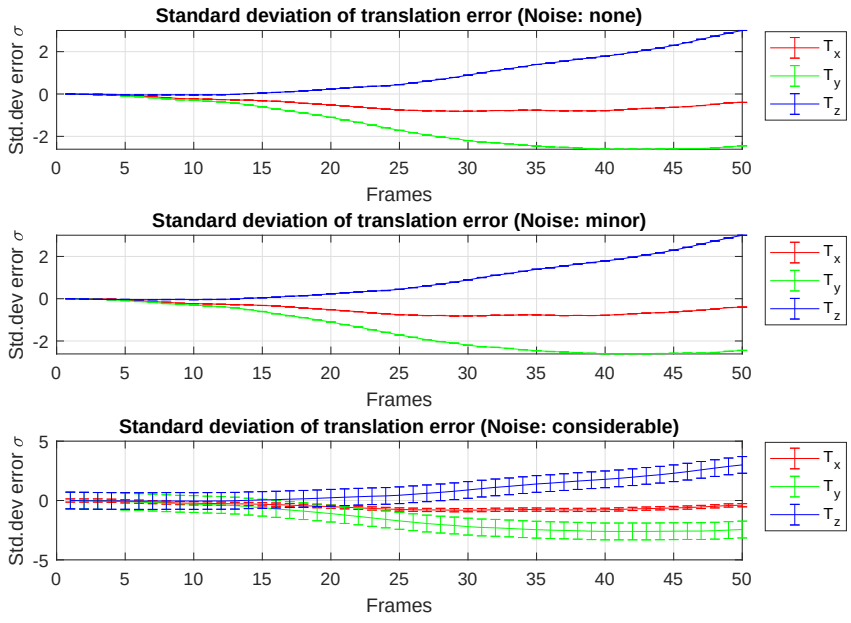


Figure 5.11: Standard deviation of the reconstructed translational error

Reconstruction results

Parameter	No noise	Minor noise	Considerable noise
E_{reproj} (px)	5.59e-27	0.145	232.85 (25.14)
E_{trans} (m)	(-2.15,3.16,6.97)e-16	(0.001,0.007,-0.002)	(0.04, 0.7, -0.5)
E_{rot} $\theta(rad)$	(-1.98,-1.29,6.25)e-17	(-0.0001, 0.0007, 0.0004)	(-0.14, 0.068, 0.002)

Table 5.3: Reconstruction error and the mean reconstructed rotational and translational error

As observed within the reconstruction framework, when no noise is introduced, the error source is extremely small and can be attributed to number representation and precision within MATLABs own language. As the noise introduced to the system becomes minor, the E_{reproj} is a highly desirable 0.145, and would be expected in a system that is well calibrated, or with a system performing auto-calibration within a feature-rich environment. Within a considerable noise environment, its expected that there will be errors in the correspondence searching algorithm and in addition to the noise subjected to the camera calibration matrix, this combines to induce a considerable E_{reproj} error value of 232.85. The reason this value is so incredibly high is due to a couple of 3-D refined triangulated points been calculated well away from the actual scene as observed in Figure 5.3c). When these explicit values are removed from the calculation, the E_{reproj} value drops to 25.14 px.

5.3 Discussion

This chapter introduced the CH-SfM framework, which incorporates the matching procedures theorised within Chapter 4, with the aim of producing robust feature correspondences when ascertaining the 3-D geometry of constrained, homogeneous environments. The fundamental reason for developing a bespoke reconstruction environment was due to many openly available software for 3-D reconstruction is unsuitable for challenging environments, with many generalised approaches such as VisualSFM [25] and COLMAP [21] relying heavily on the target scenes being feature rich and containing lots of unique, salient feature points to produce accurate 3-D reconstructions. Funda-

mentally, most reviewed methods incorporate reconstruction methodologies that focus primarily on the completeness of a scene rather than the accuracy, resulting in the scene reconstructions suffering from considerable noise or just failing to produce a reconstruction altogether - this is further reviewed in the following chapter. Consequently, this provided the motivation in order to produce an incremental SfM framework which focuses on the accuracy of the output, rather than needlessly densifying the reconstruction at the expense of noise and obfuscation of the underlying scene.

To ensure transparency in the work, this chapter focuses on validating the reconstruction capability of the framework, and demonstrating the interdependencies of different important parameters and how this can cause a ripple with regards to the reconstruction and camera pose estimation accuracy. In association with this chapter, Appendix A goes through the framework at a lower-level with regards to the intrinsic dependencies of the framework and how each core part of the framework explicitly operates, and in Chapter 6, the framework is introduced to a real-world example using AGR RVI inspection footage. To evaluate the reconstruction accuracy of the framework, synthetic data is generated with additive noise applied to determine the real-world capabilities of the framework when handling potential instrumentation noise and general residual error from the matching and reconstruction procedure - this is of particular importance due to the constrained, homogeneous environments in which the reconstructions are performed. In the following chapter, an evaluation is performed of the CH-SfM framework using real world examples of footage acquired within a constrictive, feature poor environment against the current commercial and academic state-of-the-art.

Chapter 6

3-D Visualisation of Advanced Gas-cooled Reactor Fuel Channels

Summary of chapter contributions

This chapter aims to coalesce the research and proposed novelties in previous chapters and apply it to the target problem space of this thesis; the reconstruction of AGR fuel channels using RVI source imagery. To evaluate image-based 3-D reconstructions of the AGR fuel channels, both in-core and laboratory footage captured using bespoke experimental apparatus which emulates a sub-section of the fuel channel was utilised. Firstly, a distortion analysis is performed to determine the robustness of the proposed MIFM approach, in addition to determining the limitations of the CH-SfM framework, by performing matching and reconstructions with increasingly deficient imagery. Although the state-of-the-art outperforms the proposed MIFM method using high quality source imagery, the MIFM approach exhibits a 138.53% performance improvement in match density when evaluating challenging, feature-poor imagery. The reconstruction quality of the CH-SfM framework is quantified by point density, and reconstruction MSE via cylinder fitting based on start-of-life AGR fuel channel dimensions, with in-core footage obtaining 2.78-8.42N_X/cm^2 with an approximate MSE of 4.69 - 5.01 and laboratory footage obtaining 23.8N_X/cm^2 with an approximate MSE of 3.91. The assessment indicates that the CH-SfM framework is suitable for image-based 3-D reconstruction of even most challenging of footage for diagnosis and prognosis of the derived AGR fuel channel structure.

6.1 Introduction

The purpose of this chapter is to take the theory and practical elements developed in previous chapters and apply it to a specific application domain which is the underlying motivation for the work performed in this document. In Chapter 2.2.2 - 2.2.3, the manual inspection of an AGR fuel channel is detailed and how, over time, the inspection methodologies have evolved in parallel with the increasing need for visual inspection. At reactor start-of-life, the condition of the graphite has been subjected to minimal degradation and therefore inspection of the core was limited. As the reactor ages, there is an increasing need to inspect and track the core health. Over time, different inspection tools such as NICIE2/CBIU were introduced, improving inspection capability [2, 44] in order to investigate and quantify structural defects as the reactors begin to age and exceed original design life-time estimations.

With many of the reactors planned to close over the next two decades, development of new inspection tools incorporating modern technologies to perform visual inspection is uneconomical. Consequently, it has become prudent to augment the tools hardware visualisation limitations with clever software-based solutions to facilitate the extraction of supplementary practical information to improve an inspection engineers understanding of the structural morphology of life-limited reactor components. For that reason, this provided the primary motivation for the development of a 3-D visualisation mechanism which would take pre-existing inspection footage and extract tangible structural geometry of the AGR fuel channel. This chapter is split into several sections; Firstly, an analysis is performed using the matching procedure developed in Chapter 4 to determine the robustness of the framework against different levels of challenging imagery. Secondly, a comparative analysis is performed against pre-existing incremental SfM frameworks with regards to its ability to extract representative geometry of the fuel channel using both in-core and laboratory footage. Validation is performed to ascertain the reconstruction density capability of the framework using bespoke experimental apparatus which emulates a section of the AGR fuel channel. Finally, the concept of circumferential reconstructions are introduced by combining orientational reconstructions

together into a singular cohesive reconstruction.

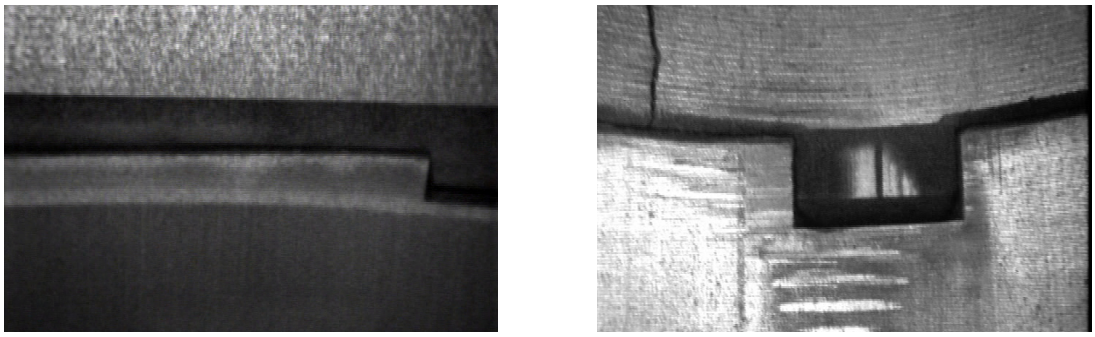
6.1.1 Footage sources

The frameworks tested within this chapter are assessed by an assortment of footage gathered from within AGR fuel channels of varying quality and are discussed below with associated image stills in Figure 6.1 of in-core footage and Figure 6.2 for laboratory footage respectively:

NICIE/CBIU footage The majority of footage obtained within the reactor inspection is gathered either by the NICIE inspection tool family or the CBIU tool. The usable resolution of this camera is 720×400 captured at 25 FPS, with the FoV guides and inspection overlay being cropped out the source footage.

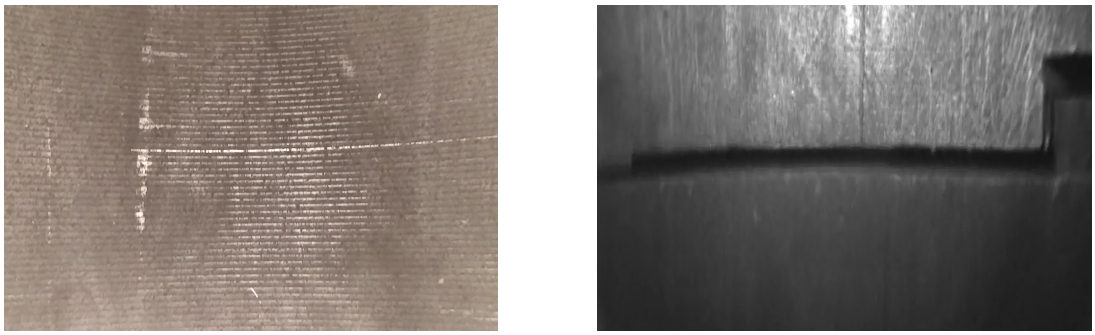
CRVW footage The CRVW is used for channels that exist on the outer rim of the fuel channel lattice structure. The CRVW footage is captured using the same format as the NICIE/CBIU, however the inspection methodology with this tool is different from the inspection protocol used by NICIE/CBIU. The footage quality produced is much sharper and captures more textural information from the AGR fuel channel interior, despite being the same image resolution as the NICIE/CBIU.

Laboratory footage Laboratory footage pertains to image sets that have been captured either using a manually translated camera system capturing footage with a resolution of 1920×1080 (1080p) at 30 FPS or a translation stage with the NICIE/CBIU inspection tool mounted onto it to simulate the inspection protocol observed in NICIE/CBIU in-core inspection imagery.



(a) (b)

Figure 6.1: In-core image stills from different inspection tools of the AGR graphite brick inner surface. a) NICIE/CBIU image still; b) CRVW image still;

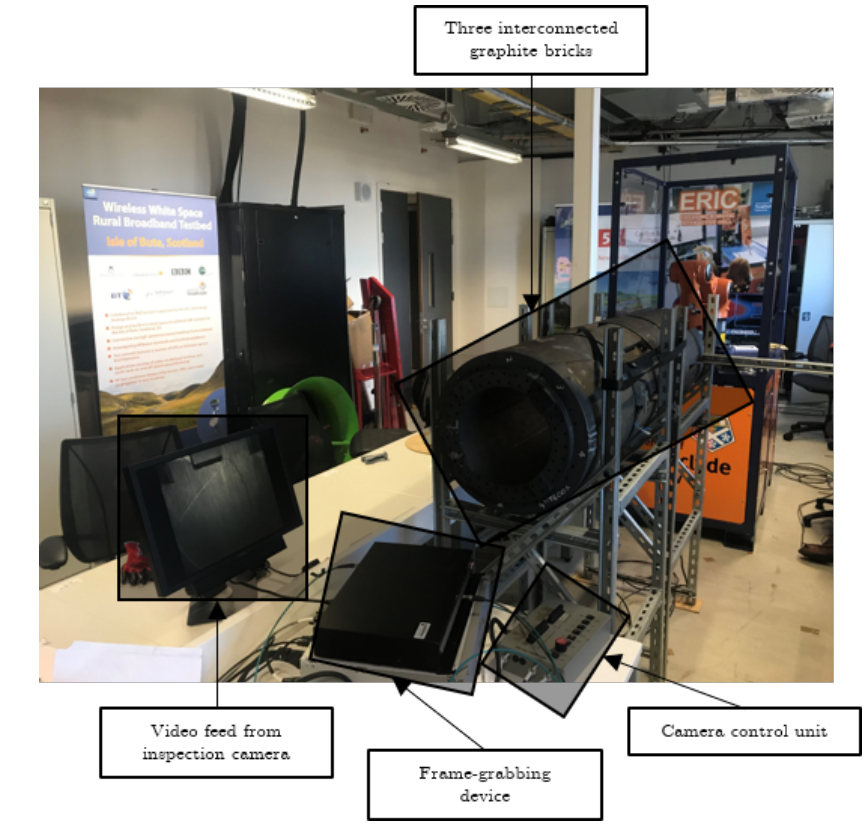


(a) (b)

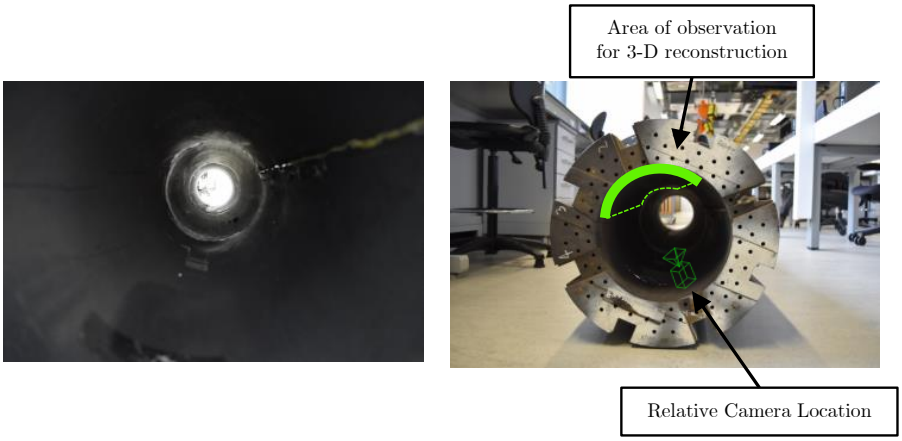
Figure 6.2: Laboratory image stills from different inspection tools of the AGR graphite brick inner surface. a) 1080p image still captured using Toshiba T4KA7 camera sensor in laboratory conditions; b) NICIE/CBIU image still captured in laboratory conditions.

To capture footage within the laboratory, 3 AGR graphite bricks sourced from different reactor designs were interconnected to form a small representative sub-section of an AGR fuel channel which normally consists of 10-12 layers depending on the reactor design. Some of the graphite bricks contained cracks from prior experiments. The laboratory footage was gathered using both the NICIE inspection tool, with the tool being moved through the AGR fuel channel sub-section mounted onto a translational stage. Additional footage was also gathered using an high quality, low-light Toshiba T4KA7 camera sensor which was manually translated, capturing a singular orientation at approximately 2.5× the normal AGR footage resolution. The experimental setup is shown in Figure 6.3a). The inspection procedure is similar to how an orientational scan

is performed during real in-core fuel channel inspections where the NICIE inspection tool observes a single orientation during the scanning procedure. Where the approaches differ however is that during in-core inspection, the tool is suspended into the channel using a crane with an umbilical wire, whereas our approach transforms the camera on a translational stage or manually. As a result, the in-core footage is subjected to unwanted rotation whilst the tool is being suspended and pulled up and down the channel whereas the translation in laboratory conditions is controlled. The manual footage captured on the hand suffered abrupt motions and slight rotations was utilised to further test the framework capability.



(a)



(b)

Figure 6.3: Images of the experimental apparatus used to evaluate the 3-D reconstruction framework. a) Image of the framework and the associated hardware used to inspect the rig; b) Down-channel view of the channel and diagram highlighting the area of observation of the lab footage.

6.1.2 Reconstruction methodology

To perform 3-D reconstructions of the channel, the image dataset is first manually processed and split into 6 different folders, each representing a scan performed at each discrete orientation. After this procedure is complete, reconstructions are generated using the CH-SfM framework for each individual orientation. An orientational reconstruction represents a reconstruction produced using imagery obtained from a singular scan within the AGR fuel channel. Once 6 orientational reconstructions have been produced using the CH-SfM framework, the CH-SfM framework combines all 6 reconstructions into a single circumferential reconstruction, and performs post-processing on the point-cloud to remove noisy 3-D points based on expected geometry. A simplified version of the framework process described in Figure 6.4:

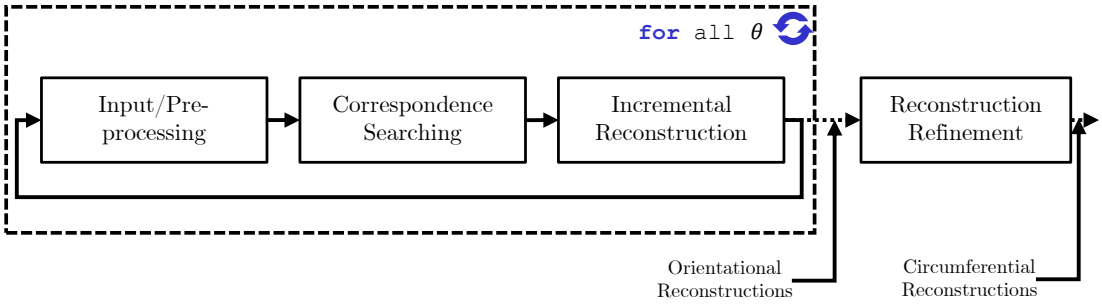


Figure 6.4: Simplified block diagram of the CH-SfM with the additional reconstruction refinement module

To assess the effectiveness of the CH-SfM framework, it is compared against two other state-of-the-art incremental reconstruction frameworks - VisualSFM and COLMAP. Since circumferential reconstructions are application-specific, the evaluation is performed using only orientational reconstructions. The primary objective of this comparison is to evaluate the proposed frameworks ability at producing accurate and representative 3-D scene geometry.

6.2 Distortion analysis

Repetitive texture, low resolution and non-uniform illumination are identified as the key challenges to performing 3-D reconstruction using AGR footage. Not only are the reconstructions distorted, but areas that exhibit such properties often suffer from a low reconstruction density as a result of the matching scheme deployed being unable to derive a robust set of matches within this region. Consequently, it is important to understand at what stage does the correspondence searching procedure described in Chapter 4 begin to suffer from match degradation. To perform this assessment, images are convolved by a Gaussian smoothing kernel with the incremented standard deviation of the Gaussian distribution value $\sigma_i = [0.1, 10]|i = 1...50$ and 3 different texture-analysis measures are used to quantify the "uniqueness" of the assessed images convolved by the variable-Gaussian:

Entropy Shannons entropy in an image processing context can be used to determine the statistical randomness of an image. Images with a high entropic value are associated with heterogeneous imagery consisting of high information content. Conversely, an image with a low entropic value are associated with homogeneous imagery which consists of low information content [59]. The entropy is calculated in this application using the following equation:

$$E = -\sum_{k} p_k \log_2(p_k) \tag{6.1}$$

where $k = 256$ image bins are calculated for 8-bit grayscale imagery (0-255) using a 2-D histogram of image gradients to ascertain the probability p_k.

Variance The variance helps quantify to the average deviance from the image statistics. A image with high variance is associated with heterogeneity whereas an image with low variance is indicative of textural uniformity and homogeneity.

Standard deviation Like variance, the standard deviation quantifies the statistical spread of the image statistics from the average value and is effectively the square root of the variance.

Three sets of image data are used for the evaluation; a generic scene which comprises of images taken of a computer desk with mixed degrees of saliency, laboratory footage taken from within the AGR fuel channel brick apparatus of a mid-brick region as shown in Figure 6.3 and real in-core inspection footage also of a mid-brick region. The results produced from the 3 datasets can be seen in Figure 6.5 for each respective dataset, and Table 6.1 for the numerical extrema produced for each texture analysis measure.

As observed in both Figure 6.5 and Table 6.1, the desk scene contains a high degree of heterogeneity with the entropy, variance and standard deviation of the underlying imagery vastly exceeding the footage captured within the AGR fuel channel. This arbitrary scene is used to effectively contextualise the values obtained from both the laboratory and the in-core brick footage. When comparing the laboratory and in-core footage, the entropic values are considerably closer and this is due to the images conveying repetitive structures. This is indicated further with the extracted variance extrema, with the high quality laboratory footage demonstrating higher levels of variance, with the maximum variance of the laboratory footage being nearly twice that of the in-core footage. To test the matching performance of Motion-based Iterative Matching (MIFM), the correspondence searching algorithm proposed in Chapter 4, both the laboratory and in-core footage will be the basis for testing.

	Entropy		**Variance**		**Standard Deviation**	
Image source	*Min*	*Max*	*Min*	*Max*	*Min*	*Max*
Desk scene	5.7	7.86	508.83	7798.44	22.56	88.31
Laboratory footage	6.2	6.82	461.22	1237.67	21.47	35.1806
In-core footage	6.43	6.53	531	626.47	23.04	25.03

Table 6.1: Minimum and maximum entropy, variance and standard deviation of the test footage convolved with variable standard deviation Gaussian Kernel ($\sigma = [0.1, 10]$)

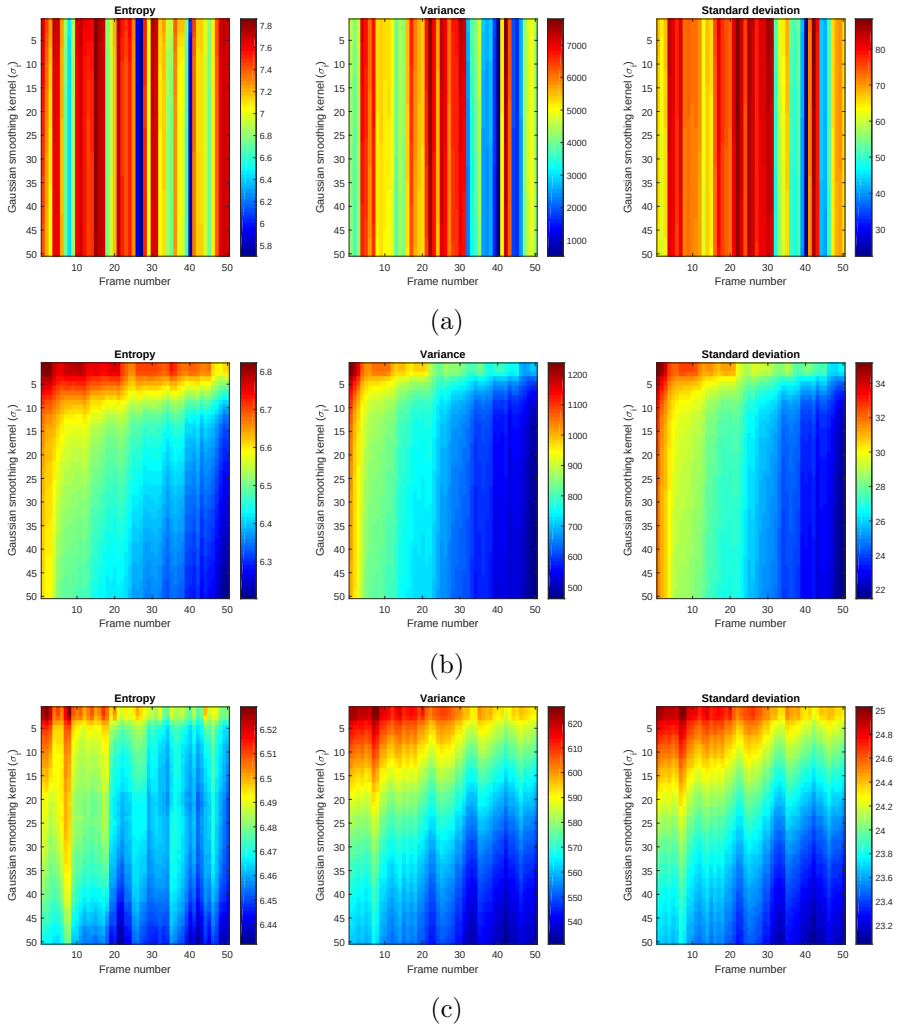


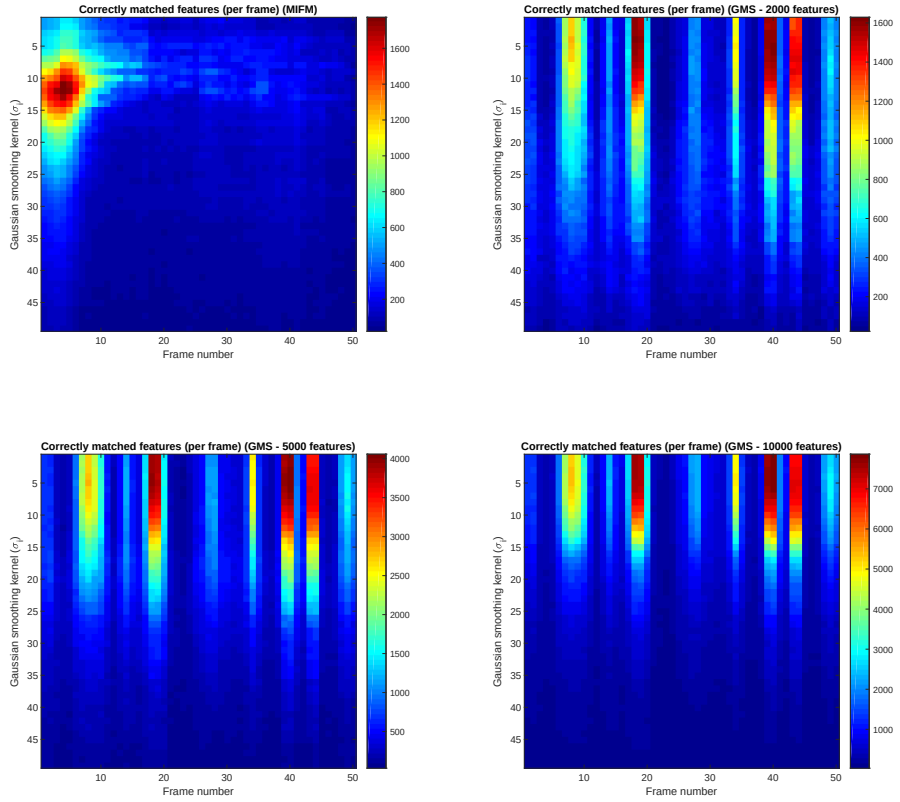
Figure 6.5: Entropy, variance and standard-deviation of a) High quality dataset of mixed saliency scene (Computer desk); b) High quality imagery of the AGR graphite brick; c) In-core inspection footage

6.2.1 Feature matching performance

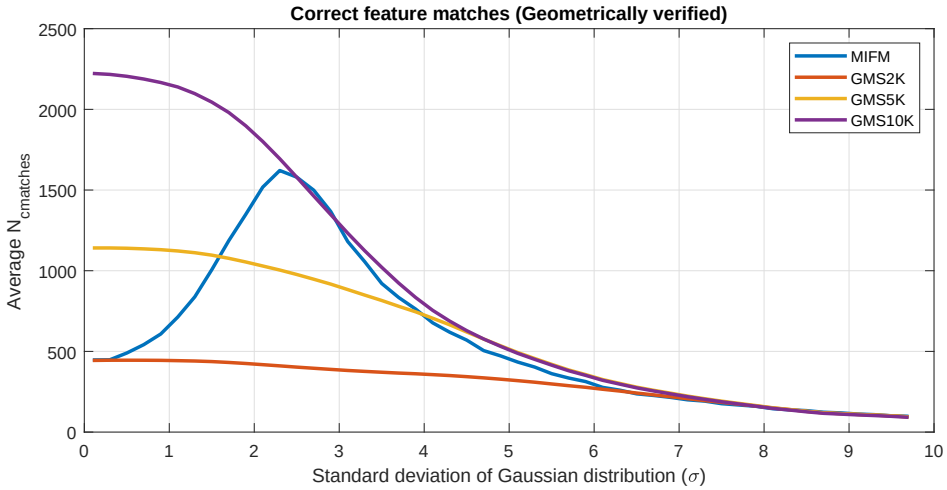
To perform representative 3-D reconstructions of the AGR fuel channels, the incremental reconstruction algorithm is heavily reliant on the underlying robust correspondence searching algorithm which can detect and match features. When the image is structurally repetitive, typical matching issues can occur as observed in Figure 4.15 where the resulting correspondence searching scheme either produces an insufficient number of matches or the robust estimation procedure used to geometrically verify the

matches converges on a incorrect matching consensus. Therefore, it is crucial that the underlying algorithm produces accurate 2-D correspondences in which the resulting reconstruction is derived from. To quantify the performance of the proposed novel semi-dense MIFM algorithm when dealing with challenging imagery, a comparison is made to its closest approach from a theoretical standpoint, the Grid Motion Statistics (GMS) algorithm [63] using AGR RVI imagery.

To simulate the varying levels of image quality observable within an AGR reactor, a sub-set of 50 images from both laboratory footage and in-core inspection footage is convolved within a variable-sigma Gaussian kernel to blur the images as described in the previous section, whereas $\sigma = 0.1$ is a very weak blurring filter and does little to the underlying imagery, to $\sigma = 10$ where the image loses considerable amounts of textural information as shown in Figure 6.5. The purpose of this approach is to demonstrate the matching capability of the novel MIFM against one of the best matching paradigms in literature, the GMS algorithm, using challenging AGR RVI footage. The GMS algorithm, like the MIFM approach, partitions the matching procedure into a series of spatial bins and determines region strength based on the matching strengths of its neighbouring matches and constrains the matching procedure accordingly. The MIFM method, however, exploits both sparse and dense image representations, performing matching based on the regional motion statistics to help constrain the problem, taking into account the intra-image variation in order to obtain a robust set of matches. To evaluate the matching procedure, the MIFM procedure will operate with Dense Motion Statistics (DMS) enabled, SLIC segmentation selected and an initial matching threshold of 2 (more information can be found in Chapter 4.3 and Appendix A.2.2 on these parameters). The GMS algorithm has been evaluated with 2000, 5000 and the recommended value of 10000 detected ORB features [109] per image [63]. Results are shown in Figure 6.6, 6.7 and Table 6.2:

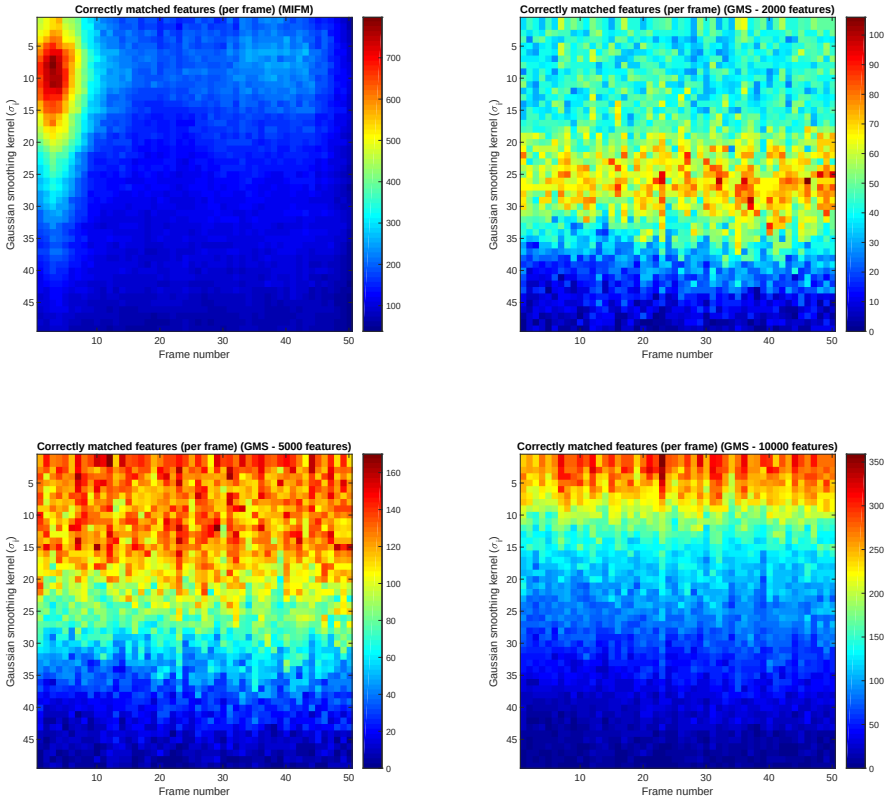


(a)

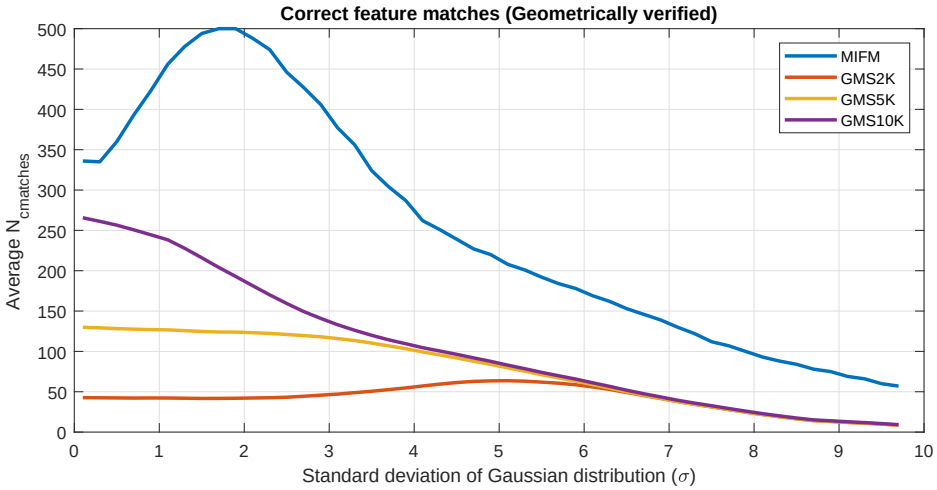


(b)

Figure 6.6: Mean feature observability and true feature matches across variable-Gaussian smoothed laboratory footage



(a)



(b)

Figure 6.7: Mean feature observability and true feature matches across variable-Gaussian smoothed in-core footage

	Average true positive matches			
Footage source	**MIFM (Proposed)**	**GMS (2K)**	**GMS (5K)**	**GMS (10K)**
Laboratory footage	568.26	299.4	586.58	**873.61**
In-core footage	**251.75**	41.5	76.93	105.54

Table 6.2: Average true positive geometrically verified feature matches using filtered laboratory and in-core footage

As observed in Figure 6.6 (laboratory footage) and Figure 6.7 (in-core footage), the proposed MIFM feature matching approach attains better performance when the underlying image is smoothed by a Gaussian filter in both scenes with a $\sigma = 0.8 - 2.2$, with the peak being observable in both Figure 6.6 and Figure 6.7 with regard to feature observability and the resulting average of correctly geometrically verified matches. It is assumed that by pre-filtering the image with a Gaussian filter will negate SIFT points located at noise in the DoG scale space and also improve the resulting motion field produced by the DMS component of the MIFM algorithm. This observation will integrated as a standard pre-filtering method for the MIFM approach in future work.

One of the notable strengths of the MIFM framework is the consistency of the framework with its ability to observe the detected SIFT features across a number of frames, making the matching system considerably more robust. On the other hand, the GMS matching approach is excellent at deriving pairwise matches across small baselines however struggles at maintaining the observability across more than 2-3 images. This is clearly observable in Figure 6.6 where there is vertical bands pertaining to stronger image pairs, whereas the method struggles with poorer quality images at the same assessed σ value. With the GMS approach using densely sampled ORB features[1], which trades off descriptive properties for computational speed [159], resulting in a lower number of correct matches as the baseline is increased.

When assessing the laboratory footage, the GMS algorithm on average outperforms the MIFM algorithm by 53.73% when the GMS algorithm utilises 10,000 (GMS10K) features. On average, the SIFT algorithm in the MIFM approximately calculates around 3,000 features per image, and this provides the motivation to provide the matching

[1]The ORB feature descriptor is baked into the GMS algorithm `https://github.com/JiawangBian/GMS-Feature-Matcher/)`, and it relies on the properties of the ORB feature descriptor to function

results when GMS operates with 2,000 (GMS2K) and 5,000 (GMS5K) ORB features respectively to provide parity to the results. When operating with 5,000 features, it attains around 3.22% more feature matches and with 2,000 features, the MIFM attains a 47.31% performance improvement. As observed in Figure 6.6b), when the image consists of strong anchor points ($\sigma < 2.5$), the GMS matching algorithm performs very strongly and attains considerably more matches than the proposed MIFM algorithm. The fundamental reason to this is that the GMS is heavily reliant on strong intra-image regions to effectively guide the rest of the matching procedure and this is possible when the image suffers minimal distortion. However, as the image is subjected to a Gaussian filter with $\sigma > 2.5$, the effectiveness of the GMS10K variant of the algorithm approximates the performance of the MIFM algorithm across the rest of the evaluated range of σ_i. It also demonstrates that once $\sigma > 6.5$, the MIFM and the tested GMS variants (GMS2K/GMS5K/GMS10K) all converge on similar performance with an average of around 100-200 matches attained even when the underlying images are significantly blurred.

The in-core footage results in Figure 6.7 demonstrate the strength of the proposed MIFM approach when dealing with considerably more challenging image footage, where the entropy and underlying statistical variance is considerably lower than the laboratory footage. On average, the MIFM approach produces on average 251.75 true positive geometrically verified matches as shown in Table 6.2, whereas the GMS10K variant only manages on average around 105.54 true positive geometrically verified matches. This indicates that the MIFM algorithm can achieve 138.53% more matches on average over the GMS10K variant of the algorithm. With the in-core footage, the MIFM approach consistently outperforms the GMS variants across the entire assessed σ spectrum, with the GMS variants all converging in performance when ($\sigma > 6$) at around 60 matches obtained per frame. As ($\sigma > 8$), the GMS algorithm begins to struggle, producing less than 25 matches per frame. On the other hand, the MIFM algorithm never drops below 50 matches regardless of σ value, indicating its strength across the potential spectrum of in-core mid-brick layer obfuscation, and its ability to operate in deficient feature spaces.

6.2.2 Reconstruction density and limitations

The next stage of validation is to determine the quality of the feature matches derived using the MIFM feature matching scheme with respect to resulting 3-D reconstruction density acquired using the CH-SfM framework described in Chapter 5 and Appendix A. To evaluate the reconstruction quality, 3-D reconstructions were performed on progressively stronger Gaussian-smoothed laboratory footage dataset evaluated in Section 6.2.1 with the σ value being increased in steps of 0.8 in a range between $[0, 4]$. The reason this range being selected is as the image becomes increasingly blurred, although the features can be still be derived based on the properties of SIFT and ORB for the respective MIFM and GMS algorithms, the underlying depth cues derived from the features get increasingly obscured. Consequently, as the $\sigma > 4.0$, the CH-SfM framework fails to reconstruct any representative geometry. Therefore, limitations with the framework can be quantified by using statistical properties of the image, given that $\sigma > 4$ equates to $i = 21$, by taking the average values for the laboratory image on the 21^{st} row of the entropy, variance and standard deviation matrices in Figure 6.5, the approximate lower-limits of the framework can be derived:

	Approximate lower limits
Entropy	6.46
Variance	700.18
Standard deviation	26.33

Table 6.3: Approximate lower limits for Entropy, Variance and Standard deviation for CH-SfM framework to operate

It is important to note that both the variance and standard deviation lower limits proposed above in Table 6.3 are somewhat misleading due to the global nature of the calculations, which can obscure localised statistics associated with salient regions of the image. Entropy gives a stronger indicator as it captures the local statistics and the pixel co-occurrence during the image histogram procedure, and provides a better metric for determining a numerical lower limit for the system. Ergo, it can be expressed that the input images must have an minimum entropy value of at least 6.46 for the reconstruction procedure to operate successfully.

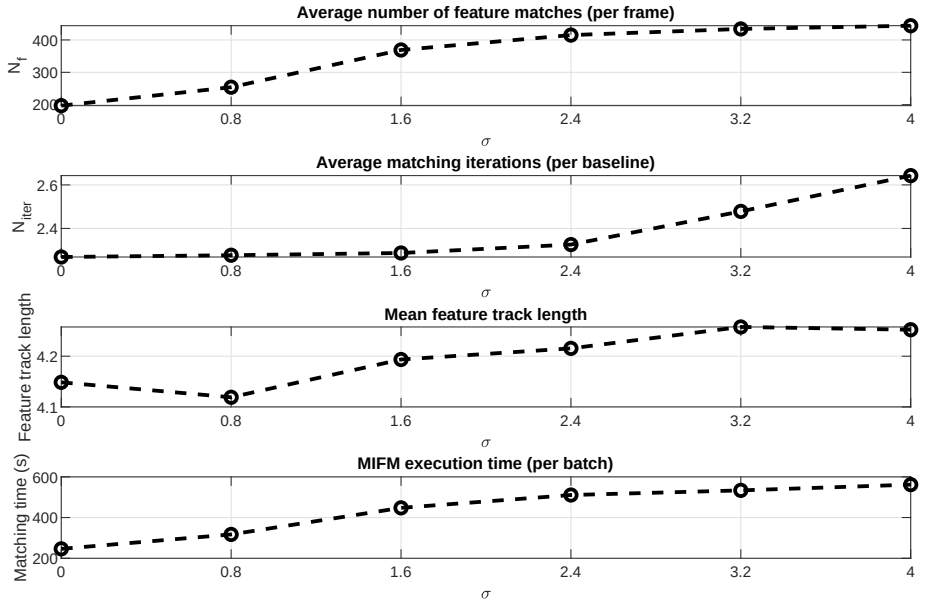
Parameter name	Parameter value
`mC.seqMatch`	10
`mC.minMatches`	150
`mC.bline`	3
`mC.initThresh`	1.8
`mC.DMS`	1

Table 6.4: Key parameters used for the CH-SfM framework to evaluate 3-D reconstruction density and accuracy of laboratory footage

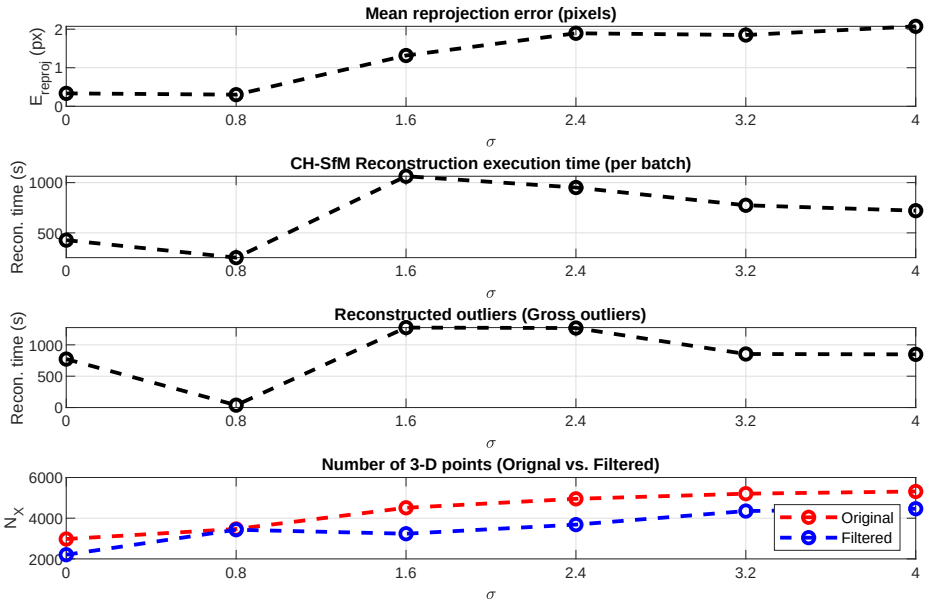
To evaluate the reconstruction procedure and resulting accuracy, the CH-SfM reconstruction parameters selected for the evaluation can be viewed in Table 6.4, with more information about the parameters can be found in Appendix A.2.1 - A.2.2. To determine reconstruction density, the feature points were spatially binned into a 10x10mm grid (cm^2), with the average point density for the orientational scan being captured. The Mean Square Error (MSE) of the 3-D reconstructions were captured using a cylinder fitting procedure, where a single orientation was replicated and rotated to form a cylindrical model which would then be the object of the cylinder fitting procedure, where the base model was constructed using start-of-the-life brick geometry. The numerical results can be viewed in Table 6.5, reconstruction and matching metrics shown in Figure 6.8 and 6.9 below:

Sigma	$\bar{E}_{reproj}$**(px)**	N_X	**3-D point density** (N_X/cm^2)	**Mean Square Error (MSE)**
0	0.3350	2977	8.8160	4.0850
0.8	0.3013	3470	13.7280	4.0949
1.6	1.3168	4512	12.9480	4.6240
2.4	1.8937	4953	14.7400	4.3540
3.2	1.8483	5205	17.4040	4.2121
4.0	2.0737	5313	17.8600	4.6304

Table 6.5: Reconstruction results for Gaussian-smoothed imagery

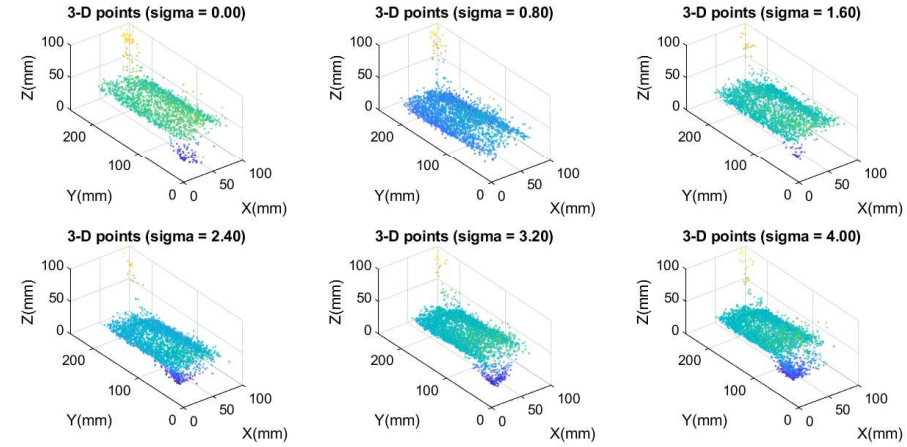


(a)

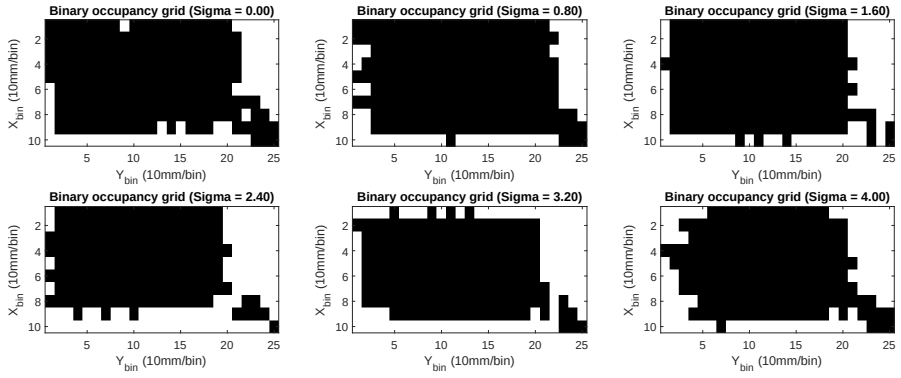


(b)

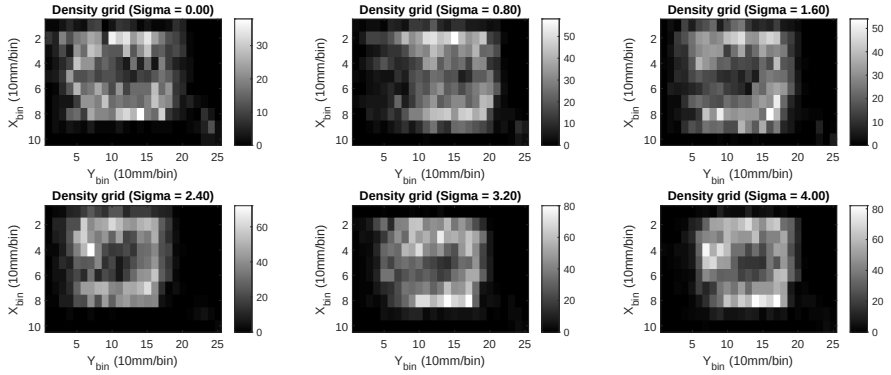
Figure 6.8: CH-SfM reconstruction results when assessing laboratory footage with increasing blurring strength; a) Matching metrics produced by the CH-SfM framework; b) Reconstruction metrics from the CH-SfM framework.



(a)



(b)



(c)

Figure 6.9: Reconstruction results using Gaussian-smoothed imagery; a) 3-D point cloud reconstructions; b) Binary occupancy grids; c) Density grids for each σ value.

As observed in Figure 6.8a), as σ increases, the number of feature matches rises and

this is due to the constrained iterative matching procedure having to perform more iterations on average in order to find more matching points. Consequently, not only does this increase the number of features and feature observability of the scene, it comes at a cost with an increase in MIFM matching and CH-SfM reconstruction times. By increasing the number of matches within the system, although improving point-cloud density as shown in Table 6.5, this is induces more error into the scene with the mean reprojection error $\bar{E_{reproj}}$ value increasing as σ increases. With σ increasing, and the MIFM approach having to work harder to find 2-D correspondences, the number of reconstructed outliers also increases significantly, with nearly every reconstructed 3-D point cloud resulting in approximately 1000 outliers with the exception of $\sigma = 0.8$. When $\sigma = 0.8$, the matching procedure finds a optimal point, where the Gaussian filter is blurring spurious noise and the resulting reconstruction has minimal outliers, low $\bar{E_{reproj}}$ error of 0.3013 and a highly representative point cloud as shown in Figure 6.9a). Interestingly, every point cloud regardless of σ value produces noisy 3-D points which is visible in the bottom right of all binary occupancy grids produced in Figure 6.9, and this is the result of a blurry surface visible just before the camera is inserted into the laboratory brick. With regard to 3-D point density, it is clear that as the MIFM becomes more aggressive in trying to find matches, this results in more reconstructed 3-D points. However, there is a trade-off, where the point-cloud density may increase but the resulting error in the 3-D point cloud becomes more significant. This is an issue present in the test cases evaluated using both VisualSFM and COLMAP in Chapter 3.4. Consequently, the best achieved result would be that of $\sigma = 0.8$, which exhibits a point cloud density of $13.72 N_X/cm^2$, and retains a high reconstruction accuracy. This corresponds with the results presented in Section 6.2.1 where the use of minor Gaussian blurring improved the matching accuracy. The high density value of the scene also corresponds to the high resolution imagery (1920×1080) used for the laboratory footage, allowing for significantly higher reconstruction resolution than active inspection approaches discussed in Chapter 2.3.1. Fundamentally, however, it is important to attain a balance between reconstruction density and representation, as sparse reconstruction methods can be augmented by MVS to produce fully dense,

photometric representations of the inspected scene [88].

6.3 Comparative analysis between CH-SfM and state-of-the-art incremental SfM frameworks

To evaluate the CH-SfM framework effectively, a comparative analysis is prepared using subsets of imagery from both the laboratory and in-core inspection footage against state-of-the-art incremental reconstruction frameworks VisualSFM and COLMAP. The aim is to determine the effectiveness of the CH-SfM framework and its underlying MIFM algorithm described in Chapter 4.3 and evaluated in Chapter 6.2. During the evaluation procedure, no *a-priori* geometric constraints or initial camera calibration were used by any framework, and the matching parameters were matched as closely as possible. Evaluation of the reconstruction was performed using known geometry of the AGR graphite bricks.

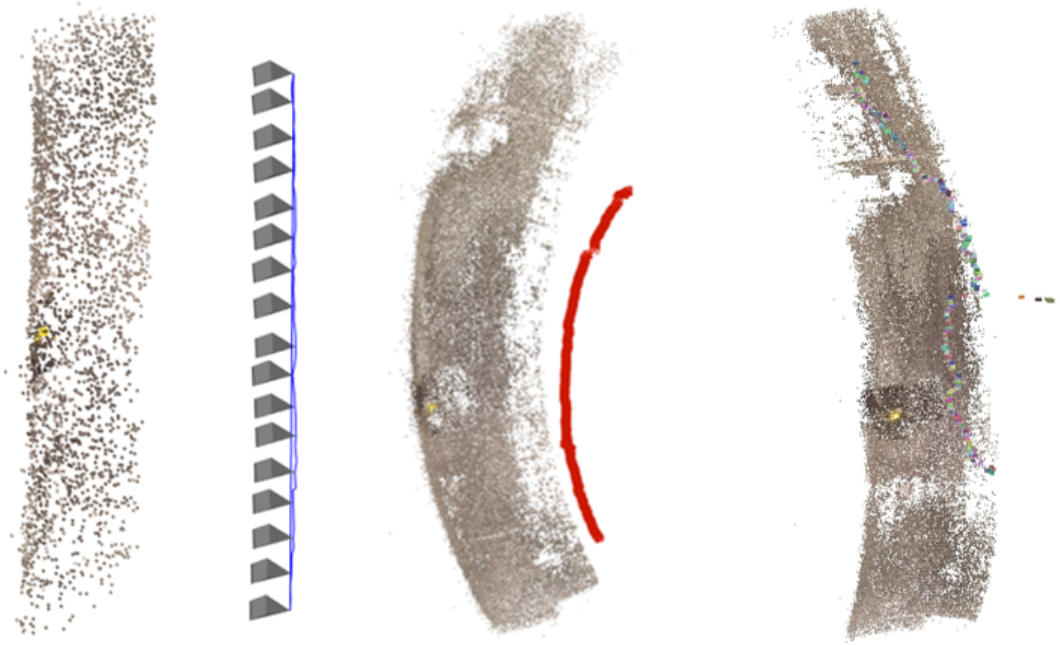
6.3.1 Laboratory footage reconstructions

Figure 6.10 and Figure 6.11 shown the results for each respective framework. The COLMAP and VisualSFM reconstructions in Figure 6.11b) and c) respectively struggle to interpret the curvature of the brick geometry correctly. The point clouds suffer from a large degree of geometric error with the resulting point clouds produced by COLMAP and VisualSFM bending from the beginning to end, forming a crescent shape (a.k.a. a lune in planar geometry). By using the top down-view of the point-cloud, the resulting point cloud in 2-D should approximate an arc shape, which the proposed framework does in Figure 6.11a). The error that VisualSFM and COLMAP are both subjected to is often termed cumulative or incremental drift, which is a direct result of incorporating mismatches or very poor features matches either through a non-robust correspondence searching mechanism or the use of re-triangulation [21, 25] to densify the 3-D point cloud.

The cumulative drift of the point clouds produced using COLMAP and VisualSFM can be better observed from Figure 6.11b) and c) with a side on profile view of the

(a) (b) (c)

Figure 6.10: Top down view of lab-based orientational reconstruction of a single AGR brick. a) CH-SfM reconstruction; b) COLMAP reconstruction; c) VisualSFM reconstruction. Note that the proposed framework in a) adheres to the expected geometry of the brick.



(a) (b) (c)

Figure 6.11: Side on view of lab-based orientational reconstruction of a single AGR brick. a) The proposed framework reconstruction; b) COLMAP reconstruction; c) VisualSFM reconstruction. Note that the proposed framework in a) adheres to the expected geometry of the brick.

point clouds. As observed, the reconstructions from both methods arch forwards and backwards respectively whereas in comparison to our framework result in Figure 6.11a), it has minimal positional drift with the camera location correctly following a near linear path at a set distance away from the triangulated 3-D points and correctly represents the expected curvature of the brick. It is noticeable that in Figure 6.11, the reconstructions produced by the CH-SfM framework are noticeably sparser with 31,737 3-D points while the COLMAP and VisualSFM reconstructions containing 88,289 and 144,331 3-D points respectively. This is due to the CH-SfM framework tracking more reliable and stable feature points across a large number of views whereas COLMAP and VisualSFM inherently have a larger number of poor or ambiguous matches at the expense of reconstruction accuracy. This is reflected in the computation time of our framework being $9m : 9s$ whereas VisualSFM and COLMAP took $38m : 27s$ and $3h : 43m : 30s$ respectively. The reconstruction metrics produced by the CH-SfM framework are tabulated in Table 6.6 below, with the resulting visibility matrix, binary occupancy grid and the density grid being shown in Figure 6.12.

Reconstruction metrics	
E_{reproj}(px)	0.8350
N_X (Number of 3-D points)	31,737
3-D point density (N_X/cm^2)	23.8025
Mean Square Error (MSE)	3.91

Table 6.6: Reconstruction results for laboratory footage using the CH-SfM framework

As observed, the 3-D point cloud density is significantly higher than the methods proposed in Section 6.2.2 with a reconstruction density of $23.8025 N_X/cm^2$ with only 80 images. The reason for this is that the evaluated baseline between feature points is higher than the parameters used in Table 6.5, with the minMatches parameter also being increased to 300 over the previously evaluated 150 matches, meaning that the point cloud density would be significantly higher. As suggested by [57], one of the key principles for accurate triangulation is to increase the baseline of the matched images to ensure that rays used for determining the point of intersection come from wider angles to ensure that triangulation isn't prematurely determined by extremely close baseline shifts. By introducing an artificial baseline between images, this results in a

significantly stronger reconstruction with more 3-D points being able to be resolved in close proximity, in addition to reducing the resultant MSE error to 3.91.

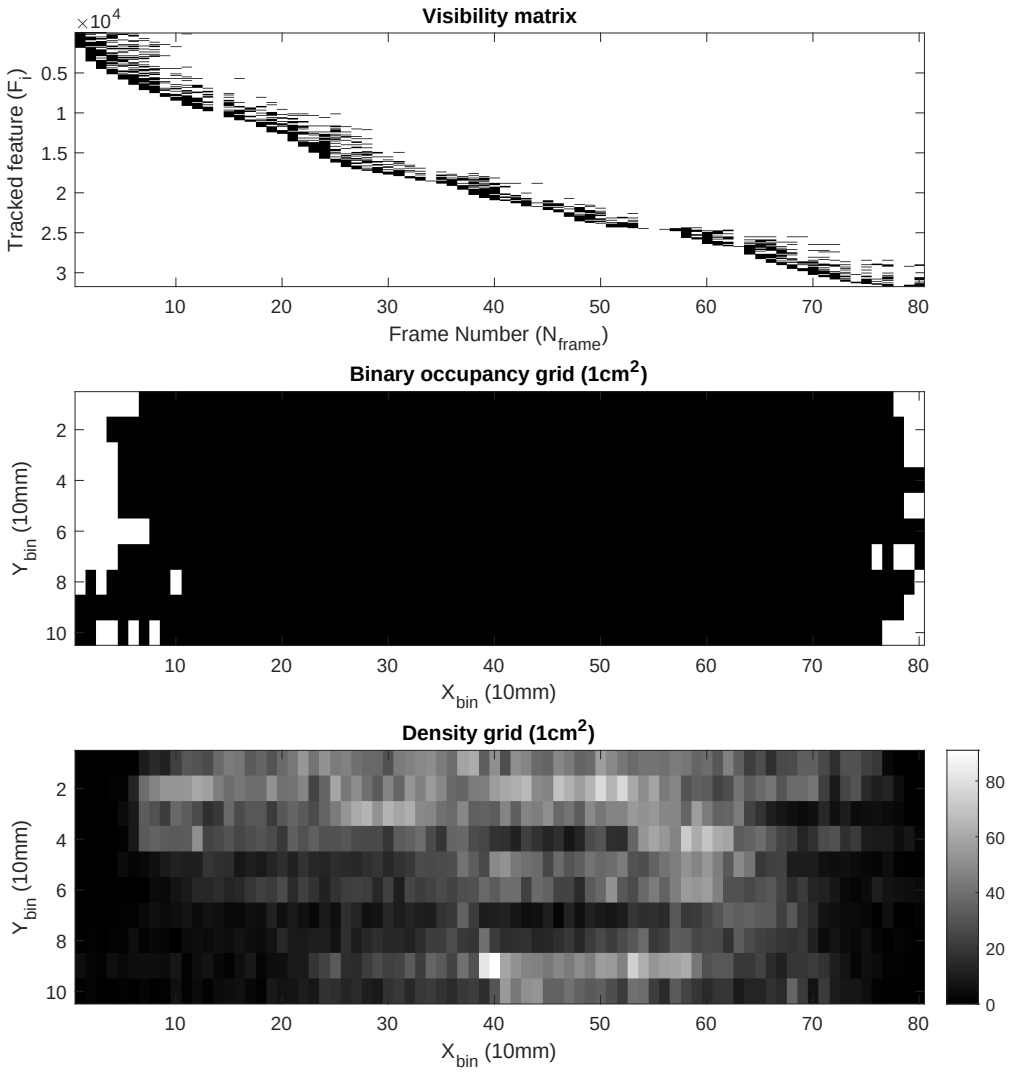


Figure 6.12: Feature visibility matrix, binary occupancy grid and density grid of the laboratory orientational 3-D reconstruction

6.3.2 In-core footage reconstructions

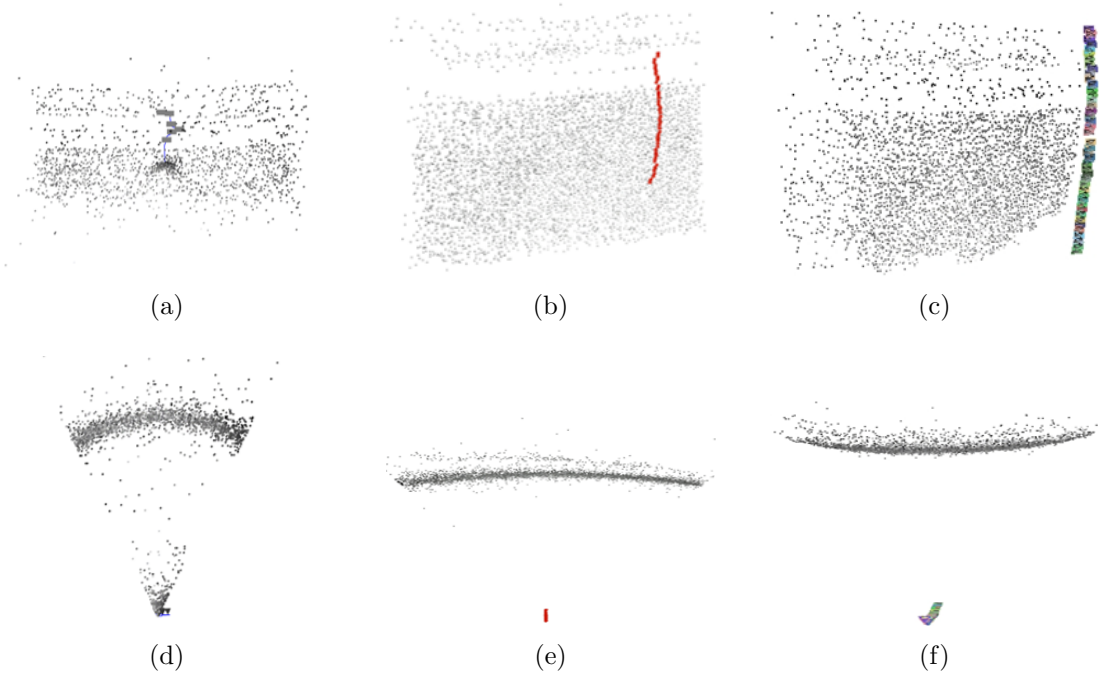


Figure 6.13: In-core reconstructions of a brick interface with challenging motion. a) & d) CH-SfM; b) & e) COLMAP; c) & f) VisualSFM

In-core reconstructions on a technical level are extremely challenging in comparison to reconstructions performed using laboratory footage since the general image fidelity is much poorer. As demonstrated in Table 6.1, the mid-brick layer images can be challenging due to the average entropic, variance and deviation of the imagery being considerably lower than laboratory footage. Consequently, the number of derived matches drops significantly as the entropy of the images decreases towards the lower stated bound of the framework. For the purposes of evaluation, two specific subsets were used to assess the framework. The first set of footage was based around a brick interface where there is a slight jolt in the camera movement. The framework however has a strong visible anchor point, the brick interface due to the dark contrast to the mid-brick layer texture, making it distinctive and easier to track. The second set of imagery focuses on the mid-brick layer footage. Performing mid-layer brick reconstructions are significantly more difficult and pose a different challenge to the brick

interface images due to the low contrast and texture of the imagery which results in a distinct lack of feature correspondences [44]. Due to the absence of camera calibration information and with no information with regards to the area of surveillance, 3-D reconstruction metrics cannot be reliably extracted and analysed in a manner similar to Section 6.3.1. Therefore, objective observation of in-core inspection reconstructions is the manner of determining the reconstruction accuracy. With regards to the later presented CRVW values, the brick start-of-life geometry is used to profile the accuracy. Consequently, reconstructions demonstrate high degrees of sparsity in addition to an abundance of noisy 3-D points. The results from both these image data-sets tested using all 3 frameworks can be observed in Figure 6.13 for the brick interface images and Figure 6.14 for the mid-brick layer images.

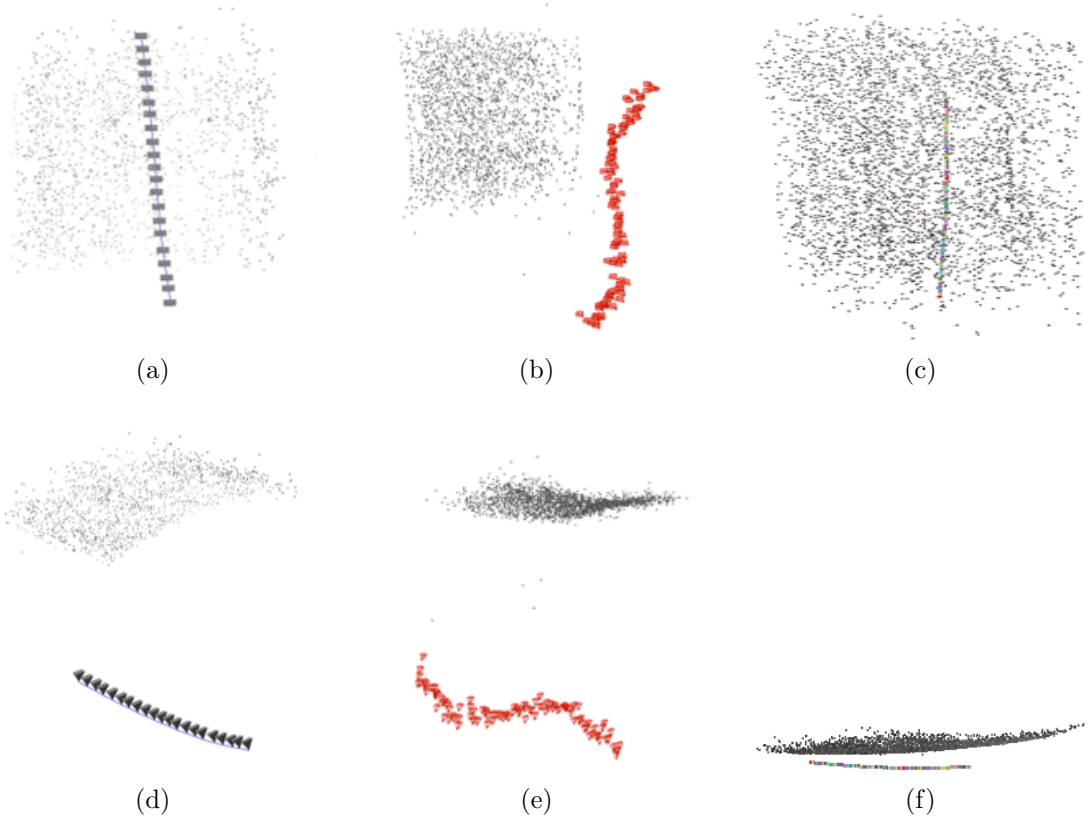


Figure 6.14: In-core reconstructions of a mid-brick layer with challenging feature space. a) & d) CH-SfM; b) & e) COLMAP; c) & f) VisualSFM

When applying reconstruction techniques to brick interface images, the COLMAP

and VisualSFM reconstruction frameworks results as seen in Figure 6.13b) and c) highlight an ability to easily reconstruct a point cloud due to the density of salient features available across all the images of the brick interface. In Figure 6.13e) and f) nonetheless, it can be observed that both COLMAP and VisualSFM struggle to determine the curvature of the observed brick interface even with a high number of detected features. Furthermore, inaccurate bending of the brick interface correlates to the reconstruction results observed using laboratory footage with these frameworks. In contrast, the proposed reconstruction framework is capable of accurately extracting this curvature of the brick interface at the expense of noisy 3-D points as visualized in Figure 6.13a) and d). In retrospect, by convolving a Gaussian filter with the images as presented in Section 6.2 would be highly likely to negate such noisy points. The resulting reconstructions contain 2,292, 4,809 and 3,885 3-D points for the proposed reconstruction framework, COLMAP and VisualSFM respectively with execution times of $53s$, $101s$ and $68s$ [44].

Reconstructions using the mid-brick layer footage presents a fundamentally different challenge. With a lack of features, correspondence searching becomes difficult and this is compounded at the model initialisation stage, which requires an excellent 2-pair reconstruction to ensure the entire reconstruction is accurate. As observed in Figure 6.14b) and c), both COLMAP and VisualSFM suffer from cumulative drift and do not exhibit the brick curvature observed in Figure 6.14a) or d) produced by the proposed framework. Furthermore, the resulting camera pose trajectory in Figure 6.14b) and e) reproduced by the COLMAP framework is erratic whereas the trajectories determined by VisualSFM and the proposed framework in Figure 6.14c) and f) respectively correctly exhibit a smooth vertical motion which is expected. Due to the nondescript nature of mid-brick layers, reconstruction density is inherently sparse; 1706, 2422 and 4319 3-D points for the proposed framework, COLMAP and VisualSFM respectively with execution times of $158s$, $74s$ and $167s$.

CRVW Reconstructions

CRVW footage, although restricted in capture volume, gives an insight to how reconstructions could be performed if the in-core source footage was of higher fidelity.

Subjectively, the CRVW footage can be classed as a medium quality footage, whereas in-core and laboratory NICIE footage is low quality footage and the camera footage obtained using the manually translated Toshiba T4KA7 camera sensor which has excellent low-light abilities as the high quality footage. With CRVW footage, the texture of the graphite is more prominent and distinct, ensuring that reconstructions can be produced with almost any framework. Once looking at the results however, the CRVW footage reconstructions suffer from the same symptoms that afflict the subjectively lower and higher quality image sources. Cumulative, incremental drift error is the most prominent error source which results in deformation of the 3-D reconstructions. Since the point cloud has minimal noise, it can be scaled to the start-of-life brick size, and used as a basis of approximating the reconstruction accuracy using in-core footage. The 3-D point cloud results from the proposed framework, COLMAP and VisualSFM can be observed in Figure 6.15 and Figure 6.16 in addition to tabulated results provided in Table 6.7 and the feature visibility matrix, occupancy grid and density grid of the reconstruction provided by the CH-SfM framework in Figure 6.17.

From the results in Figure 6.15, it is clear that from the top-down view from Figure 6.15a) and d), the proposed framework approximates the expected curvature of the brick and corresponding rigid body motion. The results from the COLMAP framework b) and VisualSFM c) however demonstrate a large degree of geometric deformity resulting in the reconstructions to bend. The extent of the bending of the reconstruction is conspicuous in Figure 6.16b) and c) where both COLMAP and VisualSFM produce 3-D point clouds that are very misrepresentative of the AGR fuel channel structure. Furthermore, although the associated camera pose estimation appears to tracks well in Figure 6.15e) and f), this is subverted in Figure 6.16b) and c) where the camera pose estimation can be viewed as wildly inaccurate. With regards to the reconstruction statistics, the reconstruction produced by the proposed framework only has 6,743 3-D points whereas COLMAP and VisualSFM comprise of 288,503 and 399,550 3-D points with execution times of $7m : 45s$, $102m : 53s$ and $105m : 27s$ respectively for each framework. With the proposed framework results in both Figure 6.15 and Figure 6.16, the reconstructions do suffer from minor incremental drift in addition to recon-

struction sparsity. However, as previously mentioned in prior subsections, the proposed framework is obtaining strong and reliable feature matches, resulting in improved reconstruction accuracy, which is critical within a nuclear domain [9, 44]. Consequently, the results produced by the proposed framework directly produces approximately 15% of the 3-D points ascertained by both COLMAP and VisualSFM. However, the framework also operates much quicker than the other two frameworks in addition to producing a representative sparse 3-D point cloud.

Reconstruction metrics	
E_{reproj}(px)	0.3101
N_X (Number of 3-D points)	6,743
Approximate 3-D point density (N_X/cm^2)	8.4288
Approximate Mean Square Error (AMSE)	4.6917

Table 6.7: Reconstruction results for CRVW footage

As observed in the associated metric approximations, the number of 3-D points detected drops considerably from previous examples, with the approximate point density being $8.42N_X/cm^2$ and the resulting Approximate Mean Square Error (AMSE) score is 4.69, considerably higher than the laboratory footage reconstruction provided in Section 6.3.1. It is clear that, although this is an excellent example of in-core reconstructions, the bowing of the CH-SfM framework results in a higher AMSE score. The CRVW footage, unlike other in-core footage, is not gathered in an circumferential manner and is only used to inspect potential cracks in fuel channels at the outer rim of the graphite core. In the following section, a newer module is introduced to allow for circumferential reconstructions to be attained using normal AGR in-core inspection footage.

6.4 Circumferential Reconstructions

Circumferential reconstructions are limited to in-core inspection footage that has been gathered from multiple orientations and require six orientational reconstructions in order to produce a full circumferential reconstruction. Since the functionality is inherently application-specific, the CH-SfM framework will be used in order to produce

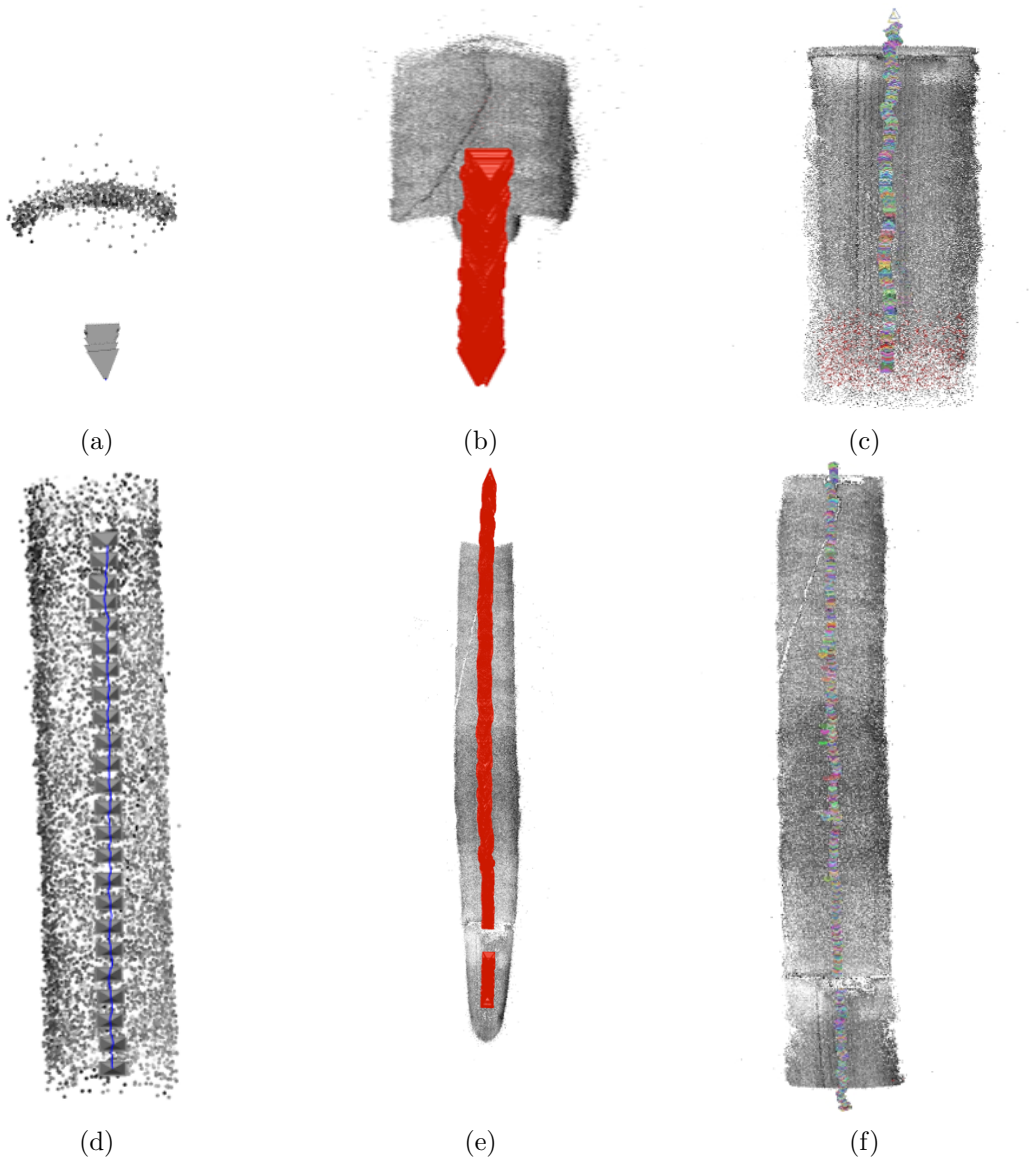
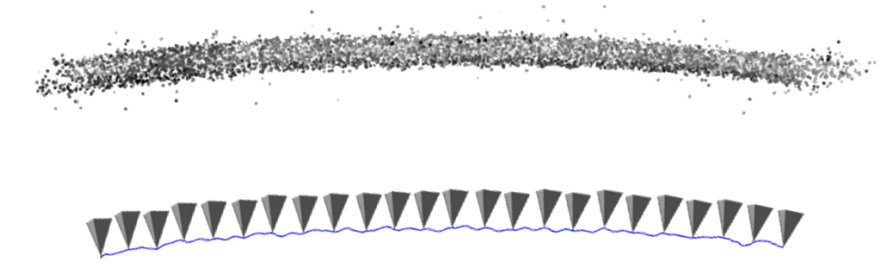
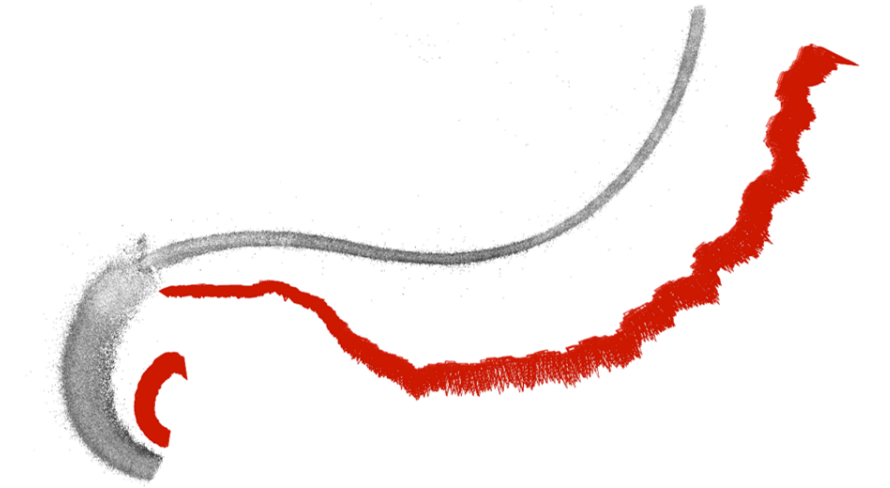


Figure 6.15: Top-down a)-c) and frontal view d)-f) of CRVW orientational reconstruction of a full AGR brick layer. a/d) The proposed framework reconstruction; b/e) COLMAP reconstruction; c/f) VisualSFM reconstruction.

these reconstructions [32]. Figure 6.18 shows the key steps to the refinement procedure applied to the point clouds, with the following sub-sections going into greater detail about each task.



(a)



(b)

(c)

Figure 6.16: Full side on view of CRVW orientational reconstruction of a full AGR brick layer a) The proposed framework reconstruction; b) COLMAP reconstruction; c) VisualSFM reconstruction.

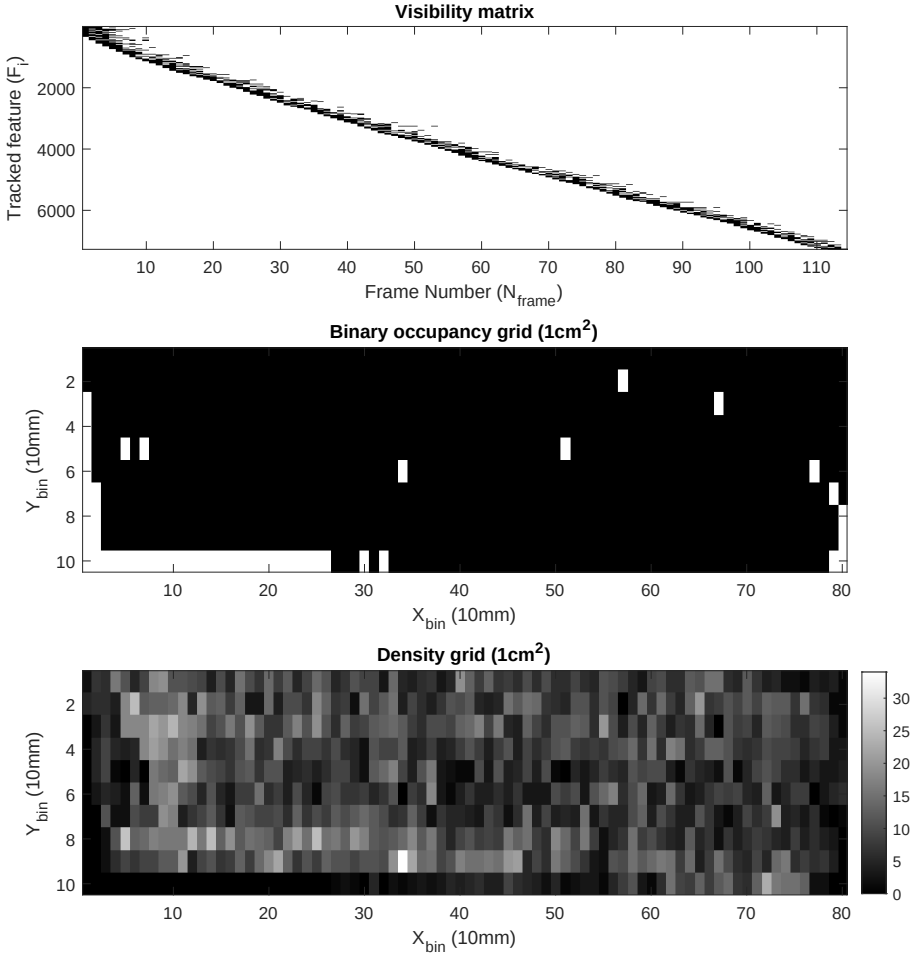


Figure 6.17: Feature visibility matrix, binary occupancy grid and density grid of the CRVW orientational 3-D reconstruction

6.4.1 Reconstruction refinement operation

The reconstruction refinement procedure is an addendum to the framework proposed in Chapter 5 and described functionally in Appendix A where it takes multiple orientation reconstructions and brings them together into a single point-cloud for further processing. Since each orientational scan may exhibit different feature characteristics from each other, some orientational scans may be affected by noisy 3-D outliers that need to be removed. This denoising procedure is initiated on the combined reconstruc-

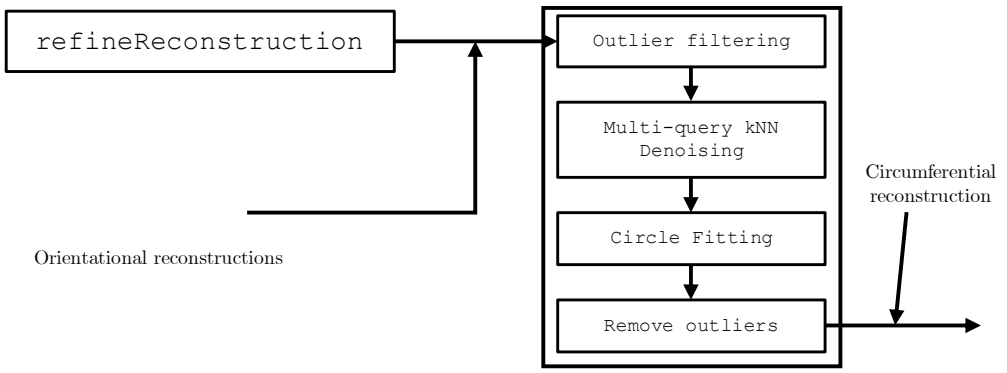


Figure 6.18: Block diagram of the core functions within the refineReconstruction function.

tion to exploit the approximate geometry of the channel. Due to the minimal overlap of neighbouring scans, trying to merge 3-D point clouds that pertain to each scan which all exhibit similar feature properties causes 3-D matching algorithms to simply overlap the orientational reconstructions on top of each other [32, 168]. To avoid this, since the orientation of each scan is known beforehand, each orientational point cloud p_θ and the associated 3-D points $\mathbf{X} \equiv (\acute{x}, y, \acute{z})^T$ are rotated around the y-axis where $\acute{x} = x\cos\theta + z\sin\theta$ and $\acute{z} = z\cos\theta - x\sin\theta$ and merged into a singular 3-D point cloud. Due to the nature of reconstruction, the orientations used for manoeuvring the reconstructions requires manual intervention. A typical result of this can be seen in Figure 6.19a).

Outlier filtering and denoising

Since the reconstructions are metric and initialised at a depth of 1, the first stage is to remove gross outliers from the point cloud model due to the triangulation procedure struggling to obtain the correct depth. This is performed by filtering out 3-D points using the Median Absolute Deviation (MAD). The second stage is to apply 3-D specific denoising by applying multi-query k-Nearest Neighbours (kNN) denoising, which exhaustively calculates the mean distance of 3-D points to all neighbouring points and

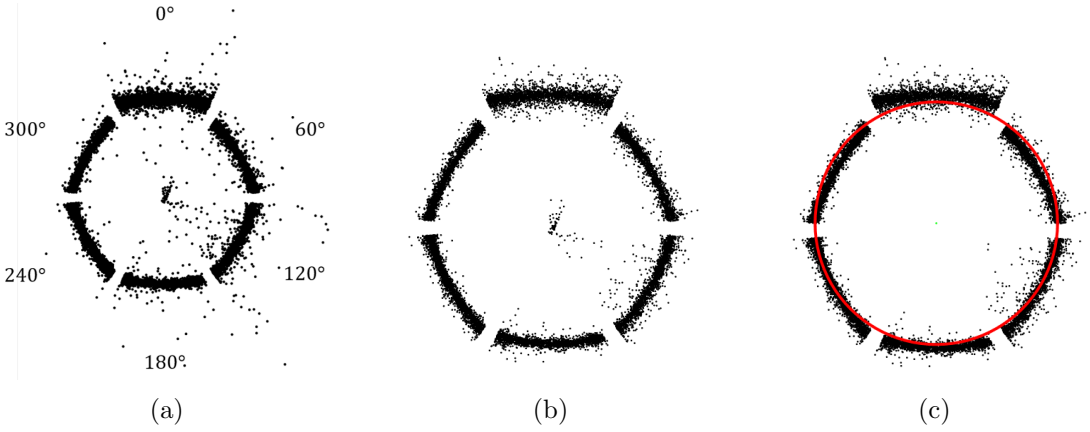


Figure 6.19: The typical reconstruction refinement processes at various stages of processing. a) The unrefined circumferential point cloud from a top-down view; b) A point cloud denoised by multi-query kNN denoising; c) Further point-cloud denoising using circle fitting

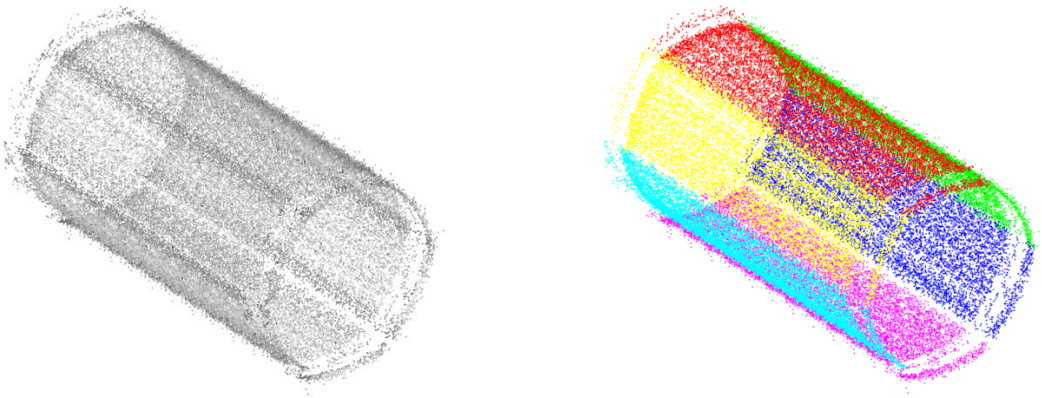
detects outliers that exceed a specified distance threshold [169]. The typical result of the previous outlier filtering and the denoising process can be observed in Figure 6.19b).

Producing circumferential reconstructions

Once the point-clouds have been sufficiently pruned of outliers, the next stage is to remove outliers that can exist near the centroid of the cylinder as a result of triangulation of weak feature points that can be observed in Figure 6.19a) and in-core reconstruction results as seen in Figure 6.13d). Since the general shape of the AGR fuel channel is known (i.e. near cylindrical), outliers can be removed by exploiting *a-priori* knowledge about the channel structure by evaluating the channel cross-section. Furthermore, the dimensionality of the problem can be reduced by performing 2-D circle fitting on the $\acute{x}$ and $\acute{z}$ co-ordinates. To do this, the method proposed by Pratt [170] incorporates the use of a robust circle fitting method based on direct least-squares to further remove outliers and produce a clean cylindrical 3-D point cloud.

The final stage is to effectively bring the orientational reconstructions together toward the centroid and align them to produce a cylindrical, circumferential reconstruction. At this stage of implementation, merging the individual orientational reconstructions is a purely manual process that involves shifting the co-ordinates of the

respective point clouds to bring them together into a singular, cohesive 3-D point cloud. The results of a single and multi-layered circumferential reconstruction of AGR graphite bricks can be observed in Figure 6.20 and Figure 6.22 respectively.



(a) (b)

Figure 6.20: Circumferential, geometrically constrained reconstruction of a single AGR brick layer using in-core footage. a) Colour representative point-cloud; b) Orientation-based coloured point cloud - each colour representing a different orientation.

6.4.2 Circumferential reconstructions using in-core footage

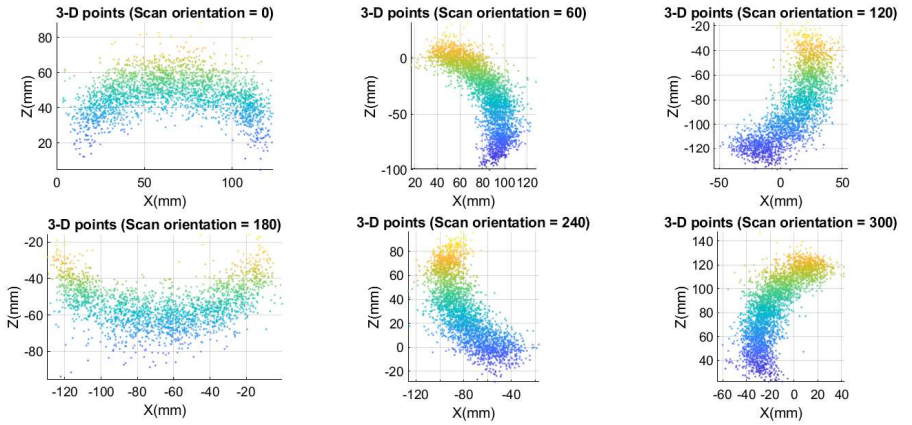
Circumferential reconstructions using in-core footage is an on-going challenge and is an area for future work. The fundamental reason for this is that the reconstruction mechanism can be inconsistent with reconstruction quality due to the inspection protocol sometimes changing between different orientations. This results in the underlying parameters used for reconstruction having to be manually tweaked and re-run in order to obtain a representative reconstruction. To produce representative circumferential reconstructions, considerable manual intervention is required, with geometric constraints often being used to flatten the point cloud surface as observed in Figure 6.20. In this section, the original source reconstructions from Figure 6.20 are used and the approximate density and AMSE is used by rescaling the reconstruction to approximate the expected structure size. As demonstrated in Figure 6.18, the requirement is to produce 3-D point clouds of every assessed orientation and rotate it based on the acquisition of said rotation - the result of this is shown in Figure 6.21a) (significant outliers have

p_θ	N_X	Approximate 3-D point density (N_X/cm^2)	Approximate Mean Square Error (AMSE)
0	2506	3.1325	5.0115
60	3342	4.1775	
120	2785	3.4813	
180	2228	2.7850	
240	3063	3.8288	
300	3481	4.3513	

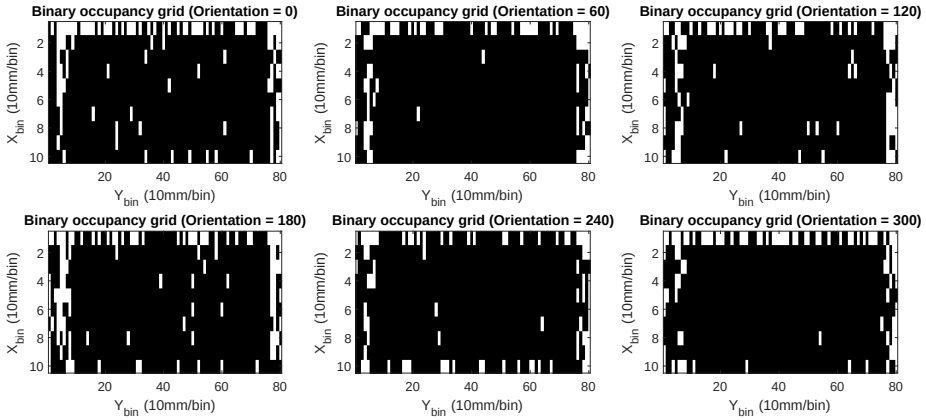
Table 6.8: Circumferential reconstruction results produced at each evaluated orientation

been removed for visualisation purposes). As shown in Table 6.8, the derived number of 3-D points for each in-core orientation reconstruction can vary significantly, in this case from 2,228 points to 3,481 points and this effects the underlying point density. Consequently, the point density varies from $2.7850N_X/cm^2$ to $4.3513N_X/cm^2$, in a range much lower than that extracted from in-core CRVW footage and laboratory footage. Consequently, the AMSE score is also higher than previously evaluated reconstructions, with an AMSE score of 5.0115 and this can be viewed in Figure 6.21a) where the point cloud cross-section shows deviation in the derived depth.

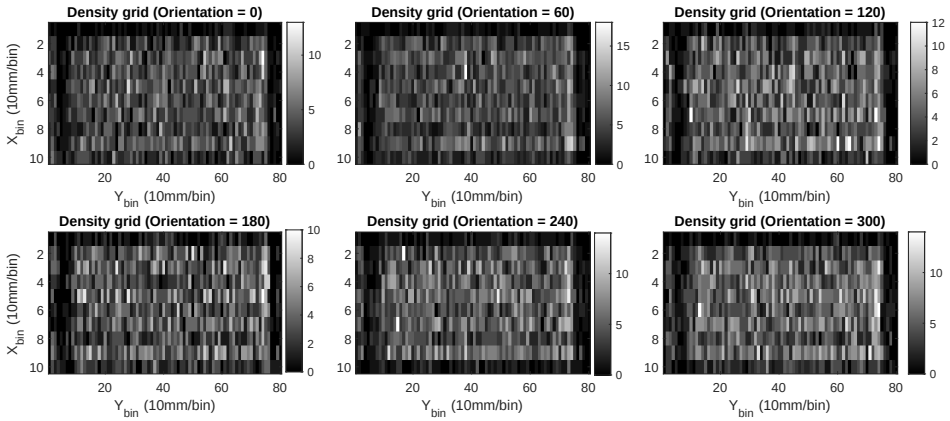
The occupancy grids shown in Figure 6.21b) also demonstrate the difficulties in some scans to acquire point reconstructions on the outer edges of the images and this is due to the uniform illumination and lens distortion of the camera (shown in Figure 2.7) making it increasingly difficult to find points, even with MIFM matching in place. Interestingly, the brick interfaces prove challenging as visualised on the far left and far right regions of the occupancy grids in 6.21b) where the reconstruction fails to find points in the dark regions, indicating these areas may not be as stable for reconstructions as previously thought. The density grids provided in Figure 6.21c) further indicate the aforementioned inadequacies with many of the derived 3-D point locations residing in the center of the assessed images.



(a)

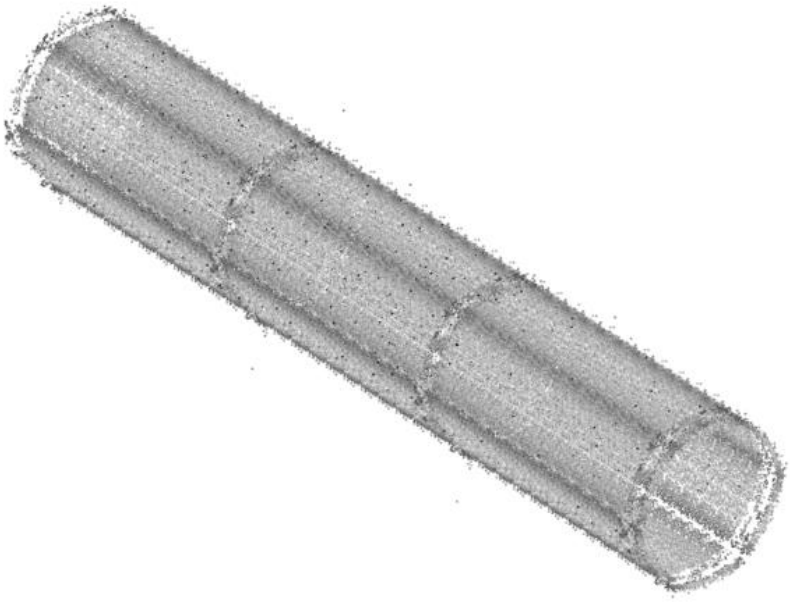


(b)

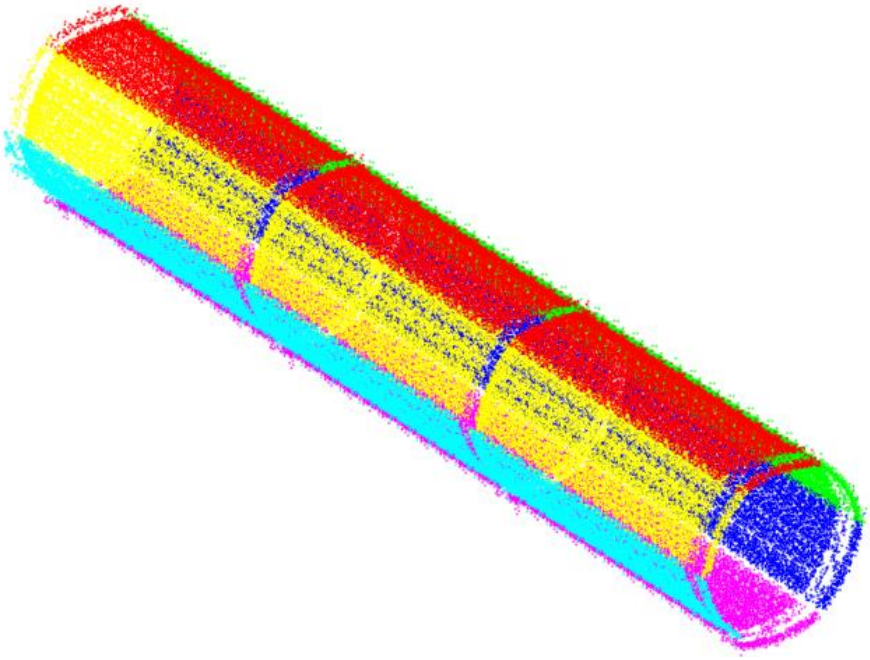


(c)

Figure 6.21: Reconstruction results of AGR in-core footage from each assessed orientation; a) 3-D point cloud reconstructions; b) Binary occupancy grids; c) Density grids for each θ value.



(a)



(b)

Figure 6.22: Circumferential reconstruction of a three AGR bricks using in-core footage. a) Colour representative point-cloud; b) Orientation-based coloured point cloud - each colour representing a different orientation.

6.5 Discussion

This chapter has introduced the application basis for this work which is producing a 3-D visualisation of AGR fuel channels, with the work performed in this chapter representing the first time that 3-D data has been extracted from an AGR fuel channel using only image data. Due to the variation of reactor imagery that can be acquired, it is important to establish what are the practical limits to applying the CH-SfM to real in-core inspection imagery. To ascertain this, the chapter begins by introducing a distortion analysis aimed at defining the functional limits of the proposed CH-SfM framework and the underlying MIFM approach with regards to the imagery on which it operates. This is quantified through the usage of 3 distinct texture analysis metrics; Entropy, Variance and Standard Deviation. The motivation for performing this analysis is to quantify the point of failure, an evaluation that very little reconstruction frameworks in literature perform. By providing this quantification, it potentially allows for images to be passed through a pre-assessment procedure using the texture-analysis metrics, and identify potential images that may cause issues for the framework and filter them accordingly. To test the MIFM approach, the matching procedure is evaluated using mid-brick layer imagery from bespoke laboratory apparatus and in-core inspection footage convolved with a Gaussian filter with an incrementally higher σ value, which results in the image being increasingly blurred, with the entropy, variance and standard deviation of the images captured at each stage. The proposed matching procedure MIFM is outperformed by a similar rival, the GMS algorithm, by on average 53.73% on the higher quality laboratory imagery. However, on the challenging in-core imagery, the MIFM algorithm extracts more than double the number of correct feature matches produced by the GMS algorithm, exhibiting a 138.53% performance improvement in match density. The MIFM approach demonstrates excellent matching capabilities even when the underlying imagery is strongly blurred, supporting the novel thesis presented in Chapter 4 and its potential applicability across not only AGR imagery, but imagery acquired in different and more challenging constrained environments. It is important to note however, that the reconstruction density was evaluated

using the matches obtained using the MIFM approach, and it became clear that the resulting matches suffered in quality as σ increases, with the resulting reconstructions being distorted as a result of the underlying depth cues being lost during blurring. This established effectively a lower bound of the framework accuracy, requiring an average image entropy of approximately 6.4624 in order to operate effectively.

Metric	**Laboratory footage**	**CRVW (In-core)**	**NICIE2/CBIU (In-core)**
3-D point density (N_X/cm^2)	23.8025	8.4288	2.785 - 4.3513
Mean Square Error (MSE)	3.91	4.6917	5.0115

Table 6.9: Reconstruction metric scores across the varying levels of footage quality assessed using the CH-SfM framework. Note that both the CRVW and NICIE/CBIU scores are approximate values.

An evaluation was then performed by comparing the CH-SfM framework against the state-of-the-art incremental frameworks COLMAP and VisualSFM, with the CH-SfM framework demonstrating its capability of extracting improved representative channel geometry over both COLMAP and VisualSFM, despite having considerably less dense point clouds. It is important to note that during the constrained iterative procedure within the MIFM method could produce more matches if the higher bound is increased. However, by increasing the number of possible matches, there is a considerable chance of noise being induced into the matching and subsequent reconstruction, and therefore a balance must be maintained. To quantify the performance of the CH-SfM framework, the 3-D point density per cm^2 and the Mean Square Error from a cylinder fitting procedure using brick dimensions is used to quantify the reconstruction density and error. In Table 6.9, the results produced using different sources of footage. As observed, as the footage quality is reduced, this is reflected in a sparser reconstructed 3-D point cloud density and a higher MSE/AMSE score. With the laboratory footage being acquiring at a significantly higher resolution, the resulting 3-D point cloud density is significantly higher than the in-core inspection imagery, which is not only a lower resolution but a lower frame rate. The reconstruction density of the 3-D point clouds produced from laboratory footage would be suitable for further investigation of channel defects. However, the in-core footage reconstruction density, even with the MIFM matching scheme, struggles to pull out enough features in challenging areas,

making it potentially unsuitable for defect inspection. On the other hand, the reconstruction density could be vastly improved across all footage if the usage of MVS was augmented onto the system using an approach described by Locher et al [171] to obtain photorealistic reconstructions using the underlying sparse point cloud produced by the CH-SfM framework. However, this is an area of future work. The capability of producing circumferential reconstructions, a feature unique to the CH-SfM framework is demonstrated in Chapter 6.4, where orientational reconstructions produced can be effectively combined into a singular reconstruction, where it can be effectively filtered, manually manipulated and constrained to produce a representation similar to that of Figure 6.20 and 6.22. The current procedure however takes considerable manual intervention in order to accurately reconstruct the cylindrical shape of the system and producing this autonomously is also an area of future work.

Chapter 7

Conclusions

7.1 Summary and discussion of contributions

This thesis has presented a novel approach to performing Structure-from-Motion in deficient feature spaces. It augments existing approaches in 2 key areas and demonstrates the novelty on an application for inspection of AGR fuel channels. The application domain is introduced in both Chapter 1 and Chapter 2, and why performing 3-D reconstructions of the AGR fuel channel can be beneficial for reasons of prognosis and diagnosis of potential and pre-existing structural defects initiated by reactor operation in addition to discussing the limitations and challenges associated with performing 3-D reconstructions of the AGR channel. A review of the state-of-the-art in image-based 3-D reconstruction is presented with an emphasis on Structure-from-Motion, with a preliminary evaluation of pre-existing closed and open-source SfM frameworks being performed to determine suitability. The key outcome is that generic SfM approaches were insufficient, in addition to many techniques relying on additional hardware and sensory data in order to reduce the complexity of the 3-D reconstructive process such as laser profiling approaches such as LiDAR or integration of *a-priori* geometry. With the costs associated with adopting new hardware into the AGR fuel channel being cost-prohibitive, this drove the need to explore image only based reconstruction approaches.

The challenges of performing SfM in challenging, feature-poor environments resulting in correspondingly deficient feature representation are presented. Feature homogeneity is highlighted as a specific challenge, as the 3-D reconstruction process is heavily reliant on the fidelity of the derived features extracted from the images used for reconstruction. These limitations were identified, with a novel feature-based approach called Motion-based Iterative Feature Matching (MIFM) being proposed, incorporating

the global descriptive power of indirect neighbourhood-based statistic feature descriptors and the direct motion estimation to derive stronger matches, even in extremely challenging environments. The approach was evaluated against a similar state-of-the-art matching method by incrementally distorting in-core inspection imagery, with the MIFM method attaining a 138.53% performance improvement in match density. The proposed MIFM method was integrated into a novel incremental SfM framework called the Constrained Homogeneous Structure-from-Motion (CH-SfM) framework, which can not only operate in general environments, but also contains novel application-specific reconstruction algorithms to refine and produce circumferential representations of the AGR fuel channels. The CH-SfM framework is verified initially using synthetic ground-truth data, in addition to being comparatively analysed against two other state-of-the-art incremental SfM approaches using both in-core RVI footage and laboratory footage gathered from a bespoke experimental apparatus emulating a sub-section of the AGR fuel channel. It is demonstrated that the CH-SfM framework developed extracts structurally representative geometry from the fuel channel from both sets of footage, whereas the state-of-the-art methods produce deformed or misrepresentative point-clouds of the AGR fuel channels. The CH-SfM framework achieves an approximate 3-D point density of 2.785 - 23.8025N_X/cm^2 and an approximate MSE of 5.0115 and 3.91 for real in-core inspection footage and high quality laboratory footage respectively, indicating its usability for structural assessment of the AGR fuel channel providing that the associated image quality is sufficient.

7.2 Future Work

There are a number of approaches proposed for extending the work proposed in this thesis:

- Reduce computational complexity of the MIFM approach.
- Preprocessing of input images for CH-SfM framework to remove weak imagery.
- Postprocessing of 3-D sparse point cloud using Multi-View Stereo (MVS) to pro-

duce photo-realistic 3-D reconstructions.

- Introduce a hierarchical SfM approach to further reduce incremental drift.

7.2.1 Framework improvements and considerations

MIFM and CH-SfM limitations

Disambiguation of repetitive elements within the imagery is a significant challenge with literature. In Chapter 4.1 - 4.2, image homogeneity and matching ambiguity is defined as a considerable problem for applications that require accurate image correspondence. The MIFM approach proposed within Chapter 4.3, although performs significantly stronger than state-of-the-art approaches as shown in Chapter 6.2 when processing challenging imagery, still has ways in which it can be improved. First of all, the MIFM approach takes a considerable amount of computation since it extracts both sparse and dense representations of the image. To accelerate this procedure, GPU processing techniques [25,113] for both the SIFT and Dual TV-l1 approaches can be deployed to allow for real-time performance. Secondly, the MIFM approach relies on the imagery being captured contiguously to extract robust feature matches to obtain highly accurate feature matches due to the dense optical flow approaches used for the DMS procedure. On high quality contiguous imagery, similar state-of-the-art matching methods outperform MIFM as seen in Figure 6.6 and further work should be implemented at the constrained iterative stage of the MIFM approach to reliably increase match density. Another point to make with regards to MIFM is that as shown in Chapter 6.2, applying a Gaussian filter with a standard deviation distribution value of $\sigma = [0.8, 2.2]$ seems to actually improve the matching performance and this warrants further investigation.

The CH-SfM framework proposed in Chapter 5 and evaluated in Chapter 6 also warrants additional future work, particularly at the pre-processing stage. One of the biggest issues is the introduction of weak or poor images into a reconstruction can also drastically affect the end result if these images make up a considerable portion of the data. Weak image in this context pertains to images that are subjected to motion blur, extreme intra image under/overexposure due to non-uniform illumination

of the inspection tool or defocus aberration during the AGR fuel channel inspection. Consequently, if the image is weak, the textural component of the image is subdued and performing pairwise matching with said images will result in the reconstruction framework being able to discern the depth as discussed in Chapter 6.2.2. To solve this issue, a pre-emptive assessment strategy of the input imagery is proposed for future work, where techniques derived for Shape-from-focus can be used to detect blurry images such as the reduced DCT energy ratio [172], diagonal laplacian [173], steerable filters [174] or even common place methods such as the Fast Fourier Transform (FFT) or Laplacian of Gaussian (LoG) [43]. This approach, in combination with a next best view strategy [21] and assessment of the image entropy at the reconstruction stage would be highly advantageous at removing imagery which challenge the correspondence searching and reconstruction stages, allowing for weak images to be detected and removed.

At the reconstruction stage, the CH-SfM framework proves to be considerably more robust than the state-of-the-art at producing representative scene geometry by searching for accurate feature matches at the expense of matching density. As a result, the reconstruction density produced by the CH-SfM framework is observably much lower than other evaluated approaches, with a reconstruction density of 2.785 - 23.8025N_X/cm^2 for low quality in-core inspection footage and high resolution laboratory footage respectively. At a density of 2.785N_X/cm^2 for the most challenging of in-core footage, the reconstruction is simply not dense enough in its own right to determine accurately structural defects within the AGR fuel channels. The CRVW footage, with an approximate density of 8.4288N_X/cm^2, indicates the higher bound of in-core footage reconstruction quality that can be obtained at this moment in time. Due to the sparse nature of the reconstructions, having a extremely dense sparse point-cloud is somewhat redundant, as Multi-View Stereo (MVS) approaches can utilise the underlying sparse reconstruction to model a fully dense, accurate photo-realistic reconstruction of the target scene providing the sparse point cloud is representative of the scene geometry. In future work, it is suggested that a progressive dense reconstruction method as proposed in Locher et al [171] be used to progressively densify the reconstruction.

SfM - A Hierarchical approach

The CH-SfM framework, even with the MIFM approach deriving more accurate reconstructions, is still subject to incremental drift in the reconstruction. Without time constraints, the entire reconstruction framework would have been redeveloped from the ground-up in C++, incorporating a hierarchical reconstruction approach [96] where each brick-layer is clustered into an atomic model and incrementally reconstructed within each said cluster [99]. This would have been done by utilising an image retrieval pre-processing approach where it would sequentially search for brick interface images then subsequently partition the RVI image stills into sub-sets pertaining to each layer in a manner diagrammatically visualised in Figure 7.1. This could be combined with the aforementioned weak image detection approach previously described, as the brick interface images would be easily identifiable due to the high number of distinct features available in the images, to effectively split the image datasets autonomously. By breaking up the reconstruction procedure, the incremental drift error introduced is effectively compartmentalised across the entirety of the reconstruction, allowing for the reduction of error drift within the resulting model. Furthermore, this opens up the CH-SfM framework to being able to assess multiple layers in parallel, vastly speeding up the reconstruction execution time.

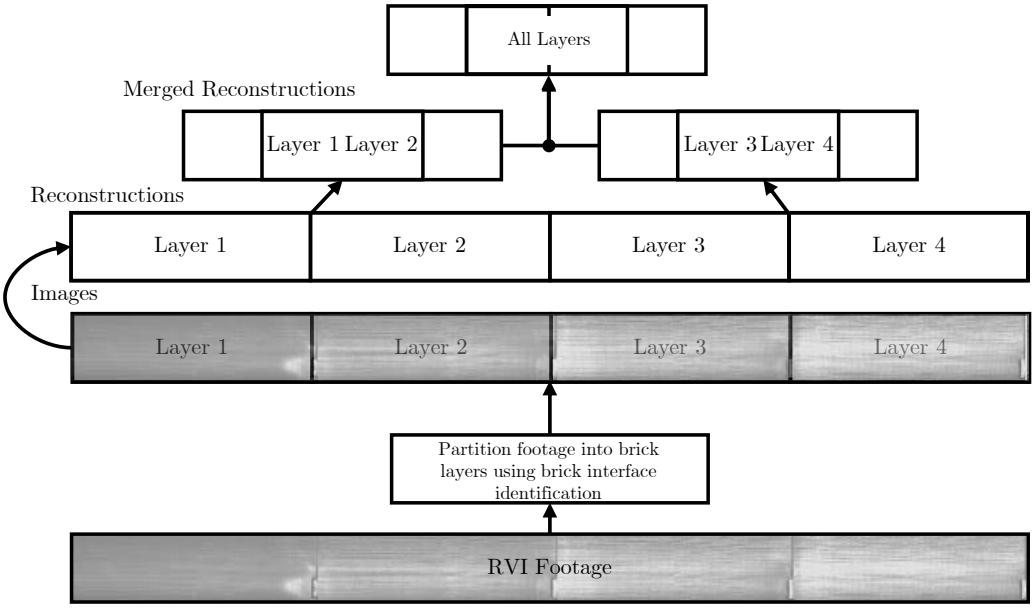


Figure 7.1: Hierarchical reconstruction approach for cyclic environments using four layers as an example.

With the footage being captured in a sequential manner, graph optimisation algorithms or query expansion techniques would be nonsensical - however by partitioning the reconstruction process, reconstructions in predefined cyclic environments as observed in the AGR fuel channel would also remove issues such as the brick layers being reconstructed as the same layer over and over again as observed in Figure 3.8 whilst using VisualSFM [25]. This would mean that the reconstructions can simply be appended onto one another, in addition to allowing inspection engineers to evaluate particular bricks that warrant further investigation.

7.2.2 Conclusion

In conclusion, the work produced in this thesis demonstrates capabilities of being able to accurately reconstruct 3-D point clouds of AGR fuel channel interiors that can be used for the purposes of diagnosis and prognosis using pre-existing inspection footage. The novel image correspondence matching scheme and reconstruction framework proposed within this thesis outperform state-of-the-art approaches, with future potential to enhance the capabilities and the technology readiness of the proposed approach. The

approaches demonstrate that extraction of representative geometry can be applied to passive sensory data successfully, and indicate a way forward for methods inspecting constrained environments where active sensory approaches cannot be applied. With the energy demands of the world increasing, the need for improved inspection methodologies in harsh environments such as nuclear reactors will become increasingly necessary.

Appendices

Appendix A

CH-SfM User Guide

The purpose of this section is to provide a low-level user guide to the operation of the 3-D reconstruction framework. In Chapter 5, the incremental reconstruction element of the framework is validated using synthetic data, and in Chapter 6, the CH-SfM framework is assessed against other state-of-the-art incremental reconstruction frameworks using AGR RVI imagery.

A.1 Framework operation - an overview

The framework described in this chapter was fully implemented within `MATLAB` in the associated language. The folder structure for the toolbox contains three core folders - "FUNCTIONS", "INPUT DATA" and "RESULTS". To initialise the framework, the user must first execute `setup_framework.m` at the top level of the FUNCTIONS folder. The purpose of this function is two-fold; the first is to add all the necessary folders to the `MATLAB` path to allow for execution of functions located in a variety of locations and the second is to setup and compile external dependencies. The framework has the following dependencies:

1. `MATLAB` Image Processing Toolbox is required throughout the framework and leans heavily on many of these toolbox functions.

2. VLFeat developed by Vedaldi et al is required for the SIFT feature descriptor and SLIC [134, 175].

3. OpenCV [176] and an associated `MATLAB` wrapper called "mexopencv" [177]

4. VLG Toolbox developed by Taehee Lee [178] for the bundle adjustment and visualisation features of the toolbox.

These dependencies are automatically compiled by the script for the user however for mexopencv to operate, the file location of OpenCV pre-compiled binaries must be specified. To begin operating the framework, the user must place a folder containing the input images that must be processed into the INPUT DATA folder. In the proceeding section, each component of the framework will be simplified into four key components - Input/Pre-processing, correspondence searching, incremental reconstruction and reconstruction refinement. Each module operates in a sequential fashion, requiring the inputs of the previous module in order to continue. A simple block diagram demonstrating the framework structure can be observed in Figure A.1.

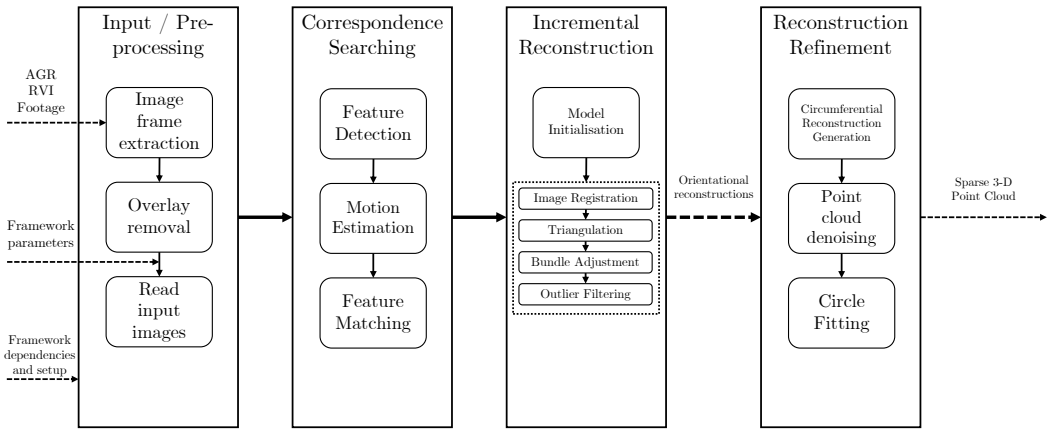


Figure A.1: Simplified block diagram of the framework operation

A.2 Input and Pre-processing

Before the framework is initialised, there is several pre-processing mechanisms required, especially if AGR imagery is utilised. Firstly, the framework operates predominantly on ordered images which means that within the folder that the framework is reading from, the images must be already ordered corresponding to the capture method. Each AGR RVI video must be processed by an video encoder such as FFMPEG into individual frames, and this is done by simply allowing naming the frames by the associated extracted frame number. The second step is to ensure that there is no overlays or constant scene features (i.e. a part of the inspection tool protruding into the camera

images) that could be misconstrued as a permanent member of the rigid scene. With the AGR footage, there is a text-based overlay used to show the current time, the current tool-height and the station at which the footage is captured at. Furthermore, in some footage, there can be FoV guides protruding at the top of the footage which both have to be cropped out which results in a smaller image.

```
%% General matching parameters
mC.seqMatch = 6;            % Window of sequential matched images
mC.minMatches = 150;        % Mininum number of initial matches
mC.Shift = [15,100];        % Shift parameters for translation filter
mC.RSNCiter = 100;          % Number of RANSAC iterations
mC.gInliers = 30;           % Mininum geometrically verified inliers
mC.bline = 3;               % Baseline between images
mC.minViews = 2;            % Mininum number of feature observations

%% Deficient feature space specific matching criterion
mC.kMatch = 1;              % 1 - Apply trans. filter/ 0 - Don't apply
mC.initThresh = 1.8;        % Initial matching threshold
mC.threshstep = 0.2;        % Step-size when iterating for better matches
mC.maxIter = 5;             % Max amount of matching iterations
mC.DMSinit = 0.75;          % DMS enabled if minMatches*DMSinit > matches
mC.seg = 1;                 % 1 - SLIC seg. / 0 - Spatial seg.
mC.SLICregsize = 50;        % SLIC region size
mC.SLICregulariser = 0.1;   % SLIC regulariser size
mC.spatialBins = 100;       % Number of spatial bins for spatial seg.
mC.DMS = 1 ;                % 1 - DMS enabled / 0 - DMS disabled
mC.est_motion = [0,0];      % Initialised estimated motion parameter
mC.percentile = 20;         % Range percentile for region strength ident.
mC.rejperc = 5;             % Rej. matches based on match score
```

Figure A.2: Code listing containing general matching and deficient feature space matching parameters that must be selected before initialisation. Each parameter is explained briefly using the MATLAB comment syntax next to each mC parameter.

The next stage to the framework is the specification of the input parameters which are contained within a structure (struct) variable called mC which stands for "matchingCriterion". Within the mC structure, there is two effective groupings of the parameters; the general matching parameters and the low-feature space matching parameters. In this section, these parameters and their effect will be discussed:

A.2.1 General matching Parameters

mC.seqMatch The sequential matching parameter is used to specify the window size of neighbouring images used during the matching scheme. This approach is the most effective for ordered image data-sets since many scene points (depending on the rigid body motion of the camera and the observed environment) will only be observed within a number of frames. By introducing a restrictive, sequential matching scheme, it will also reduce matching ambiguity if the observed environment is inherently cyclic or contains repetitive elements that would be introduced using other forms of matching such as exhaustive. Therefore, if `mC.seqMatch` > 1 and is say for example, `mC.seqMatch`=6, then I_1 will be matched to every image until I_7 providing there is enough images. If `mC.seqMatch` $== 1$, then pairwise matching will be executed. If pairwise matching is executed, care must be taken to ensure `mC.minViews` ≤ 2.

mC.minMatches During the correspondence stage, each image is matched initially to each other using a matching threshold of `mC.initThresh`. This value is usually set high to ensure that only very strong matches are incorporated at the initial matching stage. This number should be manually tweaked based on the input image data set. If the image dataset is high in quality, then this variable have very little to no effect. If the input image dataset is very challenging then `mC.minMatches` should be set lower. During the iterative matching scheme, as the threshold is continually relaxed, the number of matches currently obtained must match or exceed `mC.minMatches`.

mC.Shift The shift is a 2-vector which contains the estimated horizontal and vertical motion component expressed in pixels to be used with the translation filter

mC.RSNCiter This parameter is used whilst removing outliers through the use of the epipolar constraint. This parameter directly relates to the number of iterations undertaken to try and produce the consensus of feature matches with the lowest Sampson error - which in itself is a first order approximation of the geometric distance - given a pair of normalised feature point correspondences and the determination of the Fundamental matrix $\mathbf{F}$ using the 8-point algorithm [66]. Increased number of iterations will likely result in a better result, but will increase computation. Conversely, if the number of set iterations is low, it will be quick to execute but mismatches may end up being included.

mC.gInliers This parameter is used to effectively constrain ambiguous images and their corresponding feature points. This ensures that only feature match sets that exceed the number of `mC.gInliers` are accepted and subsequently added to the detected feature tracks defined by the visibility matrix. This parameter is helpful when a blurry or very poor image is introduced to the dataset and as a result, has minimal geometrically verified matches.

mC.bline This parameter is utilised right at the start of the framework and is used to induce an artificial baseline between images by simply taking every $\texttt{mC.bline}^{th}$ image in the dataset. This is particular useful if the rigid-body motion of the capture device is small and therefore minimises problems induced by narrow-baseline matching such as depth computation which is inherently sensitive to image coordinates precision for viewpoints acquired that have very high degrees of visual overlap. Furthermore, by inducing a baseline, redundant images can be removed - increasing computation whilst having minimal effect on the resultant reconstruction.

mC.minViews The last general matching parameter is `mC.minViews` which is

simply used to retain feature points that are observed in set number of images as a minimum requirement. By increasing this parameter, the feature points will be increasingly robust however this may result in reduced reconstruction density.

A.2.2 Deficient feature space matching parameters

mC.kMatch This parameter is a binary flag that enables knowledge-based matching (i.e. the use of the translation filter) if the value is '1'. The translation filter will utilise the estimated 2-D motion parameters specified in `mC.Shift` that have been manually entered by the user. If the value is '0', then no knowledge-based matching is implemented.

mC.initThresh By defining `mC.initThresh`, this value is the matching distance threshold used by Lowes ratio test [60] to remove ambiguous matches. The higher this threshold value is, the stronger and more selective the matching procedure is. Conversely, if the initial threshold specified by `mC.initThresh` is low, this will increase the chances of inducing noisy or erroneous matches into the reconstruction at the initial stage.

mC.threshstep During the iterative matching scheme, the threshold used to match features will be continually relaxed by a step-size defined in `mC.threshstep`. Setting this as a relatively large number ensures faster convergence on satisfying the `mC.minMatches` criteria but may introduce unnecessary or noisy points whereas making the value too small may result in negligible differences in the matching set whilst increasing the computation time required for matching.

mC.maxIter This parameter is used to define the maximum number of iterations used during the matching procedure whilst relaxing

the threshold. This is primarily used to minimise computation whilst also ensuring that the matching threshold does not drop too low, ergo introducing noisy feature matches.

mC.DMSinit Like `mC.kMatch`, `mC.DMSinit` is a binary parameter used to initialise Dense Motion Statistics (DMS) if the value is '1' or ignore if the value is '0'. If the value is '0' then the `motionEstimation` sub-routine inside the `correspondenceSearchir` function will not be utilised.

mC.seg This parameter is used to define the image segmentation algorithm used to extract image regions. If the value is '1', then SLIC segmentation [134] is performed. If '0', spatial grid-based binning of the image is used to extract regions.

mC.SLICregsize If SLIC segmentation is being utilised (`mC.seg` = 1), `mC.SLICregsize` is used as an approximation of the initial size of the super-pixel. As the parameter value increases, larger regions will be produced however this will result in potentially disparate regions of the image being grouped together. Conversely, if the region size specified is very small, too many regions will be produced.

mC.SLICregulariser This parameter is used to determine spatial regularity of the super-pixels generated by SLIC. If the value is increased, the spatial regularisation is increased to ensure no over-fitting when determining clusters.

mC.spatialBins This defines the number of spatial bins that are extracted providing (`mC.seg` = 0). The smaller number of regions may result in local motion-based statistics being obscured.

mC.DMS A binary flag that either enables ('1') or disables ('0') DMS.

mC.est_motion This parameter is used to give an initial estimated motion parameter if known a priori or effectively store the calculated DMS and the resulting 2-D motion estimation.

mC.percentile The percentile parameter is utilised within the region identification module of the program and dictates the percentage of regions with the highest and lowest density of matches, and the highest density of rejected matches. So, if `mC.percentile` = 20%, then 20% of the highest and lowest density of accepted matches and 20% of the highest density regions of rejected matches will be retained. By increasing the parameter, more regions will be processed however this may allow for weaker regions to be incorporated into the region strength determination. Conversely, keeping this value small effectively restricts the number of eventual matches produced by the system.

mC.rejperc This property is used if there is no rejected matches based on the initial region identification and matching procedure. `mC.rejperc` will be a parameter used to reject a number of matches based on their matching score that is calculated using the squared euclidean distance.

It is important to note that throughout the program, there are other parameters that get included within the `mC` parameter structure however these parameters are usually just to store variables and stop duplication of effort (e.g. `mC.sz` contains the image dimensions of the images within the dataset). Furthermore, many of the input parameters must be set heuristically based on the environment observed in the input dataset.

```
% Specify image directory and file extension
image_dirs = {'UoS_R1_0deg_test'};
image_ext = '/*.png ';
% Determine number of scans
numScans = size(image_dirs,2);
% Read images from specified directories
[org_filelist,filelist,n,mC.is64bit] = ...
    readAGRimages(numScans,...
image_dirs,...
image_ext,...
mC.bline);
```

Figure A.3: Input parameters and the associated function used to read image data-sets.

A.2.3 Input images

After the input matching parameters have been decided, the next stage is to specify the image directory that must be stored in the "INPUT_DATA" folder and file extension that is used by the images within. With the variables `image_dirs` and `image_ext`, the user can specify these options in the form of string variables and a code snippet is shown in Figure A.3. The input parameter `image_dirs` is a string-based variable that can point to multiple directories - this is particularly useful when a user wants to perform multiple reconstructions of different orientations within the fuel channel. Contrary to the name of the function, the `readAGRimages` function is utilised to generate two file-lists - the primary filelist is the contents of the entire folder that the user is pointing to that has the file extension specified by `image_ext` ($|filelist| = N_{images}$) and the second list is a file-list which has the artificial base-line applied to it ($|filelist| = \frac{N_{images}}{\texttt{mC.seqMatch}}$). The original filelist is retained for two reasons - firstly, because the motion estimation will require all images to be performed and secondly, future work was to do further pre-processing on the images and to automatically detect and remove images that were considered detrimental to the framework. If `mC.bline` > 1, then this would also reduce computation at the next stage, the correspondence searching process.

A.3 Correspondence Searching

The `correspondenceSearching` function is rather elaborate and contains a myriad of functionality pertaining to deriving an robust set of feature correspondences from the input image data-set. The function itself contains several core functions;

1. `performFeatureDetection`
2. `motionEstimation`
3. `performFeatureMatching`

In this section, each core function will be demonstrated in extensive detail with the associated contribution of each function highlighted. `correspondenceSearching` is diagrammatically illustrated in Figure A.4:

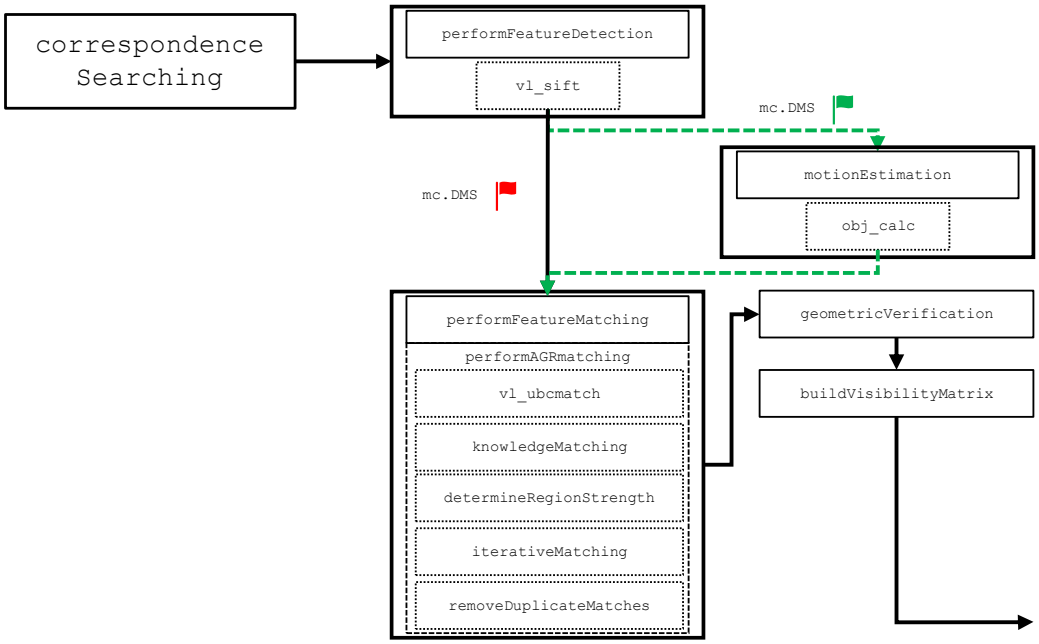


Figure A.4: Block diagram breakdown of the correspondence searching function containing only the most important functions to illustrate functionality.

A.3.1 performFeatureDetection()

The first function performed is `performFeatureDetection` which in itself is self-explanatory. With this function, the primary functionality is to perform SIFT feature detection with the "`vl_sift`" function, a MATLAB implementation [175] of Lowes SIFT algorithm [60]. The input images to this function utilise the baseline file list directory which intrinsically minimises computation. The function then extracts the colour of each associated feature point and this is primarily for visualisation of the feature-points at a later stage.

A.3.2 motionEstimation()

Once the feature detection has been completed, a conditional binary check is performed to ensure that `mC.DMS` == 1 before entering the `motionEstimation` function. If this condition is not met, the function is simply ignored. If the condition is met, the function is entered. Upon entry, a preliminary check is performed to ensure that there hasn't been a prior run of motion estimation on the dataset. The principle reasoning behind this decision is that performing the Dual TV-L^1 algorithm is computationally intensive, especially when dealing with a large image data-set. If it has been performed previously, the function simply loads a prior run into the `MATLAB` workspace before exiting the function. If there has been no prior run detected then the motion estimation is executed. This process is demonstrated in Figure A.5.

To perform motion estimation, the mexopencv wrapper to the OpenCVs implementation of the Dual TV-L^1 dense optical flow algorithm is used and a code snippet of how this works is demonstrated in Figure A.6.

A.3.3 performFeatureMatching()

The final and most intricate function within the correspondence searching process is `performFeatureMatching`. Within this function, there is a multitude of sub-routines such as `performAGRmatching`, `geometricVertification` and `buildVisibilityMatrix` that in themselves exceed the complexity of the previous

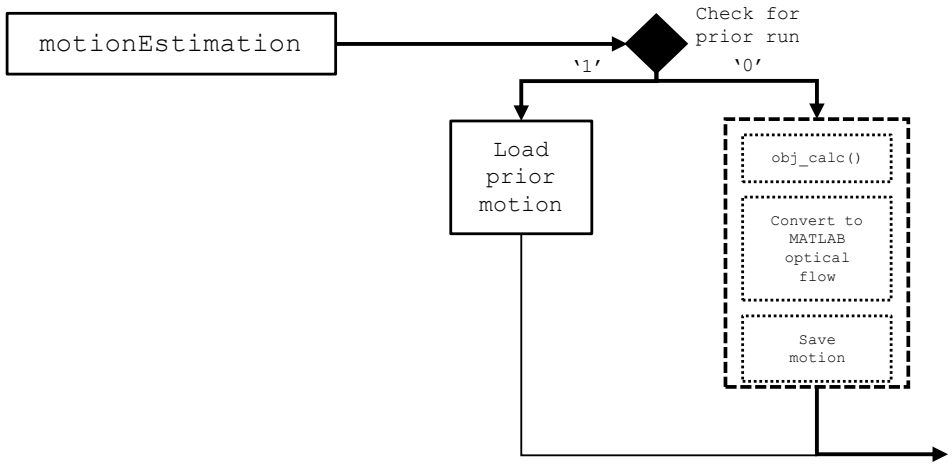


Figure A.5: Block diagram breakdown of the motion estimation function.

```
% Perform DMS
tic
nimg = length(org_filelist);
MFlow(1:nimg) = opticalFlow;
% Declare Dual TV-l1 instance
obj = cv.DualTVL1OpticalFlow();
% Perform optical flow on all the images within the dataset
for i = 2:nimg
% Read an image
    I1 = rgb2gray(imread(org_filelist(i-1).name));
    I2 = rgb2gray(imread(org_filelist(i).name));
    % Compute the Dual TV-l1 algorithm on the image
    flow = obj.calc(I1,I2);
    % Convert into MATLAB object container (for comparison with
    % other MATLAB inclusive optical flow methods)
    MFlow(i) = opticalFlow(double(flow(:,:,1)),double(flow(:,:,2)));
end
% Get performance statistics
motion_time = toc;
% Save estimated motion so that it can be reused
save(searchpath,'MFlow','motion_time');
```

Figure A.6: Code snippet illustrating the use of the mexopencv wrapper function, the conversion into a MATLAB compatible variable and how the function is saved for future reusage.

aforementioned functions from a coding stand-point and therefore require sub-sections of their own for further explanation. These functions operate in a sequential fashion and a pseudo-code demonstration of the functionality is demonstrated in Figure A.7

below:

```
% Perform feature matching between all frames
for i = 2:n_img
    % Initialise feature track
    % Perform a sliding window matching process
    for j = max(1,i-(mC.seqMatch)):i-1
        % Apply low feature space matching
        [matches,stats{j,i}] = performAGRmatching(...);
        % Perform geometric verification of matched features
        inlier = geometricVertification(...);
        % Only add the tracks if they exceed the minimum number of ...
            inliers
        if stats{j,i}.n_inlier > mC.gInliers
            % Add feature track
        end
    end
end

% Build Visibility Matrix and ascertain matching statistics
[featx,featy,vis,colour,stats] = buildVisibilityMatrix();
```

Figure A.7: Simplified pseudo-code snippet of how the feature matching procedure operates and how the functions operate in a sequential fashion.

performAGRmatching()

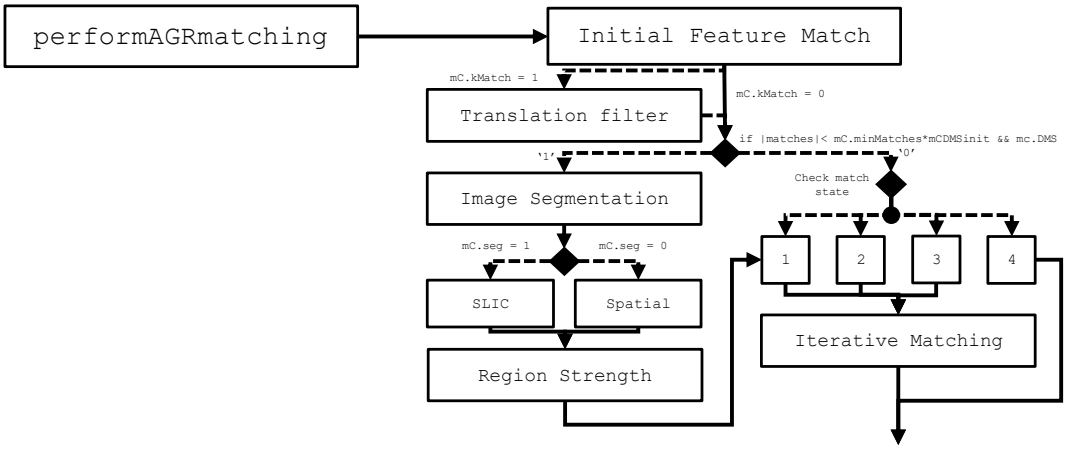


Figure A.8: Block diagram breakdown of the low feature space matching function, demonstrating the different states and execution paths.

The `performAGRmatching` function is inherently an implementation of the pro-

posed novelty in Chapter 4 of this document. Due to the various states of operation within this function, a block diagram is presented in Figure A.8 which provides an overview of how the function operates.The first stage of the matching procedure is to perform an initial match between two descriptors pertaining to the two assessed images using the `vl_ubcmatch()` matching function in VLFeat [175] at the pre-specified `mC.initThresh` value. If `mC.kMatch` $== 1$, the initial matches are then filtered using the expected translation parameters entered in the variable `mC.Shift`. Conversely, if `mC.kMatch` $== 0$, no filtering of the initial matches take place. Subsequently, the second stage of the matching procedure is to evaluate whether or not region identification and DMS is necessary through the use of logical operation:

```
if (length(matches)<mC.minMatches*mC.DMSinit) && mC.DMS
```

The first value initiates a check to see if there is enough matches produced during the initial matching stage. If the parameters entered are as observed in Figure A.2 and if the number of matches is 200, then $200 < (150 * 0.75) == 0$. With logical AND, this ensures that even if `mC.DMS` $== 1$, image segmentation, region identification and the incorporation of motion estimation is avoided since there is a sufficient number of matches found. Contrariwise, if the number of matches is insufficient (e.g. 50) and DMS is enabled, then $(50 < (150 * 0.75)) \wedge 1 == 1$, meaning the function is entered and is subjected to further processing.

If further processing is required, the subsequent stage is to perform "image segmentation"[1]. As described previously, there are two forms of segmentation available - spatial binning or superpixels generated by SLIC [134]. If `mC.seg` $== 1$ then SLIC is selected otherwise if `mC.seg` $== 0$ then spatial binning is utilised. Both forms of segmentation operate in a similar fashion and can be summarised with the following steps:

1. Segment the image into clusters (SLIC)/ spatial bins

2. Determine which cluster or bin the feature location resides in.

[1]If spatial segmentation is selected, then the feature locations are simply spatially binned using MATLABs `hist3()` function and no image processing is required.

3. If there is no rejected matches - reject `mC.rejperc`% candidates with the largest squared euclidean distance.

4. Determine regions with highest/lowest density of accepted matches, highest density of rejected matches and regions with high density of both accepted/rejected matches (ambiguous regions).

5. Produce binary region map pertaining to the strongest regions.

6. Determine interstitial 2-D translational motion between image frames and apply region map to remove estimated motion of weakest regions.

7. Utilise descriptive statistics of both horizontal and vertical motion components in strongest regions and obtain maximum estimated motion for further processing.

Once further processing has been complete, the results will be relayed to a state machine where there is four possible operational states, after which it will then perform the `iterativeMatching` function accordingly:

State 1 The first state utilises the motion determined from the further processing stages associated with DMS. This motion is incorporated and used to filter the feature match set initially before continuing to relax the matching threshold at each iteration until either the number of matches exceed `mC.minMatches` or $n_{iter} =$ `mC.maxIter` - whichever comes first.

State 2 The second state operates very differently and is only entered if `mC.kMatch` is enabled, yet the number of matches produced by the initial translation filter is ineffective with regards to producing enough matches. This state incorporates the values specified in `mC.Shift` and performs iterative matching as described in `State 1`.

State 3 The third stage is entered if both `mC.kMatch` and `mC.DMS` are both disabled - leaving the system to iteratively match with no pre-filtering of the feature set. This stage is useful for when the motion is completely unknown

or the footage hasn't been gathered in a contiguous fashion and is instead acquired with a naturally wide base-line.

`State 4` The final state is reserved for when the initial matching phase was sufficient and there is no requirement to obtain further matches.

Once each state has been completed, the resultant feature matches are then subjected to a check to ensure that there is no duplicate matches - this is caused where one feature set in the reference image is successfully matched to multiple different features sets in the processed image. This causes issues at the geometric verification stage when using consensus techniques such as RANSAC [68] which have a uniqueness constraint (i.e. 1-to-1 match) that requires that each row is linearly independent for RANSAC sampling. To avoid this, the feature set is checked for duplicate matches and are subsequently removed for geometric verification.

geometricVertification()

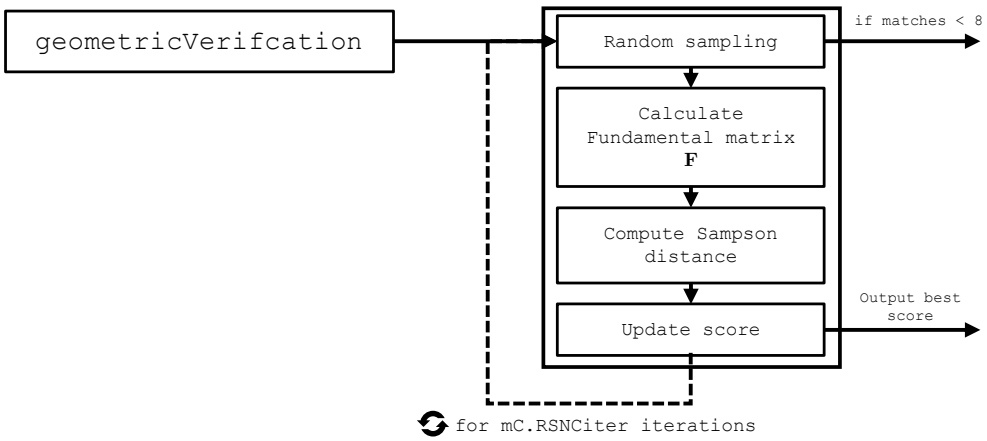


Figure A.9: Block diagram breakdown of the geometric verification process, computing the Fundamental matrix with RANSAC.

Once matches have been obtained from `performAGRmatching`, the next stage is to geometrically verify the produced matches and remove any remaining outliers. To do this, a random sample of feature points is obtained which is then used to compute

the Fundamental matrix, using the normalised 8-point algorithm [66]. Afterwards, the Sampson distance is used to obtain the squared distance between the feature point to the corresponding epipolar line [57] and is subsequently stored. This process is then continually repeated by `mC.RSNCiter` iterations, with the score of the previous run being compared to the current run, ensuring that once the process is complete that the resultant consensus and computed **F** is accurate. This process is visualised in Figure A.9.

buildVisibilityMatrix()

The final stage of the `correspondenceSearching` function is to generate a visibility matrix (a.k.a. a visibility map or a feature track) which is used to track feature points appearing across multiple image and is stored in the binary matrix where each column represents an image or view-point and every row represents an extracted feature. This function acts as a post-processing function by first checking that the generated feature tracks pertaining to each feature can be observed in more than `mC.minViews`. Secondly and finally, the function will organise and extract matching statistics from the previous functions such as the mean number of inliers, punitive matches, iterations and correct matches which can be utilised to assess the framework functionality.

A.4 Incremental Reconstruction

With a robust set of feature correspondences, the final stage of the framework[2] comprises of producing a sparse 3-D point cloud model using incremental reconstruction. The functionality of this section describes the reconstruction element of the framework. The sub-routines contained within this function are diagrammatically demonstrated in Figure A.10 below:

The first stage within the `performIncrementalReconstruction` function is to check if there is a calibration matrix **K** containing the intrinsic parameters. If $\nexists$ **K**,

[2]There is a reconstruction refinement functionality also available for AGR reconstructions only which produces circumferential reconstructions from multiple orientation-based reconstructions. This will be discussed in another chapter.

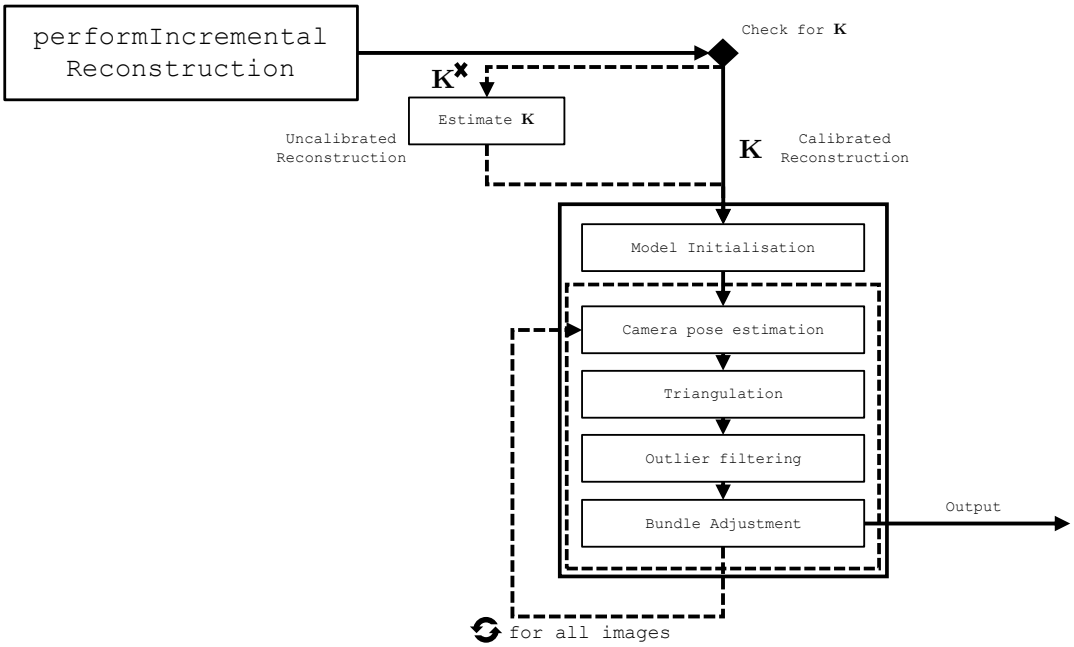


Figure A.10: Block diagram breakdown of the incremental reconstruction process.

a crude estimation using feature locations is performed to obtain the principal point $\mathbf{p} = [c_x, c_y]$ and the focal length $f = [f_x, f_y]$ as defined below:

$$\mathbf{K} = \begin{bmatrix} 2 \times c_x - \min(\mathbf{x}_x) & s & \frac{\max(\mathbf{x}_x) - \min(\mathbf{x}_x)}{2} \\ 0 & 2 \times c_x - \min(\mathbf{x}_x) & \frac{\max(\mathbf{x}_y) - \min(\mathbf{x}_y)}{2} \\ 0 & 0 & 1 \end{bmatrix} \tag{A.1}$$

Once $\mathbf{K}$ is attained (either via a calibration process or generated synthetically), the next stage is to initialise the model. Model initialisation comprises of carefully selecting a strong two-view reconstruction, as due to the incremental nature of the reconstruction, a bad initialisation will result in a poor and misrepresentative point cloud [21].

After model initialisation, new images can now be added to the initialised 3-D model through the use of Perspective-n-Point approach which aims to estimate the camera pose pertaining to each image given a set of 3-D points and the corresponding 2-D feature locations. Since the camera pose estimation process is subject to potential outliers, a simple process similar to geometricVertification() is taken where RANSAC

obtains random samples, performs the Direct Linear Transformation (DLT) algorithm for 2-D→3-D correspondences [57], computes the reprojection error for each random consensus and takes the best scoring consensus. Once more than 3 views have been assessed, the system will aim attempt to auto-calibrate [70] by perturbing the focal length parameters in order to minimise the number of outliers when performing the DLT.

With each new image or view-point being registered, they must observe 3-D scene points that were reconstructed during the model initialisation process. Therefore, by using the visibility matrix where each column pertains to features observed by each image as previously discussed in Chapter 5.4.3 - buildVisibilityMatrix(), the newly registered image can be processed to see if previous scene points are observed and therefore add new 3-D points to the scene using triangulation. Once camera pose estimation and triangulation is complete, a non-linear refinement procedure is undertaken to further refine the camera pose estimation and the triangulated points [21].

Bibliography

[1] P. Murray, G. West, S. Marshall, and S. Mcarthur, "Automated image stitching for enhanced visual inspections of nuclear power stations," *Cm & Mfpt*, 2013.

[2] P. Murray, G. West, S. Marshall, and S. McArthur, "Automated in-core image generation from video to aid visual inspection of nuclear power plant cores," *Nuclear Engineering and Design*, vol. 300, pp. 57–66, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.nucengdes.2015.11.037

[3] G. M. West, C. J. Wallace, and S. D. McArthur, "Combining models of behaviour with operational data to provide enhanced condition monitoring of AGR cores," *Nuclear Engineering and Design*, vol. 272, pp. 11–18, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.nucengdes.2013.12.067

[4] T. Lardner, G. West, G. Dobie, and A. Gachagan, "Automated sizing and classification of defects in CANDU pressure tubes," *Nuclear Engineering and Design*, vol. 325, no. January, pp. 25–32, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.nucengdes.2017.09.029

[5] F.-C. Chen and R. M. R. Jahanshahi, "NB-CNN: Deep Learning-based Crack Detection Using Convolutional Neural Network and Naïve Bayes Data Fusion," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4393–4400, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/8074762/

[6] B. H. Cho, S. H. Byun, C. H. Shin, J. B. Yang, S. I. Song, and J. M. Oh, "KeproVt: Underwater robotic system for visual inspection of nuclear reactor internals," *Nuclear Engineering and Design*, vol. 231, no. 3, pp. 327–335, 2004.

[7] J. H. Kim, J. C. Lee, and Y. R. Choi, "LAROB: Laser-guided underwater mobile robot for reactor vessel inspection," *IEEE/ASME Transactions on Mechatronics*, vol. 19, no. 4, pp. 1216–1225, 2014.

[8] K. Nagatani, S. Kiribayashi, Y. Okada, K. Otake, K. Yoshida, S. Tadokoro, T. Nishimura, T. Yoshida, E. Koyanagi, M. Fukushima, and S. Kawatsuma, "Emergency Response to the Nuclear Accident at the Fukushima Daiichi Nuclear Power Plants using Mobile Rescue Robots," *Journal of Field Robotics*, vol. 30, no. 1, pp. 44–63, 2013.

[9] A. Seki, O. J. Woodford, S. Ito, B. Stenger, M. Hatakeyama, and J. Shimamura, "Reconstructing Fukushima: A Case Study," in *2014 2nd International Conference on 3D Vision*. Tokyo, Japan: IEEE, 2014, pp. 681–688. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7035885

[10] P. Nanekar, N. Jothilakshmi, and T. Jayakumar, "Ultrasonic phased array examination of circumferential weld joint in reactor pressure vessel of BWR," *Nuclear Engineering and Design*,

vol. 265, pp. 366–374, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.nucengdes.2013.09.015

[11] C. J. Wallace, G. M. West, G. J. Jahn, S. D. J. McArthur, D. Towle, and G. Buckley, "Control Rod Monitoring Of Advanced Gas Cooled Reactors," in *Seventh American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies*. Las Vegas, Nevada, USA: ANS, 2010, pp. 1–10.

[12] S. El Kahi, D. Asmar, A. Fakih, J. Nieto, and E. Nebot, "A vison-based system for mapping the inside of a pipe," in *International Conference on Robotics and Biomimetics*, no. 1. Phuket, Thailand: 2011, 2011, pp. 2605–2611.

[13] 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. 1–27, 2014.

[14] R. Summan, G. Dobie, G. West, S. Marshall, S. Member, C. Macleod, and S. G. Pierce, "The Influence of the Spatial Distribution of 2-D Features on Pose Estimation for a Visual Pipe Mapping Sensor," *IEEE Sensors Journal*, vol. 17, no. 19, pp. 6312–6321, 2017.

[15] J. Kannala, S. S. Brandt, and J. Heikkilä, "Measuring and modelling sewer pipes from video," *Machine Vision and Applications*, vol. 19, no. 2, pp. 73–83, 2008.

[16] S. Esquivel, R. Koch, and H. Rehse, "Reconstruction of sewer shaft profiles from fisheye-lens camera images," in *Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM) Symposium*, vol. 5748 LNCS. Jena, Germany: Springer-Verlag Berlin Heidelberg, 2009, pp. 332–341.

[17] M. Pollefeys, L. Van Gool, M. Vergauwen, K. Cornelis, F. Verbiest, and J. Tops, "3D recording for archaeological fieldwork," *IEEE Computer Graphics and Applications*, vol. 23, no. 3, pp. 20–27, may 2003.

[18] S. Fuhrmann, F. Langguth, and M. Goesele, "MVE-A Multi-View Reconstruction Environment," in *Eurographics Workshop on Graphics and Cultural Heritage (2014)*. Darmstadt, Germany: Eurographics Association Aire-la-Ville, 2014, pp. 11–18. [Online]. Available: http://diglib2.eg.org/EG/DL/WS/GCH/GCH2014/011-018.pdf.abstract.pdf;internal{&}action=action.digitallibrary.ShowPaperAbstract

[19] N. Snavely, S. Seitz, and R. Szeliski, "PhotoTourism: Exploring Photo Collections in 3D," in *SIGGRAPH Conference Proceedings*. Boston, MA, USA: ACM, 2006, pp. 835—-846.

[20] S. J. D. E. F. J.-m. Heinly, Jared, "Reconstructing the World * in Six Days," in *Computer Vision and Pattern Recognition*. Boston, MA, USA: IEEE, 2015, pp. 3287–3295.

[21] J. L. Schönberger and J.-M. Frahm, "Structure-from-Motion Revisited," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, Nevada, USA: IEEE, 2016, pp. 4104–4113.

[22] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1052–1067, 2007.

[23] F. Santoso, M. Garratt, M. Pickering, and M. Asikuzzaman, "3D-Mapping for Visualisation of Rigid Structures: A Review and Comparative Study," *IEEE Sensors Journal*, no. c, pp. 1–1, 2015. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7322186

[24] M. Bradford, "Reactor Inspection and Monitoring and its Role in the Reactor Safety Case," in *Proceedings of the 4th EDF Energy Generation Ltd Nuclear Graphite Conference*, May 2014.

[25] C. Wu, "Towards linear-time incremental structure from motion," in *International Conference on 3D Vision, 3DV 2013*. Seattle, Washington, USA: IEEE, 2013, pp. 127–134.

[26] E. Nonbøl, *Description of the Advanced Gas Cooled Type of Reactor (AGR)*. NKS, Nordisk Kernesikkerhedsforskning, 1996.

[27] C. J. Wallace, G. M. West, S. D. J. McArthur, and D. Towle, "Distributed Data and Information Fusion for Nuclear Reactor Condition Monitoring," *IEEE Transactions on Nuclear Science*, vol. 59, no. 1, pp. 182–189, 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6123182/

[28] C. Wallace, "Distributed Data Fusion for Condition Monitoring of Graphite Nuclear Reactor Cores," Ph.D. dissertation, University of Strathclyde, 2013.

[29] C. Wallace, G. West, G. Jahn, S. Mcarthur, D. Towle, and G. Buckley, "Control rod monitoring of advanced gas-cooled reactors," in *7th International Topical Meeting on Nuclear Plant Instrumentation, Control and Human Machine Interface Technologies (NPIC&HMIT 2010)*, 11 2010.

[30] C. E. Watson, P. C. Robinson, and P. R. Maul, "The role of data visualisation in core inspection decision making," in *Modelling and Measuring Reactor Core Graphite Properties and Performance*. The Royal Society of Chemistry, 2013, pp. 149–157. [Online]. Available: http://dx.doi.org/10.1039/9781849735179-00149

[31] C. E. Watson, P. C. Robinson, J. F. Burrow, and P. R. Maul, "The application of portfolio selection to fuel channel inspection in advanced gas-cooled reactors," *Nuclear Engineering and Design*, vol. 328, no. January, pp. 145–153, 2018. [Online]. Available: https://doi.org/10.1016/j.nucengdes.2018.01.013

[32] K. Law, G. West, P. Murray, and C. Lynch, "3-d visualization of agr fuel channel bricks using structure-from-motion," *Nuclear Engineering and Design*, vol. 359, p. 110472, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0029549319305035

[33] G. West, P. Murray, S. Marshall, and S. Mcarthur, "Improved Visual Inspection of Advanced Gas-Cooled Reactor Fuel Channels," *International Journal of Prognostics and Health Management*, 2015.

[34] B. L. Luk, T. S. White, D. S. Cooke, N. D. Hewer, G. Hazel, S. Chen, B. Centre, and G. Gl, "Climbing Service Robot for Duct Inspection and Maintenance Applications in a Nuclear Reactor," in *Proceedings of the 32nd ISR(International Symposium on Robotics)*, no. April, 2001, pp. 19–21.

[35] K. Seungho, H. J. Seung, U. L. Sung, H. K. Chang, C. S. Ho, C. S. Yong, H. L. Nam, and M. J. Kyung, "Application of robotics for the nuclear power plants in Korea," in *2010 1st International Conference on Applied Robotics for the Power Industry, CARPI 2010*. Montreal, Canada: IEEE, 2010, pp. 1–5.

[36] K. Yoshida and S. Tadokoro, "Improvements to the Rescue Robot Quince Toward Future Indoor Surveillance Missions in the Fukushima Daiichi Nuclear Power Plant," in *Field and Service Robotics: Results of the 8th International Conference*, vol. 92. Springer Berlin Heidelberg, 2014, pp. 19–32.

[37] M. Trelinski, "Inspection of CANDU Reactor Pressure Tubes Using Ultrasonics," in *17th World Conference on Nondestructive Testing*, Shanghai, China, 2008, pp. 1–8.

[38] C. Choi, B. Park, and S. Jung, "The design and analysis of a feeder pipe inspection robot with an automatic pipe tracking system," *IEEE/ASME Transactions on Mechatronics*, vol. 15, no. 5, pp. 736–745, 2010.

[39] C. Neri, L. Bartolini, A. Coletti, M. F. De Collibus, G. Fometti, S. Lupini, F. Pollastrone, L. Semeraro, and C. Talarico, "Advanced Digital Processing for Amplitude and Range Determination in Optical RADAR Systems," *IEEE Transactions on Nuclear Science*, vol. 49, no. 2, pp. 417–422, 2002.

[40] P. Rossi, C. Neri, M. F. D. Collibus, G. Mugnaini, F. Pollastrone, and F. Crescenzi, "Ivvs probe mechanical concept design," *Fusion Engineering and Design*, vol. 98-99, pp. 1597 – 1600, 2015, proceedings of the 28th Symposium On Fusion Technology (SOFT-28). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0920379615003063

[41] R. S. Adrain, I. A. Armour, and J. H. Bach, "Laser scanning cameras for inreactor inspection," *Sensor Review*, vol. 7, no. 2, pp. 68–76, 1987. [Online]. Available: http://dx.doi.org/10.1108/eb007716

[42] J. Li, X. Wu, T. Xu, H. Guo, J. Sun, and Q. Gao, "A novel inspection robot for nuclear station steam generator secondary side with self-localization," *Robotics and Biomimetics*, vol. 4, no. 1, p. 26, Dec 2017. [Online]. Available: https://doi.org/10.1186/s40638-017-0078-y

[43] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*. Thomson-Engineering, 2007.

[44] K. Law, G. West, P. Murray, and C. Lynch, "Towards Extracting 3-D Structural Representations of AGR Core Fuel Channels from 2-D In-Core Inspection Videos," in *International Symposium*

on Future I&C for Nuclear Power Plants (ISOFIC 2017). Gyeongju, Republic of Korea: KNS, 2017, pp. 1–10.

[45] ——, "3-D Advanced Gas-cooled Nuclear Reactor Fuel Channel Reconstruction using Structure-from-Motion," in *International Topical Meeting on Nuclear Plant Instrumentation, Control and Human Machine Interface Technologies*. San Francisco, CA, USA: ANS, 2017, pp. 67–76.

[46] R. Summan, G. Dobie, G. West, S. Marshall, C. MacLeod, and S. G. Pierce, "The influence of the spatial distribution of 2-d features on pose estimation for a visual pipe mapping sensor," *IEEE Sensors Journal*, vol. 17, no. 19, pp. 6312–6321, Oct 2017.

[47] S. Y. Bao and S. Savarese, "Semantic structure from motion," in *Computer Vision and Pattern Recognition*. Colorado Springs, CO, USA: IEEE, 2011, pp. 2025–2032.

[48] F. Baqué, G. Rodriguez, N. Jardin, J. M. Carpreau, J. M. Augem, and J. Sibilo, "Generation IV nuclear reactors - R&D program to improve sodium-cooled systems inspection," in *ANIMMA 2009 - 2009 1st International Conference on Advancements in Nuclear Instrumentation, Measurement Methods and their Applications*. Marseille, France: IEEE, 2009, pp. 1–8.

[49] C. Mei and P. Rives, "Single view point omnidirectional camera calibration from planar grids," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, April 2007, pp. 3945–3950.

[50] V. Usenko, N. Demmel, and D. Cremers, "The double sphere camera model," in *Proc. of the Int. Conference on 3D Vision (3DV)*, September 2018.

[51] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," Massachusetts Institute of Technology, Cambridge, MA, USA, Tech. Rep., 1976.

[52] A. Mazumdar, M. Lozano, A. Fittery, and H. H. Asada, "A compact, maneuverable, underwater robot for direct inspection of nuclear power piping systems," in *2012 IEEE International Conference on Robotics and Automation*, May 2012, pp. 2818–2823.

[53] B. L. Luk, A. A. Collie, D. S. Cooke, and S. Chen, "Walking and climbing service robots for safety inspection of nuclear reactor pressure vessels," *Journal of Measurement and Control*, vol. 39, no. 2, p. 43, 2006. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.4661{&}rep=rep1{&}type=pdf

[54] M. Dong, W. Chou, B. Fang, G. Yao, and Q. Liu, "Implementation of remotely operated vehicle for direct inspection of reactor pressure vessel and other water-filled infrastructure," *Journal of Nuclear Science and Technology*, vol. 53, no. 8, pp. 1086–1096, 2016. [Online]. Available: https://doi.org/10.1080/00223131.2015.1091395

[55] T. E. Lee and N. Michael, "State Estimation and Localization for ROV-Based Reactor Pressure Vessel Inspection," in *Field and Service Robotics*. Springer International Publishing, 2018, pp. 699–715.

[56] P. Murray, G. M. West, K. Law, S. Buckley-Mellor, G. Cocks, and C. Lynch, "Automated Video Processing and Image Analysis Software to Support Visual Inspection of AGR Cores," in *5th EDF Energy Generation Ltd Nuclear Graphite Conference*, 2016. [Online]. Available: https://pure.strath.ac.uk/portal/en/publications/automated-video-processing-and-image-analysis-software-to-support-visual-inspection-of-agr-cores(5a833906-8e3b/export.html

[57] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision.* Cambridge University Press, 2004, vol. 2.

[58] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models.* SpringerVerlag, 2003.

[59] R. Szeliski, *Computer Vision - Algorithms and Applications.* Springer, 2012. [Online]. Available: http://link.springer.com/10.1007/978-1-84882-256-6

[60] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[61] J. L. Schönberger, H. Hardmeier, T. Sattler, and M. Pollefeys, "Comparative Evaluation of Hand - Crafted and Learned Local Features," in *Computer Vision and Pattern Recognition.* Honolulu, Hawaii: IEEE, 2017, pp. 1–10. [Online]. Available: https://www.researchgate.net/profile/Johannes{_}Schoenberger/publication/315788135{_}Comparative{_}Evaluation{_}of{_}Hand-Crafted{_}and{_}Learned{_}Local{_}Features/links/58e4f20145851547e17d4dee/Comparative-Evaluation-of-Hand-Crafted-and-Learned-Local-Features.pdf

[62] J. Maier, M. Humenberger, M. Murschitz, O. Zendel, and M. Vincze, "Guided Matching Based on Statistical Optical Flow for Fast and Robust Correspondence Analysis," in *European Conference on Computer Vision (ECCV)*, M. Leibe, Bastian and Matas, Jiri and Sebe, Nicu and Welling, Ed. Springer International Publishing, 2016, no. 1, pp. 101–117. [Online]. Available: http://www.nature.com/doifinder/10.1038/nature14539{%}5Cnhttp://www.nature.com/doifinder/10.1038/nmeth.3707{%}5Cnhttp://link.springer.com/10.1007/978-3-319-46448-0

[63] J. Bian and W.-y. Lin, "GMS : Grid-based Motion Statistics for Fast , Ultra-robust Feature Correspondence," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 302, 2017, pp. 2828–2837.

[64] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein, "The Generalized PatchMatch Correspondence Algorithm," in *European Conference on Computer Vision (ECCV)*, 2010, pp. 29–43.

[65] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "DeepMatching: Hierarchical Deformable Dense Matching," *International Journal of Computer Vision*, vol. 120, no. 3, pp. 300–323, 2016.

[66] H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature*, vol. 293, pp. 133–135, Sep. 1981.

[67] Q. Luong and O. D. Faugeras, "Determining the fundamental matrix with planes: unstability and new algorithms," in *Computer Vision and Pattern Recognition*, vol. Volume 1. New York, NY, USA: IEEE, 1993, pp. 489–494.

[68] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981. [Online]. Available: http://doi.acm.org/10.1145/358669.358692

[69] E. Brachmann, F. Michel, A. Krull, M. Y. Yang, S. Gumhold, and C. Rother, "Uncertainty-Driven 6D Pose Estimation of Objects and Scenes from a Single RGB Image," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),*, 2016.

[70] B. Triggs, "Autocalibration and the absolute quadric," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, Jun 1997, pp. 609–614.

[71] J. J. Koenderink and a. J. van Doorn, "Affine structure from motion." *Journal of the Optical Society of America. A, Optics and image science*, vol. 8, no. 2, pp. 377–385, 1991.

[72] O. D. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig?" in *European Conference on Computer Vision (ECCV)*. Santa Margherita Ligure, Italy: IEEE, 1992, pp. 563–578. [Online]. Available: http://www.springerlink.com/content/72q87065781071r0/fulltext.pdf

[73] M. Pollefeys, R. Koch, and L. Van Gool, "Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters," *International Journal of Computer Vision*, vol. 32, no. 1, pp. 7–25, 1999.

[74] F. Schaffalitzky and A. Zisserman, "Multi view matching for unordered image sets," in *European Conference on Computer Vision (ECCV)*. Copenhagen, Denmark: IEEE, 2002, pp. 414–431.

[75] B. Triggs, B. T. Autocalibration, Q. International, and B. Triggs, "Autocalibration and the Absolute Quadric," in *Computer Vision and Pattern Recognition*. San Juan, Puerto Rico: IEEE Computer Society, 1997, pp. 1–6.

[76] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle Adjustment A Modern Synthesis," *Vision Algorithms: Theory and Practice*, vol. 1883, pp. 298–372, 2000. [Online]. Available: http://dx.doi.org/10.1007/3-540-44480-7{_}21{%}5Cnhttp://link.springer.com/10.1007/3-540-44480-7{_}21

[77] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o(n) solution to the pnp problem," *International Journal of Computer Vision*, vol. 81, no. 2, p. 155, Jul 2008. [Online]. Available: https://doi.org/10.1007/s11263-008-0152-6

[78] R. I. Hartley and P. Sturm, "Triangulation," *American Computer Vision and Image Understanding Workshop*, vol. 68, no. 2, pp. 146–157, 1997.

[79] P. Lindstrom, "Triangulation made easy," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1554–1561.

[80] M. I. A. Lourakis and A. A. Argyros, "SBA: A Software Package for Generic Sparse Bundle Adjustment," *ACM Transactions on Mathematical Software*, vol. 36, no. 1, pp. 1–30, 2009. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1486525.1486527

[81] A. J. Davison, "Mobile Robot Navigation using Active Vision," Ph.D. dissertation, University of Oxford, 1998.

[82] A. J. Davison and D. W. Murray, "Simultaneous localization and map-building using active vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 865–880, 2002.

[83] M. Brown and D. G. Lowe, "Unsupervised 3D Object Recognition and Reconstruction in Unordered Datasets," in *Proceedings of the Fifth International Conference on 3-D Digital Imaging and Modeling (3DIM'05)*. Ottawa, Ontario, Canada: IEEE, 2005, pp. 56–63.

[84] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the world from Internet photo collections," *International Journal of Computer Vision*, vol. 80, no. 2, pp. 189–210, 2008.

[85] ——, "Skeletal graphs for efficient structure from motion," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, June 2008, pp. 1–8.

[86] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski, "Building Rome in a day." in *ICCV*. IEEE Computer Society, 2009, pp. 72–79.

[87] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski, "Building rome in a day," *Commun. ACM*, vol. 54, no. 10, pp. 105–112, Oct. 2011. [Online]. Available: http://doi.acm.org/10.1145/2001269.2001293

[88] Y. Furukawa and J. Ponce, "Accurate, Dense, and Robust Multi-View Stereopsis," in *Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8. [Online]. Available: http://books.google.com/books?hl=en{&}lr={&}id=7N2CAdoLtd8C{&}oi=fnd{&}pg=PR8{&}dq={%}22and+obstacles.+It+does+not+perform+any+smoothing{%}22+{%}22before+using+visibility+constraints+to+?lter+away{%}22+{%}22away+false+matches.+A+simple+but+effective+method{%}22+{%}22phot

[89] J. M. Frahm, P. Fite-Georgel, D. Gallup, T. Johnson, R. Raguram, C. Wu, Y. H. Jen, E. Dunn, B. Clipp, S. Lazebnik, and M. Pollefeys, "Building Rome on a cloudless day," in *European Conference on Computer Vision (ECCV)*, vol. 6314 LNCS, no. PART 4. Crete, Greece: Springer-Verlag Berlin Heidelberg, 2010, pp. 368–381.

[90] P. Moulon, P. Monasse, and R. Marlet, "Adaptive Structure from Motion with a contrario model estimation," in *Asian Conference on Computer Vision*, vol. 7727 LNCS, no. PART 4. Daejeon, Korea: Springer International Publishing, 2012, pp. 1–14. [Online]. Available: http://link.springer.com/10.1007/978-3-642-37447-0{%}7B{_}{%}7D20

[91] J. L. Schönberger, F. Radenović, O. Chum, and J. M. Frahm, "From single image query to detailed 3D reconstruction," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Boston, MA, USA: IEEE, 2015, pp. 5126–5134.

[92] D. Nistér, "Reconstruction from Uncalibrated Sequences with a Hierarchy of Trifocal Tensors," in *European Conference on Computer Vision (ECCV)*. Dublin, Ireland: IEEE, 2000, pp. 649–663. [Online]. Available: http://link.springer.com/chapter/10.1007/3-540-45054-8{_}42

[93] M. Farenzena, A. Fusiello, and R. Gherardi, "Structure-and-motion pipeline on a hierarchical cluster tree," in *International Conference on Computer Vision Workshops*. Kyoto, Japan: IEEE, 2009, pp. 1489–1496.

[94] R. Gherardi, M. Farenzena, and A. Fusiello, "Improving the efficiency of hierarchical structure-and-motion," in *Computer Vision and Pattern Recognition*. San Francisco, CA, USA: IEEE, 2010, pp. 1594–1600.

[95] K. Ni and F. Dellaert, "HyperSfM," in *Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*. Zurich, Switzerland: IEEE, 2012, pp. 144–151.

[96] R. Toldo, R. Gherardi, M. Farenzena, and A. Fusiello, "Hierarchical structure-and-motion recovery from uncalibrated images," *Computer Vision and Image Understanding*, vol. 140, pp. 127–143, 2015.

[97] P. Sturm and B. Triggs, "A factorization based algorithm for multi-image projective structure and motion," in *Computer Vision — ECCV '96*, B. Buxton and R. Cipolla, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1996, pp. 709–720.

[98] O. Enqvist, F. Kahl, and C. Olsson, "Non-sequential structure from motion," in *International Conference on Computer Vision Workshops*. Barcelona, Spain: IEEE, 2011, pp. 264–271.

[99] H. Cui, X. Gao, S. Shen, and Z. Hu, "Hsfm: Hybrid structure-from-motion," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 2393–2402.

[100] V. M. Govindu, "Combining Two-view Constraints for Motion Estimation," in *Computer Vision and Pattern Recognition*, vol. 2. Kauai, HI, USA, 2001, pp. 218–225.

[101] V. Govindu, "Lie-algebraic averaging for globally consistent motion estimation," in *Computer Vision and Pattern Recognition*, vol. 1. Washington, DC, USA: IEEE, 2004, pp. 1–8.

[102] Z. Cui and P. Tan, "Global structure-from-motion by similarity averaging," in *Computer Vision and Pattern Recognition*. Las Vegas, Nevada, USA: IEEE, 2016, pp. 864–872.

[103] R. I. Hartley, J. Trumpf, Y. Dai, and H. Li, "Rotation averaging," *International Journal of Computer Vision*, vol. 103, pp. 267–305, 2012.

[104] K. Wilson, D. Bindel, and N. Snavely, "When is rotations averaging hard?" in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 255–270.

[105] Z. Cui, N. Jiang, C. Tang, and P. Tan, "Linear Global Translation Estimation with Feature Tracks," in *Proceedings of the 2015 British Conference on Machine Vision*, no. 3, 2015, p. 2014.

[106] M. Kolesnik and G. Baratoff, "3-D Interpretation of Sewer Circular Structures," in *Proceedings of the 2000 IEEE international Conference on Robotics & Automation*, no. April. San Francisco, CA, USA: IEEE, 2000, pp. 1453–1458.

[107] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Procedings of the Alvey Vision Conference 1988*, pp. 147–151, 1988. [Online]. Available: http://www.bmva.org/bmvc/1988/avc-88-023.html

[108] G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces," in *IEEE and ACM International Symposium on Mixed and Augmented Reality*. Nara: IEEE, 2007, pp. 1–10. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4538852

[109] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proceedings of the IEEE International Conference on Computer Vision*, 2011, pp. 2564–2571.

[110] G. Vogiatzis and C. Hernández, "Video-based, real-time multi-view stereo," *Image and Vision Computing*, vol. 29, no. 7, pp. 434–441, 2011. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0262885611000138

[111] R. Summan, W. Jackson, G. Dobie, C. MacLeod, C. Mineo, G. West, D. Offin, G. Bolton, S. Marshall, and A. Lille, "A novel visual pipework inspection system," *AIP Conference Proceedings*, vol. 1949, no. 1, p. 220001, 2018. [Online]. Available: https://aip.scitation.org/doi/abs/10.1063/1.5031647

[112] J.-s. Gutmann, D. Freiburg, K. Konolige, and M. Park, "Incremental Mapping of Large Cyclic Environments," in *International Symposium on Computational Intelligence in Robotics and Automation*. Monterey, CA, USA: IEEE, 1999, pp. 318–325.

[113] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers, "An improved algorithm for TV-L 1 optical flow," in *Statistical and Geometrical Approaches to Visual Motion Analysis*, vol. 5604. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 23–45. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-03061-1{_}2

[114] Microsoft photosynth. [Online]. Available: https://photosynth.net/

[115] Agisoft photoscan. [Online]. Available: http://www.agisoft.com/

[116] Autodesk 123d catch. [Online]. Available: http://www.123dapp.com/catch

[117] Acute3d - capturing reality with automatic 3d photogrammetry software. [Online]. Available: https://www.acute3d.com/

[118] P. Moulon, P. Monasse, and R. Marlet, "Global fusion of relative motions for robust, accurate and scalable structure from motion," in *Proceedings of the IEEE International Conference on Computer Vision*. Sydney, Austrailia: IEEE, 2013, pp. 3248–3255.

[119] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM: A factored solution to the simultaneous localization and mapping problem," in *In Proceedings of the AAAI National Conference on Artificial Intelligence*, vol. 68, no. 2. AAAI, 2002, pp. 593–598. [Online]. Available: http://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:FastSLAM+:+A+Factored+Solution+to+the+Simultaneous+Localization+and+Mapping+Problem{#}0

[120] M. Montemerlo, S. Thrun, D. Roller, and B. Wegbreit, "FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges," in *IJCAI International Joint Conference on Artificial Intelligence*. Acapulco, Mexico: IJCAI, 2003, pp. 1151–1156.

[121] J. Civera, A. J. Davison, and J. M. M. Montiel, "Inverse depth parametrization for monocular SLAM," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 932–945, 2008.

[122] J. Engel, T. Sch, and D. Cremers, "LSD-SLAM: Large-Scale Direct Monocular SLAM," in *European Conference on Computer Vision (ECCV)*. Zurich, Switzerland: Springer International Publishing Switzerland, 2014, pp. 834–849.

[123] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a Versatile and Accurate Monocular SLAM System," *IEEE Transactions on Robotics*, vol. 31, no. 5, p. 15, 2015. [Online]. Available: http://arxiv.org/abs/1502.00956

[124] R. Mur-Artal and J. D. Tardos, "Probabilistic Semi-Dense Mapping from Highly Accurate Feature-Based Monocular SLAM," in *Robotics: Science and Systems*, Rome, Italy, 2015.

[125] H. Strasdat, J. M. M. Montiel, and A. J. Davison, "Scale Drift-Aware Large Scale Monocular SLAM," in *Robotics: Science and Systems*, vol. 2. Zaragoza, Spain: IEEE, 2010, p. 5. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.165.7975{&}rep=rep1{&}type=pdf

[126] S. Baker, I. Matthews, R. Patil, G. Cheung, I. Matthews, R. Gross, T. Ishikawa, and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, 2004. [Online]. Available: http://www.springerlink.com/openurl.asp?id=doi:10.1023/B:VISI.0000011205.11775.fdhttp://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.72.9334{&}rep=rep1{&}type=pdf

[127] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO : Fast Semi-Direct Monocular Visual Odometry," in *IEEE International Conference on Robotics and Automation (ICRA)*. Hong Kong, China: IEEE, 2014, pp. 15–22. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=6906584{%}5Cnhttp://ieeexplore.ieee.org/ielx7/6895053/6906581/06906584.pdf?tp={&}arnumber=6906584{&}isnumber=6906581

[128] J. Engel, V. Koltun, and D. Cremers, "Direct Sparse Odometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 611–625, 2018. [Online]. Available: http://arxiv.org/abs/1607.02565

[129] M. Li and A. I. Mourikis, "High-precision, consistent ekf-based visual-inertial odometry," *International Journal of Robotics Research*, vol. 32, no. 6, pp. 690–711, May 2013. [Online]. Available: http://dx.doi.org/10.1177/0278364913481251

[130] P. Saponaro, S. Sorensen, S. Rhein, A. R. Mahoney, and C. Kambhamettu, "Reconstruction of Textureless Regions Using Structure from Motion and Image-based Interpolation," *Procedings of the International Conference on Image Processing, ICIP 2014*, pp. 1847–1851, 2014.

[131] L. Zhang, B. Curless, A. Hertzmann, and S. M. Seitz, "Shape and motion under varying illumination: unifying structure from motion, photometric stereo, and multiview stereo," in *International Conference on Computer Vision*. Nice, France: IEEE, 2003, pp. 618–625. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1238405

[132] Y.-p. Wang, Z.-l. Sun, Y. Qian, Y. Jing, and D.-x. Zhang, "An Accurate Online Non-rigid Structure from Motion Algorithm," in *International Conference on Intelligent Computing*, vol. LNAI 9277, no. 3. Fuzhou, China: Springer International Publishing, 2015, pp. 318–322. [Online]. Available: http://link.springer.com/10.1007/978-3-319-22053-6

[133] K. Chen and S. Lucey, "Prior-Less Compressible Structure from Motion," in *Computer Vision and Pattern Recognition*. Las Vegas, Nevada, USA: IEEE, 2016, pp. 4123–4131.

[134] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2281, 2012.

[135] J. Shen, X. Hao, Z. Liang, Y. Liu, W. Wang, and L. Shao, "Real-Time Superpixel Segmentation by DBSCAN Clustering Algorithm," *IEEE Transactions on Image Processing*, vol. 25, no. 12, pp. 5933–5942, 2016.

[136] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, Jun. 2008. [Online]. Available: http://dx.doi.org/10.1016/j.cviu.2007.09.014

[137] S. Leutenegger, M. Chli, and R. Y. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *Proceedings of the 2011 International Conference on Computer Vision*, ser. ICCV

'11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 2548–2555. [Online]. Available: http://dx.doi.org/10.1109/ICCV.2011.6126542

[138] O. Miksik and K. Mikolajczyk, "Evaluation of Local Detectors and Descriptors for Fast Feature Matching," in *Proceedings of the 21st International Conference on Pattern Recognition*, 2012, pp. 2681–2684.

[139] M. Irani and P. Anandan, "About direct methods," in *International Workshop on Vision Algorithms Corfu, Greece, September 2122, 1999 Proceedings*, vol. 1883, 1999, pp. 267–277. [Online]. Available: http://www.springerlink.com/index/EU4GVG556RH6D7B8.pdf

[140] R. Dragon, M. Shoaib, B. Rosenhahn, and J. Ostermann, "NF-features - No-feature-features for representing non-textured regions," in *European Conference on Computer Vision (ECCV)*, vol. 6312 LNCS, no. PART 2, 2010, pp. 128–141.

[141] P. Saponaro, S. Sorensen, S. Rhein, A. R. Mahoney, and C. Kambhamettu, "Reconstruction of Textureless Regions Using Structure from Motion and Image-based Interpolation," in *Proceedings of the International Conference on Image Processing, ICIP 2014*, Paris, France, 2014, pp. 1847–1851.

[142] K. Konolige, "Small Vision Systems: Hardware and Implementation," in *Robotics Research*, Y. Shirai and S. Hirose, Eds. London: Springer London, 1998, pp. 203–212.

[143] E. Prados and O. Faugeras, *Shape From Shading*. Boston, MA: Springer US, 2006, pp. 375–388. [Online]. Available: https://doi.org/10.1007/0-387-28831-7_23

[144] R. Basri, D. Jacobs, and I. Kemelmacher, "Photometric stereo with general, unknown lighting," *International Journal of Computer Vision*, vol. 72, no. 3, pp. 239–257, May 2007. [Online]. Available: https://doi.org/10.1007/s11263-006-8815-7

[145] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry," in *Computer Vision and Pattern Recognition*, vol. 1. Washington, DC, USA: IEEE, 2004, pp. 652–659. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=1315094

[146] A. Chiuso, P. Favaro, H. Jin, and S. Soatto, "Structure from motion causally integrated over time," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 523–535, 2002.

[147] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visualinertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015. [Online]. Available: https://doi.org/10.1177/0278364914554813

[148] E. Tola, V. Lepetit, and P. Fua, "DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo," *IEEE Transanctions Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 815–830, May 2010. [Online]. Available: https://doi.org/10.1109/TPAMI.2009.77

[149] D. P. Vassileios Balntas, Edgar Riba and K. Mikolajczyk, "Learning local feature descriptors with triplets and shallow convolutional neural networks," in *Proceedings of the British Machine Vision Conference (BMVC)*, E. R. H. Richard C. Wilson and W. A. P. Smith, Eds. BMVA Press, September 2016, pp. 119.1–119.11. [Online]. Available: https://dx.doi.org/10.5244/C.30.119

[150] K. M. Yi, E. Trulls Fortuny, V. Lepetit, and P. Fua, "Lift: Learned invariant feature transform," *Computer Vision - Eccv 2016, Pt Vi*, vol. 9910, pp. 17. 467–483, 2016.

[151] A. Bursuc, G. Tolias, and H. Jégou, "Kernel local descriptors with implicit rotation matching," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, ser. ICMR '15. New York, NY, USA: ACM, 2015, pp. 595–598. [Online]. Available: http://doi.acm.org/10.1145/2671188.2749379

[152] J. Dong and S. Soatto, "Domain-size pooling in local descriptors: Dsp-sift," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 5097–5106.

[153] G. Wang, Q. Zhou, and Y. Chen, "Robust non-rigid point set registration using spatially constrained Gaussian fields," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1759–1769, 2017.

[154] K. Simonyan, A. Vedaldi, and A. Zisserman, "Learning local feature descriptors using convex optimisation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1573–1585, Aug 2014.

[155] H. Mobahi, R. Collobert, and J. Weston, "Deep learning from temporal coherence in video," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML '09. New York, NY, USA: ACM, 2009, pp. 737–744. [Online]. Available: http://doi.acm.org/10.1145/1553374.1553469

[156] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, June 2006, pp. 1735–1742.

[157] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1. New York, NY, USA: IEEE, 2006, pp. 519–528. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=1640800

[158] G. Pascoe, W. Maddern, M. Tanner, P. Pini, and P. Newman, "NID-SLAM : Robust Monocular SLAM using Normalised Information Distance," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 158. Honolulu, Hawaii: IEEE, 2017, pp. 1446–1455.

[159] J. Heinly, E. Dunn, and J.-M. Frahm, "Comparative Evaluation of Binary Features," in *European Conference on Computer Vision (ECCV)*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds., Florence, Italy, 2012, pp. 759–773.

[160] Y. Furukawa and C. Hernández, "Multi-View Stereo: A Tutorial," *Foundations and Trends in Computer Graphics and Vision*, vol. 9, no. 1-2, pp. 1–148, 2015. [Online]. Available: http://www.nowpublishers.com/article/Details/CGV-052

[161] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski, "Non-rigid dense correspondence with applications for image enhancement," *ACM Transactions on Graphics*, vol. 30, no. 4, p. 1, 2011.

[162] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "DeepFlow: Large displacement optical flow with deep matching," *Proceedings of the IEEE International Conference on Computer Vision*, no. Section 2, pp. 1385–1392, 2013.

[163] J. Revaud and P. Weinzaepfel, "EpicFlow : Edge-Preserving Interpolation of Correspondences for Optical Flow," in *Computer Vision and Pattern Recognition*, 2015, pp. 1164–1172.

[164] B. K. P. Horn and B. G. Schunck, "Determining Optical Flow," *Artifical Intelligence*, vol. 17, pp. 185–203, 1981.

[165] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259 – 268, 1992. [Online]. Available: http://www.sciencedirect.com/science/article/pii/016727899290242F

[166] D. Ceylan, N. J. Mitra, Y. Zheng, and M. Pauly, "Coupled structure-from-motion and 3d symmetry detection for urban facades," *ACM Trans. Graph.*, vol. 33, no. 1, pp. 2:1–2:15, Feb. 2014. [Online]. Available: http://doi.acm.org/10.1145/2517348

[167] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[168] K. Law, G. West, P. Murray, and C. Lynch, "3-D Reconstruction of AGR Fuel Channels using RVI footage," in *6th EDF Energy Generation Ltd Nuclear Graphite Conference*, 2018.

[169] R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha, and M. Beetz, "Towards 3D Point Cloud Based Object Maps for Household Environments," *Robot. Auton. Syst.*, vol. 56, no. 11, pp. 927–941, Nov. 2008. [Online]. Available: http://dx.doi.org/10.1016/j.robot.2008.08.005

[170] V. Pratt, "Direct Least-squares Fitting of Algebraic Surfaces," in *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '87. New York, NY, USA: ACM, 1987, pp. 145–152. [Online]. Available: http://doi.acm.org/10.1145/37401.37420

[171] A. Locher and L. V. Gool, "Progressive Prioritized Multi-view Stereo," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, Nevada, USA: IEEE, 2016, pp. 3244–3252.

[172] S. Lee, J. Yoo, Y. Kumar, and S. Kim, "Reduced energy-ratio measure for robust autofocusing in digital camera," *IEEE Signal Processing Letters*, vol. 16, no. 2, pp. 133–136, Feb 2009.

[173] A. Thelen, S. Frey, S. Hirsch, and P. Hering, "Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood-size, and height value interpolation," *IEEE Transactions on Image Processing*, vol. 18, no. 1, pp. 151–157, Jan 2009.

[174] R. Minhas, A. A. Mohammed, Q. M. J. Wu, and M. A. Sid-Ahmed, "3d shape from focus and depth map computation using steerable filters," in *Image Analysis and Recognition*, M. Kamel and A. Campilho, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 573–583.

[175] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," in *Proceedings of the 18th ACM International Conference on Multimedia*, ser. MM '10. New York, NY, USA: ACM, 2010, pp. 1469–1472. [Online]. Available: http://doi.acm.org/10.1145/1873951.1874249

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

[177] K. Yamaguchi, "kyamagu/mexopencv," Oct 2018. [Online]. Available: https://github.com/kyamagu/mexopencv

[178] T. Lee, "Vlg toolbox." [Online]. Available: http://vision.ucla.edu/vlg/