Synthetic data for damage assessment in aircraft turbines

Eduard Klein Onstenk[1] , **Jan van Gemert**[1] , **Burak Yildiz**[1]
[1]TU Delft
e.

Abstract

This paper discusses possible ways to generate synthetic data and its use cases for damage assessment in aircraft turbines. Synthetic data has many advantages such as exact ground truth and scalable data sets. Using SLAM and SfM, which are 3D construction tools, 3D models can be constructed from 2D monocular borescope videos. A 3D reconstruction of parts of the engine allows us to measure the damage if it exists. But when is the synthetic data ”good”? Using the methods SLAM and SfM synthetic data could possibly be evaluated by comparing it to real data. Using Blender, synthetic borescope videos are generated and the performance of SLAM and SfM on these videos is compared to the real videos. In general, there are many different use cases for synthetic data in damage assessment and there are multiple ways to generate the right data set. Evaluating synthetic data shows that synthetic data that qualitatively looks closer to real data does not perform closer when running SfM or SLAM on it.

1 Introduction

The inside of an aircraft engine is difficult to inspect since the blades of the propeller are often encased by a narrow construction. Yet it is vital that all components of the engine are in adequate condition to ensure safety for usage of the aircraft. To permit that technicians are still able to evaluate the condition of the the engine even on the inside, a borescope is used. This instrument is slipped into the engine and produces monocular video data that the technicians can inspect. This is quite a time consuming and error prone task, especially when it has to be done often. The risks that occur from this visual inspection task are identified in [1]. This all stimulated use of computer vision to aid the technicians.

However, to increase the performance of the used computer vision technologies data is needed for both evaluation as well as training purposes. Borescope videos of aircraft engines are not widely available, which is not uncommon for industrially applicated videos. More so, little to no ground truth data is available for these videos. This motivates the use of synthetic data, to compensate for the gap in data and have readily available ground truth data from the model. If synthetic data that behaves similarly to real data from borescope videos of aircraft engines could be generated then this allows to increase evaluation abilities of different algorithms their performance and to generate training data.

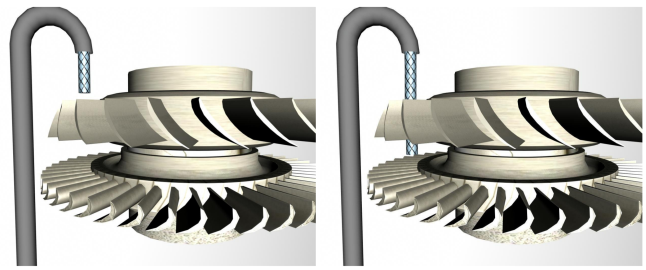


Figure 1: Example of borescope inspection scenario. The smaller braided cylinder follows a possible borescope trajectory, in each position the respective turbine is rotated to inspect the blades. Obtained from [26]

As the performance of supervised computer vision methods is greatly dependant on quality and quantity of data sets, the academic community has shown great interest in synthetically satisfying the need for more data. It can be used to simulate conditions which are not found in current data, train deep learning algorithms, have ground truth available from the model and to evaluate performance of algorithms. The research questions this paper aims to answer are:

1. How can synthetic data be created and used for borescope inspections of aircraft engines?
2. How do SLAM and SfM compare on synthetic borescope videos of aircraft engines?

Since the research exists of two partially separate questions, this paper will partly be a literature study and partly experiment with self generated synthetic borescope video data.

This paper is structured as follows. Section 2 ”Background” will go in depth on the most critical concepts needed to understand the rest of the paper. After which ”Related Work” in 3 is used to briefly mention and discuss similar studies, it is kept brief as a more in depth literature study follows. Then the ”Methodology” is briefly described in 4. 5 ”Synthetic Data” aims to answer the first question through a literature study, 5.3 motivates use of SLAM and SfM on synthetic data. 6 is used to demonstrate the experimental setup and results of this research. Then section 7 and 8 will cover ”Responsible Research” and the discussion respectively. Finally, the conclusion of this paper is presented in section 9 and ”Future Work” is discussed in section 10.

2 Background

This section is structured as follows. First in section 2.1 the two methods to evaluate on synthetic data are elaborated upon. Then in the next section 2.2 goes into the feature matches that the methods described in 2.1 use.

2.1 SLAM & SfM

Both SLAM [21] (Simultaneous Localization And Mapping) and SfM [20] (Structure from Motion) are algorithms used to estimate 3D objects from 2D film. However, when the surface lacks texture or is shiny SLAM and SfM struggle to estimate a proper 3D model.

There are several varying applications for SLAM, ranging from Unmanned Aerial Vehicles (UAV) to vacuum cleaner robots. What they all have in common is in its name, the algorithm strives to simultaneously map and steer the environment. Utilizing images to decide location and orientation in computer vision is captured by the term visual odometry (VO). For SLAM, this means that whilst a 3D reconstruction/mapping is made, the algorithm keeps track of the location of the view. This is useful for autonomous systems, because it enables the system to navigate an environment it has no data of beforehand. [9] proposed a system that showed that VO can be used for "pose estimation and dense depth map estimation" even from a monocular feed. To ensure a real time usage of SLAM often algorithms count on estimations, which emphasizes the use of usability and functionality rather than precision.

While SfM is used in similar applications and also aims to localize the view and get 3D points, the main difference lies in that it does not require real-time. Since SfM can use all data at once, more exact and denser 3D reconstruction can be achieved at the cost of needing more processing power. The relevant results for the scope of this paper that both methods produce are a 3D point cloud representing the three-dimensional model and a camera trajectory.

To summarize, a big advantage of SLAM over SfM is that SLAM can be done real time. This means that this technology could speed up the process of borescope inspection of aircraft engine blades if it works well. SfM however requires more precision, which is also a trait required for damage assessment. It is exactly this trade-off that decides which algorithm works best for our specific use case. Using synthetic data we are able to discover which of the two methods works best under what conditions in borescope videos, which allows to decide which algorithm to use then.

2.2 Feature Extraction & Matching

Both methods SfM and SLAM require feature extraction and matching of the extracted features in order to create a 3D model. From every image in a video a certain set of features is extracted, the quantity of which can vary extremely depending on the quality of the video and the algorithm that is used. The way SLAM and SfM deal with the extracted features is a little different, as became clear from section 2.1. To match features in between frames, SfM could potentially compare every single feature from every single image to all other features of all other images. SLAM however, is real time, which means that every frame can only be compared to all previously analyzed frames and their features.

The resulting features and matches can then be filtered in different ways such as removing outliers, depending on whether you are using SLAM or SfM and what version. The main algorithms used for feature extraction in this paper are SuperGlue which is a neural network introduced by [19] and LoFTR from [24], both further researched in this context by Huizer in [8]. ORB and SIFT, the last two being the two standard algorithms used by SLAM and SfM respectively in the research by Markhorst in [12] and Nonnemaker in[14].

3 Related Work

This research lies in the general space of computer vision, covering constructing synthetic data sets, visual odometry, mapping, interest point matching and training neural networks using synthetic data. Particularly relevant areas for the research conducted in this paper include research on generating and/or using synthetic photo/video data, specifically monoscopic. As well as is research on damage assessment from borescope videos of aircraft engines. Furthermore it is relevant what features make "good" synthetic data, or how to scale the amount of data that is generated since many applications require massive data sets. There is also similar research conducted by my colleagues who dive into other aspects of improving computer vision aided borescope video inspection of aircraft engines. In [14], research is conducted on the performance of SfM on the borescope videos. Research alike but for SLAM is covered in [12]. Finally [8] researched interest point detection and matching for shiny and non-textured propeller surfaces and [23] researched how to gain ground truth data, a problem which is also mentioned throughout this paper.

Both for SLAM as well as SfM research has been conducted on applications of synthetic data for improving the methods. [27] motivates the creation of a synthetic data set to keep up with the fast development of SLAM and thus the need for images and ground truth from different settings. [5] shows that in their research synthetic data performs in a similar manner to real data. They use this to evaluate performance of different algorithms used in steps of SfM using hundreds of generated data samples, as well as calculating 3D point error of the maps.

Similar research was conducted by [13], which aims to answer what makes synthetic data good for learning disparity and optical flow estimation, which have much overlap with how SLAM and SfM operate. The paper investigates usage of synthetic data for both benchmark as well as training purposes. [5; 27] Cover research on synthetic data in combination with SfM and SLAM, the used techniques differ however from research in this paper.

Contributions: Throughout the literature there are many papers that cover creating synthetic data sets, or using synthetic data in order to train or evaluate performance of algorithms. However there is little research on the comparison of SLAM versus SfM, especially in context of synthetic data. This is an interesting topic to research because as explained in section 2, they both have advantages and disadvantages dependant on the use case. Using synthetic data, their stronger and weaker points can be researched more extensively, largely due to ground truth knowledge. Furthermore this paper covers a wider variety of techniques through which synthetic data can be used to discover damage in aircraft turbines. Finally this paper covers research on evaluating the performance of synthetic data using the methods SLAM and SfM.

4 Methodology

The research questions were "How can synthetic data be created and used for borescope inspections of aircraft engines?" and "How do SLAM and SfM compare on synthetic borescope videos of aircraft engines?". The aim of the research was to find out how to create synthetic data and what the possible use cases are, as well as to generate synthetic borescope videos that share similarity to the real data, such that conclusions can be drawn from testing the algorithms on those synthetic videos. These questions were answered through a combination of literature research followed by some experiments on self generated synthetic video data. The experiments compared changes in synthetic data, which SLAM and SfM were ran on to see the effect of those changes on the respective methods. The experimental data as a result from the tests were a mixture of descriptive data to analyze the difference between methods SLAM and SfM as well as data resulting from manipulation in variables such as background light and texture.

A literature study is conducted in section 5. The aim of this literature study was to answer the first research question "How can synthetic data be created and used for borescope inspections of aircraft engines?". The results can be found in that section and the discussion, both qualitatively described. Then the second research question was answered, "How do SLAM and SfM compare on synthetic borescope videos of aircraft engines?" together with the other part of the first research question in the remainder of the paper. To answer this, the methods were be ran on varying synthetic borescope videos varying in specifications such as blade texture and light levels. Their evaluation being mainly qualitative as well, expressing the differences in performance mainly from what can be seen from the resulting 3D point clouds, such as sparsity and outliers where a denser cloud is mostly associated with a more accurate model.

All synthetic data was created using the 3D modelling software "Blender" [1], which is a common approach in the field of computer vision and related work. Using Blender my own synthetic data set was created to run the experiments on.

Example borescope videos provided by the company 'Aiir' [2] were used to base the synthetic videos off of. As well as the fan used in my computer and a miniature model of the bottom of some blades at aiir. These were used because although they don't look exactly alike, they are quite similar yet much simpler and more abstract yet capture the shape of the blades, which are useful traits in synthetic data.

[1] Blender open source 3D creation: https://www.blender.org/

[2] Aiir Innovations: https://aiir.nl/

5 Synthetic Data

In this section the first question of this research is split into the sub questions: "How to generate synthetic data?" and "What are the use cases for synthetic data?". These are answered through a literature research, this is done respectively in subsection 5.1 and section 5.2. Then in section 5.3 the experiments (sections 6.1, 6.1 and 6.1) of using methods SLAM and SfM to evaluate created synthetic data are introduced and motivated.

5.1 Use Cases

This subsection aims to answer how the created synthetic data can be used within the context of borescope videos of aircraft engines. Studying the possible use cases comes first as the application of the data greatly influences the way it should be generated. The possible use cases further motivate the research on how to generate synthetic data.

Improving interest point detection. Many interest point detection or matching algorithms use a form of neural network [6; 17; 19]. It is known that training neural networks to be accurate whilst preventing overfitting requires large data sets, which are difficult to get of borescope videos. In the context damage assessment of aircraft engines, where precision is key, manually annotating the data can result in inaccuracies. "Human judgment of interest points is subjective by nature." [7], this research shows how different annotators of interest points select different points. If the data set for training the neural network gets large multiple annotators would definitely be needed for real data, which is costly, time consuming and inaccurate. Apart from these reasons to use synthetic data, [6] concluded that training on synthetic data does generalize to real data. How much however needs to be researched (see future work 10). [13] Show in their paper that diverse data sets generalize well to other data sets and that specialized data does not. Further research would be required to decide how specialized the data set to train interest point detection networks would have to be.

Improving damage detection networks. [2] avoided the use of neural networks by falling back on other computer vision approaches to automatically detect damage in aircraft engines. The reason for abstaining from neural networks was because the required data sets would be too large. [22] Did use the neural network approach for automatic damage detection and although the two papers use different metrics for evaluating results, qualitatively the neural networks results seem to be more accurate. The latter research did not use synthetic data in training but had experts label the data. This implies that if synthetic data generalizes well to real data then the performance could be improved because of a larger data set. In [10] a more specialized neural network was developed, specifically for crack detection.

Training personnel. Synthetic data of aircraft engines could improve the skills that borescope inspectors need. [26] developed and researched a haptic virtual borescope, which could be used to train aircraft engine inspectors. Their research on the simulator did indicate however that workload remains the same and their results only include a slight improvement in speed of the inspection process. More success was found in further research [25], where an improved simulator was developed. One experiment concluded that 'constraining camera rotation leads to significantly fewer collisions of the camera with virtual environment surfaces.'. This in term prevents damage to the tip of the probe of the borescope. Similar research, [18], uses a virtual reality setup to teach students how to use a borescope.

Assess performance of algorithms. [27] presents synthetic data which naturally has ground truth available to evaluate newly developed visual SLAM methods. Some evaluation criteria for SLAM performance are proposed, relevant ones for this research include localization error of the camera, reconstruction error and outlier removal capability. However their data set is limited to underground garages, thus the set is not useful for this research. [5] introduces two new methods for evaluating the performance of SfM. 'Calculating 3D point error' and 'comparing SfM as 2D feature noise varies'. Finally, [11] proposes a synthetic data set for "evaluation of 3D reconstruction pipelines" such as SLAM and SfM discussed in this paper. Although it explains how the set was created, it is not always clarified why certain decisions are made.

5.2 Generating Synthetic Data

There are multiple tools that allow you to generate synthetic data. In [5] synthetic data is gathered from an implementation using the game engine Unity3D. [16] used the same strategy but another popular game engine, namely Unreal game engine. Their research resulted in a follow up study [28] which used the same previous technique to improve feature matching on textureless and specular surfaces. Their proposed tool can be used with multiple publicly accessible game models, which makes it generally applicable for many applications.

[4] Automatically generates large portions of their data set using the 3D modelling software tool Blender. [3] used Blender to extract ground truth data for the open source Sintel film. This allowed them to render the same scene under different conditions over and over and in term analyze where algorithms to evaluate optical flow fail. Mayer [13], researched multiple ways to generate data in Blender for training neural networks to improve optical flow. In [11] Blender is used to generate data to compare different 3d reconstruction algorithms, similarly to this paper. Comparable software 'Maya' which is not free to use like Blender is used in research on optical flow such as [15]. Many of the related research is focusing on optical flow. This is quite similar as optical flow also requires feature matching, optical flow uses it to compute motion, SLAM and SfM focus more on camera movement.

Diversity matters in training data for neural networks, meaning that diverse data may perform better as a training set than specialized data [13]. This research further concludes: realism is not all, thus the aim of generating synthetic data should not necessarily be for it to look as realistic as possible. Order of data fed to neural networks matter and knowledge of the camera and lens and their flaws improves networks. According to their paper, this holds for disparity estimation and optical flow, which could be generalizable interest point matching from monoscopic video, since these two are closely related in the field of computer vision.

Mayer further distinguishes three different ways of generating synthetic data.

- Use existing scene data
- Designing new scenes manually
- Create randomized scenes in procedural manner.

For this paper existing scene data is little to not available. Thus data is generated by designing new scenes manually. Once the data is evaluated and it performs similar to real data, randomized scaling techniques can be applied to create potentially infinite data. The easiest way to do this with the models from this research would be to use random transformations of camera angle and positions, with known ground truth data. Changing actual shapes in the model is harder as this might influence the performance of the data, more research would be required to answer this.

5.3 Evaluation using SLAM & SfM

In section 5.2 some of the synthetic data that has been generated throughout the research has been introduced. However, how can one tell whether the data is any good? In section 5.1, the idea that synthetic data can be used to assess the performance of some algorithms such as SLAM and SfM is discussed. A recurring argument in motivating this approach is that naturally, ground truth comes with the models from which synthetic videos are generated.

There are several factors that motivate the use of synthetic data for damage assessment of aircraft engines. First, both SLAM and SfM require interest point matching yet there are no public neural networks trained on borescope like videos of turbines. Neural networks require large amounts of training data and that is not available. Secondly, synthetic data can aid us to understand under what circumstances SLAM methods perform better than SfM methods and vice versa, by making minor changes to the same models.

The second research question was "How do SLAM and SfM compare on synthetic borescope videos of aircraft engines?". This is not only to find out how well SLAM and SfM perform on synthetic data, but also to use their respective performances to compare the synthetic data to real data. Some research is already available on how SLAM and SfM compare on real data, [12] and [14] respectively. As its known what results SLAM and SfM produce under certain circumstances, this motivates the approach of using SLAM and SfM to evaluate the correctness of our synthetic data. If SLAM and SfM perform similarly on real data as on synthetic, this could indicate that the synthetic data is "good". The motivation for using both SLAM and SfM is that while both methods aim to achieve similar results they operate in a different manner. If one performs well on synthetic data by coincidence but the other does not this could indicate poor performance of the generated data set.

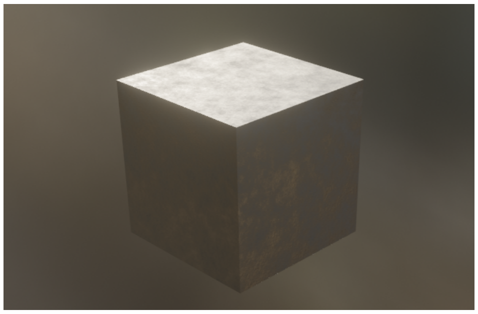


Figure 2: A metallic material with shiny, textured and less textured areas. This material will serve as the base material on the next models, where texture and shininess can be altered.

Figure 3: Example frame of a real borescope video supplied by Aiir. Although videos differ, this provides insight how synthetic data should look.

Figure 4: Frame of first version of rotating synthetic data. Highly reflective surface and little texture.

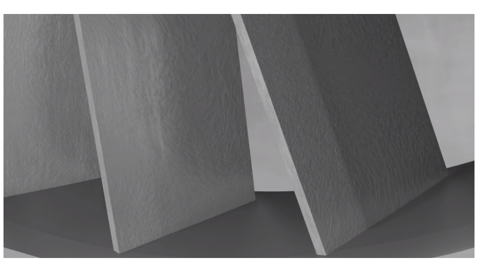


Figure 5: Frame of second version of rotating synthetic data, much more texture than most real videos for testing purposes.

6 Experimental Setup & Results

In this chapter first details of the evaluation setup are described in section 6.1. Its subsections describe the experiments that were conducted with the models. Then the results are reported in section 6.2.

6.1 Experimental Setup

Before diving into the experiments in, the basic model versions are described first, defining their underlying ideas and the workflow of creating them.

Since the ability to create a 3D model from 2D footage of an object depends greatly on the texture and shininess on the surface of that object, first the material for the model was developed. For the experiment, a shiny material with about an even distribution of textured and less textured area was created on a cube. For the first view models the colour of this material will be kept grey, however for the final model using this material would have a beige undertone to represent the real borescope videos supplied by Aiir more.

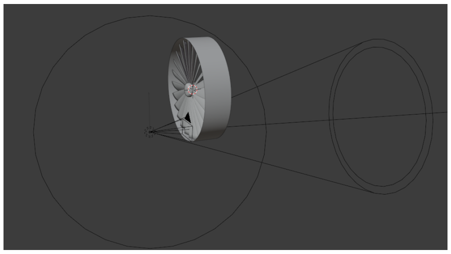
Then a basic model to apply the material on was developed. Version 1 has blades shaped as cuboids, as can be seen in 4. The last version has much more realistically shaped blades, with curves, based on actual blades. Now that the basic material has been developed, the first model is created. For this research Aiir sent 10 different borescope videos, an example of a single frame of such a video can be seen in 3. However the videos are too different to develop a single 3D model that would be similar to most. The videos differ in camera angle, shape of blades, amount of texture and more. Because of this, the model with curved blades is modeled similar to the top section of figure 1, which looks more similar to a fan than the bottom.

As discussed in section 5.2, there exist multiple ways to generate synthetic data. For the experiments conducted in this research, the manual modelling approach [13] was chosen as models could precisely be constructed to be similar to a specific borescope video. All rendered videos result in a 1280x720 display resolution, with 30 fps.

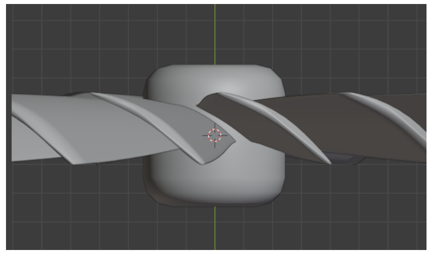
A frame from the first synthetically generated video is shown in figure 4. The aim for this video was for the footage to be much more abstract than the real borescope videos. An assumption was that a monotone pink background would result in less wrongful matches in the background. The model is similar to a watermill, which was easy to create and still looks a lot like the aircraft engine. The blades are cuboid, reflective, metallic like and don't have a lot of texture.

After the feature extraction failed from figure 4, figure 5 was created which from the same model, but has a lot more texture on the blades and much more convenient lighting. Since feature extraction was done with a colourless background, the background is changed to gray instead of pink.

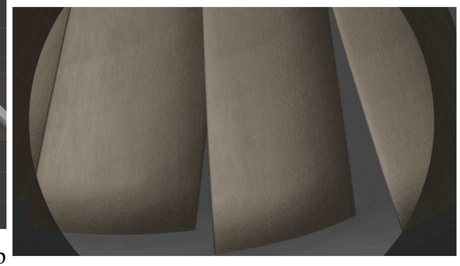
Now a more complex model is constructed 6c, where the shape of the blades is much more similar to that found on a turbine or a fan 6b. As was clear in 6c, the lighting conditions are now much more similar to that of real borescope videos, that have a light at the tip right next to the camera. 6a shows how this is achieved, namely by having a cone shaped area light in the same spot pointed in the same direction as



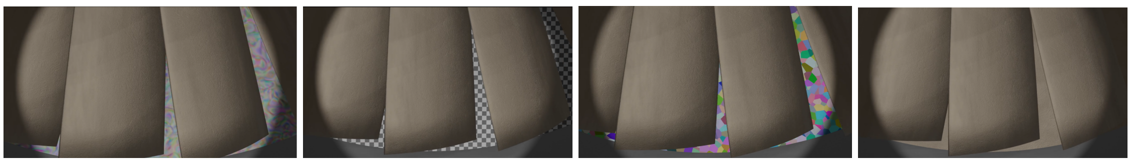
(a) Final comparison model light and camera setup. Shows how the synthetic video data was captured.



(b) Final comparison model close-up blades. Provides a better understanding of how the blades are shaped in the final model.



(c) Final model gray background single frame. Average texture and lighting, as described in experiment 3.



(a) (b) (c) (d)

the camera view. This creates the same round light spot surrounded by shadow that can be found in some of the real borescope videos.

Experiment 1: shape of blades

The first experiment aims to tell how much impact the shape of the propeller blade has on the resulting model. This is to see whether more abstract models get similar results as more complex models when running SLAM or SfM on them. To analyze this, the models with cuboid shaped blades are 3D reconstructed using both SLAM and SfM. The same goes for the new model with curved blades. The only difference between the two models is the shape of the blades. The texture, lighting and background remain the same for both. The background is monotone gray. Then the level of texture is altered by varying the roughness of the noise texture.

Experiment 2: different textures background

Then the next experiment analyzes the influence of the background in a borescope video. For this experiment the goal is too to analyze whether data more similar to real data also performs more similar under SLAM and SfM. To do this the final model was used with 4 additional textures as background, added onto a plane. Frames from this batch are shown in [a;b;c;d] above. Apart from the background the videos are exactly the same. The first background is a noise texture from Blender. This is a colourful and blurry background. The second is a simple checker background with white and gray, containing sharp lines and lots of corners. The third background is the Voronoi texture, containing multiple colours and straight lines, but lots of different random angles. Finally the last model has a background texture which is the same texture as on the blade, which is more similar to real borescope videos.

Experiment 3: lighting, texture and speed

The final experiment aims to answer how SLAM and SfM perform on synthetic data when making minor adjustments to the data. The three variables that are altered are the amount of light, the amount of texture and the speed of the rotation of the blades. From (6b, 6a), a batch of 7 videos was created, all three seconds long such that testing fit within time frame of this research. The videos differed in three aspects, namely texture on the blades, lighting conditions and speed of rotating blades. The texture on the blades was altered by changing the roughness of the texture between 0.5, 0.75 and 1.0. The light was changed by altering the power of the lamp, lowest level being 500 Watt, medium 1000 Watt and high being 1500 Watt. Finally the rotary speed of the blades was 90, 120 or 150 degrees per 10 seconds. Whilst differing one setting, the other two are always on the median value, resulting in 7 videos.

6.2 Results

The results of the first experiment show that while the abstract cuboid shaped blades result in qualitatively good model in both SLAM and SfM, see Figure 8 for the performance of SLAM. However for the new model with curved blades, also using SuperGlue, no model could be created because there were too little features detected. The result was an empty point cloud.

The second experiment ran with SuperGlue on which the real data performs well, also did not succeed in the more complex model with curved blades and more realistic colour and texture. In table 1, the amount of interest points per frame for each background are listed as an average over all frames. The last column indicates what values can be expected on average if real data were used from the videos supplied by Aiir.

When running the feature extraction and matching algorithm on the batch of 7 videos (for example 6c), SIFT works great in all instances. However, SuperGlue results in an extremely low amount of matches, namely 20 per frame. From this amount no useful model can be generated from either SLAM or SfM. On real data SIFT works bad and SuperGlue performs well.

Nonnemaker researched a multi-view stereo method for SfM, which is ran after running SfM. Using SIFT, it is shown in figure 8 that although synthetic data does not perform well in the experiments above, it is possible to create very realistic models [14].

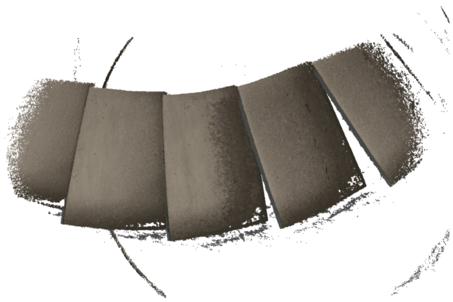


Figure 8: SfM using SIFT in combination with multi-view stereo shows that promising 3D models can be constructed from the synthetic data used in the experiments.

	Gray	Noise	Checker	Voronoi	Metal	Real
SuperGlue	312	248	466	428	93	400-800
SIFT	612	560	1420	1254	577	1500
LoFTR	9980	-	-	-	-	5k

Table 1: # of features per frame on average using different backgrounds. SuperGlue does not detect many features on the synthetic data

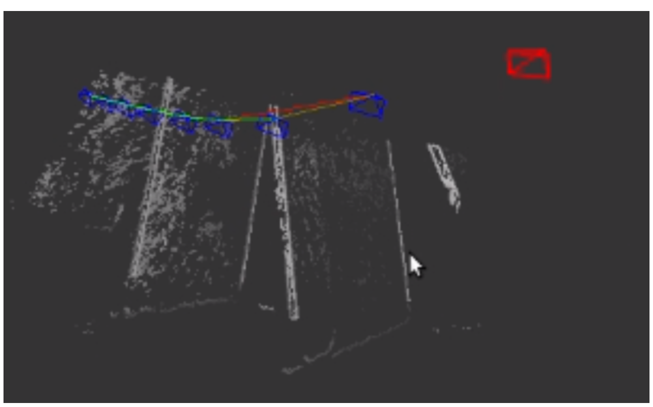


Figure 9: Result SLAM on figure 5

7 Responsible Research

There are a couple of ethical implications that follow from the research described by this paper. The two largest implications regard safety in using the technology and human replacement. A couple of use cases for synthetic data were introduced in section 5.1. One use case is to use synthetic data to train neural networks. Using synthetic data for training a neural network requires good validation data, especially since it is not always what features neural networks pick up on. It should also focus on having little to no false negatives as a result, since those are the most dangerous. Another use case was to evaluate the performance of algorithms such as SLAM and SfM. Synthetic data could aid evaluation of performance but it should not rely solely on synthetic data, as a too trusting attitude towards these algorithms without sufficient testing on real data could result in faulty aircraft engines going unnoticed during checks. Vice versa, conclusions of the performance of synthetic data evaluated using SLAM and SfM as a result from the experiments should be interpreted with care and more research is needed to see how this translates to performance as a training set for neural networks for example.

Secondly, one could argue that this research is used to automate a task that could replace a human. However, improving borescope video inspection improves safety of people on airplanes and is merely a tool to aid an inspector. For now the final responsibility should be at the human level and not the system level.

Finally, a word on the reproducability of the results of this paper. The conducted research consists partially of a literature research and partially of experiments described in 6. The most important claims from the literature research are all coupled to references that can be checked by the interested reader. As for the experiments, one could try to recreate the synthetic data as I have. All the synthetic data, including the models, textures and videos used in this research are available by request. Although the experiments include methods to quantitatively analyze results, performance is qualitatively evaluated as well, one might run the same experiments but conclude something different.

8 Discussion

This research includes three main experiments which results are discussed in this section. The purpose of experiment 1 was to identify what the influence of the shape of the propeller blade is on the results from SLAM and SfM, which are qualitatively analyzed by looking at the resulting model. From the resulting models for both SLAM and SfM it became clear that it is a lot easier for these methods to reconstruct the cuboid shaped blades with long straight edges. The synthetic data with blades with a more curved shape did not perform well, even when massively increasing the amount of texture on the blades. Only with SfM using multi-view stereo [14] or using other feature extraction/matching algorithms, promising results were found yet this is not consistent with how the real data behaves. From this its clear that the more abstract model with cuboid shape is easier to model, but more further research is required find out how more similarly behaving synthetic data can be created.

From the second experiment, where different backgrounds were tested, results are both quantitatively and qualitatively analyzed. Table 1 shows the amount of features different feature detection algorithms detect on average per frame. Here it became clear that the synthetic data behaves differently from the real data. Although LoFTR works a lot better on this data, that is not what is wanted, as the goal was to have synthetic data that behaves similarly.

For experiment 3, it also became apparent that for the test set of 7 videos, SIFT worked great as a feature extraction and matching tool however SuperGlue did not. This is the opposite in real data, but the test set is the most realistic looking test set yet. Here further research is also required to find out why. This could possibly be because the neural network SuperGlue takes context more into consideration than texture, which works great for the real data but does not work on the test batch [8]. Context lacks, background has no texture and the only real observations are the blades. The opposite is true for SIFT, this method works great on textured surfaces but does not consider context. This could explain why it would work so well on 6c.

Because in general the results were not good, there was little reason to use quantitative methods to analyze the results.

In section 5 it was discussed how to generate synthetic data and what possible use cases are for damage assessment in aircraft turbines. Four main potential use cases were discussed, improving interest point detection, improving damage detection networks, training personnel and assess performance of algorithms. For the last two, existing literature already shows benefits. The first two require further research to evaluate whether there are benefits to using synthetic data.

Then in section 5.2 it was discussed how to generate synthetic data. For the experiments only manual modeling in Blender was used, however the literature study does cover other methods that can scale the amount of data generated. This could prove to be very useful for future research but its application is outside of the scope of this research.

Limitations The fact that there was no access to aircraft engine models that were used, for example in the video, was a limiting factor in creating models.

In section 5.1, multiple use cases from existing literature were mentioned. However not all were considered when generating the data. They were left in this research as they help answer how synthetic data could be applied in the context of damage assessment from the borescope videos, but creating synthetic data for these was outside of the scope of this paper.

9 Conclusions

The main research questions of this paper were "How can synthetic data be created and used for borescope inspections of aircraft engines?" and "How do SLAM and SfM compare on synthetic borescope videos of aircraft engines". A literature study conducted investigated how to create synthetic data and how this could be used for borescope inspection. Then an experiment was run aimed to evaluate the quality of self generated synthetic data.

The literature study found multiple ways to generate synthetic data. The most important findings on generating synthetic data were that synthetic data allows for large test sets and readily available ground truth data, two valuable attributes. It was also found that within the context of borescope inspections of aircraft engines synthetic data has many different potential use cases, the main potential use cases being:

- Improving interest point detection
- Improving neural networks
- Training personnel
- Assess performance of algorithms

Finally the methods SLAM and SfM were used to evaluate self generated data. Although qualitative and quantitative methods to evaluate synthetic borescope video data compared to real borescope video data were proposed, this research was unable to consistently produce synthetic data that behaved similar to real data when running SLAM or SfM on it. Using feature matching and detection algorithms that did not perform well on real data did show promising results on the synthetic data, especially after applying multi-view stereo after creating a model with SfM.

10 Future Work

More research is required on why the more complex synthetic models did not perform similar to the real data, to properly conclude the second research question.

There are multiple reasons to develop synthetic data for related applications. One reason which was only discussed in a literature study is to train the neural networks for interest point matching such that they can construct better 3D models. However, as the used neural networks for interest point matching used in this project are only public pre-trained testing this on my own synthetic data was outside of the scope of this project. This would be an interesting application to do further research on in the future however, specifically to research what features are extracted from real data by using more abstract synthetic models.

Another interesting topic to do more research on would be the use of neural networks for classification of damage to blades such as the one described in [22]. I am interested whether these neural networks can be utilized to achieve accurate measurements, instead of merely detecting the damage.

References

[1] Jonas Aust and Dirk Pons. Bowtie methodology for risk analysis of visual borescope inspection during aircraft engine maintenance. *Aerospace*, 6(10), 2019.

[2] Jonas Aust, Sam Shankland, Dirk Pons, Ramakrishnan Mukundan, and Antonija Mitrovic. Automated defect detection and decision-support in gas turbine blade inspection. *Aerospace*, 8(2), 2021.

[3] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *European Conf. on Computer Vision (ECCV)*, Part IV, LNCS 7577, pages 611–625. Springer-Verlag, October 2012.

[4] Chen Chen, Xin Jiang, Weiguo Zhou, and Yun-Hui Liu. Pose estimation for texture-less shiny objects in a single RGB image using synthetic training data. *CoRR*, abs/1909.10270, 2019.

[5] Joseph DeGol, J. Lee, Rajbir Kataria, Daniel Yuan, T. Bretl, and Derek Hoiem. Feats: Synthetic feature tracks for structure from motion evaluation. *2018 International Conference on 3D Vision (3DV)*, pages 352–361, 2018.

[6] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[7] Helin Dutagaci, Chun Pan Cheung, and Afzal Godil. Evaluation of 3D Interest Point Detection Techniques. In H. Laga, T. Schreck, A. Ferreira, A. Godil, I. Pratikakis, and R. Veltkamp, editors, *Eurographics Workshop on 3D Object Retrieval*. The Eurographics Association, 2011.

[8] Rick Huizer. How well does interest point detection/matching work on shiny and non-textured surfaces?, 2021.

[9] Ruihao Li, Sen Wang, Zhiqiang Long, and Dongbing Gu. Undeepvo: Monocular visual odometry through unsupervised deep learning. *CoRR*, abs/1709.06841, 2017.

[10] Yadan Li, Zhenqi Han, Haoyu Xu, Lizhuang Liu, Xiaoqiang Li, and Keke Zhang. Yolov3-lite: A lightweight crack detection network for aircraft structure based on depthwise separable convolutions. *Applied Sciences*, 9(18), 2019.

[11] Ciocca G. Marelli D, Bianco S. *IVL-SYNTHSFM-v2: A synthetic dataset with exact ground truth for the evaluation of 3D reconstruction pipelines.*, 2019.

[12] Thomas Markhorst. Performance analysis of simultaneous localization and mapping used to reconstruct aircraft engines in 3d, 2021.

[13] Nikolaus Mayer, Eddy Ilg, Philipp Fischer, Caner Hazirbas, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. What makes good synthetic training data for learning disparity and optical flow estimation? *International Journal of Computer Vision*, 126(9):942–960, Apr 2018.

[14] Alec Nonnemaker. Evaluating structure-from-motion on shiny and non-textured surfaces in borescope videos, 2021.

[15] & Sappa A. D. Onkarappa, N. *Speed and Texture: An Empirical Study on Optical-Flow Accuracy in ADAS Scenarios*, 2014.

[16] Weichao Qiu and Alan L. Yuille. Unrealcv: Connecting computer vision to unreal engine. *CoRR*, abs/1609.01326, 2016.

[17] Jérôme Revaud, Philippe Weinzaepfel, César Roberto de Souza, Noé Pion, Gabriela Csurka, Yohann Cabon, and Martin Humenberger. R2D2: repeatable and reliable detector and descriptor. *CoRR*, abs/1906.06195, 2019.

[18] T. Rupasinghe, M. E. Kurz, C. Washburn, and A. Gramopadhye. Virtual reality training integrated curriculum: an aircraft maintenance technology (amt) education perspective. *International Journal of Engineering Education*, 27:778–788, 2011.

[19] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. *CoRR*, abs/1911.11763, 2019.

[20] J. L. Schonberger and J. Frahm. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4104–4113. 2016.

[21] Z. Shang and Z. Shen. *Construction Research Congress 2018*, pages 305–315. 2017.

[22] Zejiang Shen, Xili Wan, Feng Ye, Xinjie Guan, and Shuwen Liu. Deep learning based framework for automatic damage detection in aircraft engine borescope inspection. In *2019 International Conference on Computing, Networking and Communications (ICNC)*, pages 1005–1010, 2019.

[23] Devin Lieuw A Soe. Ground truth for evaluating 3d reconstruction of jet engines, 2021.

[24] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. Loftr: Detector-free local feature matching with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8922–8931, June 2021.

[25] Deepak Vembar, Andrew T. Duchowski, Anand K. Gramopadhye, and Carl Washburn. Improving simulated borescope inspection with constrained camera motion and haptic feedback. In *Proceedings of Graphics Interface 2009*, GI '09, page 141–148, CAN, 2009. Canadian Information Processing Society.

[26] Deepak Vembar, Andrew T. Duchowski, Sajay Sadasivan, and Anand K. Gramopadhye. A haptic virtual borescope for visual engine inspection training. In *2008 IEEE Symposium on 3D User Interfaces*, pages 19–26, 2008.

[27] Senbo Wang, Jiguang Yue, Yanchao Dong, Shibo He, Haotian Wang, and Shaochun Ning. A synthetic dataset for visual slam evaluation. *Robotics Auton. Syst.*, 124:103336, 2020.

[28] Yi Zhang, Weichao Qiu, Qi Chen, Xiaolin Hu, and Alan L. Yuille. Unrealstereo: A synthetic dataset for analyzing stereo vision. *CoRR*, abs/1612.04647, 2016.