3D PHOTOGRAMMETRIC INSPECTION OF RISERS USING RPAS AND DEEP LEARNING IN OIL AND GAS OFFSHORE PLATFORMS

J. D. Salazar [1,2], P. Buschinelli [1], G. Marcellino [1], M. Machado [1], H. Rodrigues [1], D. Regner [1], D. Oliveira [1], J. M. Santos [3], C. A. Marinho [3], M. R. Stemmer [2] and T. C. Pinto [1*]

[1] Mechanical Engineering Department, Labmetro/UFSC - Florianópolis, SC, Brazil
[2] Automation and Systems Department/UFSC - Florianópolis, SC, Brazil
[3] CENPES/Petrobras, Rio de Janeiro, RJ, Brazil

KEY WORDS: Pipeline inspection**,** Deep learning, Photogrammetry, RPAS, UAV, YOLO, Oil and gas.

ABSTRACT

The purpose of this paper is to show how deep learning techniques, based on CNNs, can contribute to photogrammetry process to perform geometric inspections of risers on offshore platforms. The photogrammetry process has a problematic related to the relative movements presented in the scene where the images are being acquired (dynamic photogrammetry). As an alternative solution, this work proposes the use of the YOLOv2 architecture, because this detector complies with some requirements of speed and good performance considering the functional requisites of the study executed. Thus, the purpose of this model is to detect risers and i-tubes on offshore platforms, then extract an inspection riser from the scene. Finally, with the images obtained, a 3D reconstruction is performed, followed by the results' analyses.

1. INTRODUCTION

The integrity of equipment and structures is very important to ensure the safety of the operations in the oil, natural gas, energy and biofuels sector (Mercuri et al., 2015) (Jordan et al., 2018). In this context, one of the main components of oil and gas offshore platforms are the flexible risers, which are pipelines in charge of transporting oil, gas, water and cables between subsea structures and the platform on water surface (Wang et al., 2016). The visual inspection is necessary to guarantee the integrity of the risers, check its general conditions and identify possible damages in the external coating and accessories (e.g. twisting, looping, and bending). The inspection, as shown in Figure 1, is done by industrial climbers who perform manual measurements of riser geometry and photographic record of points of interest. This operation requires a large mobilization of resources for the preparation, inspection and disassembly of auxiliary equipment and structures (Mercuri et al., 2015).

Figure 1. Riser inspection by industrial climber.

Nowadays, RPAs (remotely operated aircrafts) equipped with cameras are being used in the oil and gas industry to perform visual inspection of structures and components such as flares (Marinho et al., 2012) and risers relatively quickly and economically. However, quantitative or geometric studies of the structures have not been done yet. Thus, with the purpose of making geometric measurements of risers from images captured by RPAs, techniques such as photogrammetric 3D reconstruction can be used.

To get a good measurement result using photogrammetry, a series of requirements must be met, such as sequential and overlapping image acquisitions, effective camera positioning (network design), spatial resolution and object texture (Thomas Luhman, Stuart Robson, Stephen Kyle, 2011).

One of the problems presented by the photogrammetry process is related to the relative movements presented in the scene where the images are being acquired. This class of problems is attributed to dynamic photogrammetry (Thomas Luhman, Stuart Robson, Stephen Kyle, 2011). For example, in Figure 2 it is possible to observe a sequence of images acquired in a riser inspection process using a RPA. The relative movement between sea and riser can compromise the quality of the reconstruction, since a large portion of the scene is changing, reducing the number of homologous points between images (Thomas Luhman, Stuart Robson, Stephen Kyle, 2011) (Atkinson, 1996). Therefore, a potential solution for this, is to remove unwanted parts of the scene that hinder the measurement. This removal can be performed by manual task or by specialized tasks done through a computer vision system that captures images and, in addition, by analyzing each of the images, the system will be able to determine actions to follow, such as segmentation of unwanted regions within the image.

Figure 2. Movement of the sea region compared to the riser. This kind of situation impair measurements.

* Corresponding author - tiago.pinto@ufsc.br

Among the different computational techniques that can be used to process the images, deep learning techniques, and more specifically, those related to convolutional neural networks (CNNs), have emerged as a promising alternative that, when applied with RPAs, derive from the concept of deep learning - remotely pilot aircrafts (DL-RPAs). In this way, DL-RPAs systems can assist the photogrammetric inspection processes. In this study, the state-of-art proposal-free CNN-based model YOLO (version 2) (Redmon and Farhadi, 2017) is used for detecting risers and i-tubes present on the platform and then, extract the inspected riser from the scene. With the obtained images, a 3D reconstruction is performed in a commercial photogrammetric software. Finally, a geometric evaluation is performed using simulated 3D scene and images, comparing the 3D result with the ground truth (GT), which is a riser (with some artificial defects) CAD model.

In this paper, a total of 2400 aerial image samples of one offshore platform containing risers and i-tubes under different environment conditions were used. The images were obtained using a DJI M210 RPA (DJI, 2019a) with an DJI Z30 camera (DJI, 2019b).

One of the biggest drawbacks of the images used to perform the research is that they were acquired using zoom in and without considering aspects such as photogrammetry network design, sequential and overlapping acquisition. This, as mentioned earlier, influences the result of the photogrammetric reconstruction. In that context, considering that the final purpose of the research is to execute an analysis of the risers' photogrammetric reconstruction, a ROS/Gazebo simulation environment representing an offshore platform presented in the work of (Salazar et al., 2019), was used for acquire several virtual images (in total 600 images) considering different acquisition strategies and camera configurations.

Therefore, the image dataset used in this paper consists in a total of 3000 virtual and real aerial image samples of an offshore platform that contains risers and i-tubes under different conditions. Furthermore, manual annotations of the images were performed to generate GTs of risers and i-tubes, which are the two classes that the model will detect.

Due to the small training dataset, and to avoid overfitting by the detector, an expansion of the training data (data augmentation) is performed (Liu et al., 2018) (LeCun et al., 2015). Transfer learning is also used on the ConvNet, in this case the darknet (Redmon and Farhadi, 2017), trained in the COCO dataset (Lin et al., 2014).

2. RELATED WORK

Object detection is one of the most important and challenging tasks in computer vision. An object detection model aims to determine instances of objects of a certain class (categories) in an image or video, providing not only the classes of the objects in the image but also additional information such as spatial location of those objects trough the centroids or bounding boxes (Zhao et al., 2019) (Liu et al., 2018).

With the emergence of deep convolutional neural networks (DCNN), especially, due to the success of AlexNet (Krizhevsky et al., 2012) architecture in the Large Scale Visual Recognition Challenge (ILSVRC 2012) for image classification, many deep learning based methods have been proposed in the domain of generic object detection. In this context, generic object detection methods can be categorized into two types: (i) regions proposal-based method (two stage) and (ii) proposal-free methods (one stage).

2.1 Method based on regions proposal

The method based on regions proposal-based method is a two-stage process. In the first stage, regions proposals[1] that may contain objects are generated from an input image. In the second stage, CNN features are extracted from the regions and then, a classification process is carried out in each of the regions to determine its category labels or classes. R-CNN (Girshick et al., 2014) is one of the most popular object detection method that uses CNN (AlexNet), region proposals and a linear SVM (Chang and Lin, 2013) for classification.

Although R-CNN introduces CNN for practical object detection, it requires high computational costs since each region is processed by the CNN network separately. Fast R-CNN (Girshick, 2015) improve the efficiency by sharing the computation of convolution across all the region proposals. Thus, the model runs over the entire image only once instead of thousands of times like R-CNN.

Fast R-CNN speeds up the training and testing time, but it still depends on a set of external regions proposals, causing a bottleneck in the detection process. To solve this problem, (Ren et al., 2015) proposed Faster R-CNN. This model adopts a fast module to generate region proposals instead of the slow selective search or edgebox algorithms. Faster R-CNN consists of two modules: the first module is a fully convolutional network called Region Proposal Network (RPN) for generating region proposals. The second module is a Fast R-CNN used for classification and regression of the ROIs (regions proposals) that were generated in the first module. Thanks to the inclusion of the region proposal based in CNN architecture, Faster R-CNN obtain a frame rate of 5 FPS (Frame Per Second) in a GPU. Also achieve state-of-art object detection accuracy in PASCAL VOC 2007 and 2012, employing 300 regions proposals for image (Zhao et al., 2019).

2.2 Proposal-free (one shot) methods

The region proposals methods are composed of several correlated stages, including region proposal generation, feature extraction with CNNs, classification and bounding box regression, which are trained separately. As a result, those methods are computationally expensive for real time applications and current mobile/wearable devices (LeCun et al., 2015) (Carrio et al., 2017) (Zhao et al., 2019). In order to solve this problem, researchers introduced the second type of generic object detection methods, called proposal-free or one-shot (one-step). Proposal-free methods are based on global regression/classification idea, directly mapping the bounding box and class probabilities from the feature maps generated by a single network. Since the whole pipeline is a single network and does not involve region proposals generation and the subsequent feature resampling stage, it can reduce time expense.

[1] Object proposals, also called region proposals or detection proposals, are a set of candidate regions or bounding boxes in an image that may potentially contain an object.

The YOLO (You Only Look Once) (Redmon et al., 2016), SSD (Single Shot Multi Box Detector) (Liu et al., 2016) and RetinaNet (Lin et al., 2017) frameworks are the state-of-the-art proposal-free methods for real time object detection. In this paper, YOLO version 2 (YOLOv2) (Redmon and Farhadi, 2017) is used for risers and i-tubes detection. This approach outperforms SSD and RetinaNet in terms of speed and it is competitive in terms of accuracy in detecting large objects[2] (Liu et al., 2018) (Li et al., 2020). Moreover, recent works such as (Hossain and Lee, 2019), (Sadykova et al., 2019), (Opromolla et al., 2019) and (Chen and Miao, 2020) achieved good performance in applications involving RPAs and real-time object detection using YOLOv2.

3. METHODOLOGY AND PROPOSED WORK

3.1 YOLO and YOLOv2 architectures

YOLO (Redmon et al., 2016) adopts a single CNN backbone to directly predict bounding boxes and class probabilities from the entire images in one evaluation. As shown in Figure 3, YOLO divides an input image into a $S \times S$ grid and each grid cell is responsible for predicting only one object. If the center of the object falls into a grid cell (for example, the yellow dot represents the center of the riser in the input image), that cell is responsible for the detection of that object. Each grid cell predicts B bounding boxes and their confidence scores. Confidence scores are defined as $P_r(Object) * IOU_{pred}^{truth}$, which reflects how likely the box contains an object ($P_r(Object) \geq 0$) and how accurate is the boundary box (IOU_{pred}^{truth}). The (IOU_{pred}^{truth}) determines the IoU (intersection of union) of the bounding prediction (b_{pred}) boxes and the GT (b_{truth}). Simultaneously, C conditional class probabilities (Pr ($Class|Object$)) are predicted in each cell, regardless of the number of the bounding box number (B). The conditional class probability is the probability that the detected object belongs to a particular class (for each grid cell there are one probability per category).

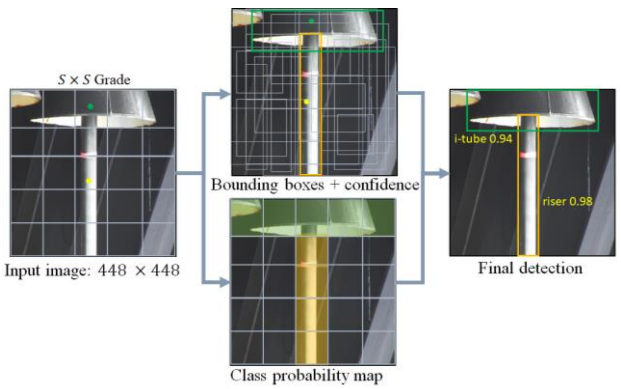


Figure 3. YOLO. The input is divided into a grid, followed by the multiplication of the confidence prediction of the bounding boxes and the class probabilities, generating a final detection result.

Each boundary box contains five components: $(x, y, w, h, conf)$. The (x, y) coordinates represent the center of the box, relative to the grid cell location (the grid cell responsible for the object). The (w, h) represents the width and height of the predicted box. The last component is the prediction confidence $(conf)$. Lastly, adding the class predictions to the output vector, the YOLO algorithm produces a tensor output with the shape $(S \times S(B * 5 + C))$. YOLO uses its own CNN, which is inspired by GoogleNet. This architecture has 24 convolutional layers followed by 2 FC. Since YOLO uses $S \times S$ grid, if an object occupies more than one grid cell, this object may be detected in more than one grid cell. Thus, there are a lot of bounding boxes without any object, that is, duplicate detections for the same object. To fix this, The Non-Maximum Suppression (NMS) (Rothe et al., 2015) method is applied at the end of the network. It consists in merging highly overlapping bounding boxes of a same object into a single one.

By eliminating the proposal region stage, YOLO can process images in real-time at 45 FPS with better results than other real-time detectors. However, its accuracy is less compared to other detectors such as Faster R-CNN and SSD. YOLO may fail to localize some objects, especially small-sized objects, possibly because of the coarse grid cell division, and that each grid cell can only contain one object (Liu et al., 2018). In order to improve the trade-off between the speed and accuracy, (Redmon and Farhadi, 2017) proposed the YOLOv2, in which the YOLO network is replaced with the simpler DarkNet19 network. Darknet has 19 convolutional layers, without FC layers, and uses mostly 3×3 filters to extract features and 1×1 filters to reduce output channels. YOLOv2 also add batch normalization (Ioffe and Szegedy, 2015) in all of the convolutional layers for prevent overfitting without using dropout. Moreover, YOLOv2 improves performance using anchor boxes[3] learned via K-means and multiscale training adjusting the input image size from 224×224 to 448×448, thus, detecting more small-sized object in the image than YOLO (Liu et al., 2018) (Li et al., 2020). Finally, for an input image on the PASCAL VOC 2007 dataset test, YOLOv2 achieves 78.6% mAP (mean average precision) at 40 FPS versus SSD500 mAP 76,8% / 19 FPS or YOLO mAP 63.4% / 45 FPS or Faster R-CNN mAP 73.2% / 7 FPS.

3.2 Data preparation

In this work, two types of images datasets representing an offshore platform (containing risers and i-tubes) were used. The first, with real aerial images to carry out the training and testing of the YOLOv2 object detector. The second, composed of synthetic images acquired from a virtual scene, to perform different types of experiments and photogrammetric analysis.

3.2.1 Real dataset: The real image dataset used in this paper (Figure 4) consists in total of 2400 aerial image samples under different conditions, such as distinct backgrounds, variation in the size target objects, illumination changes (e.g. presence of shadows, non-uniform light distribution, etc.), various types of texture and different acquisition angles. The images were obtained using a DJI M210 RPA system equipped with a DJI 30X optical zoom Z30 camera at the resolution of 1920×1080 pixels (2 MP). Furthermore, manual annotations of the images were performed to generate GTs of risers and i-tubes, which are the two classes that the model will detect.

One of the most challenging problems within deep learning area is the lack of training data. Deep learning models for mapping and inspection of components require a considerable amount of training (Nguyen et al., 2018) (Liu et al., 2018). Thus, due to the small training dataset, and to avoid overfitting by the detector, an expansion of the training data (data augmentation) is executed,

[2] In this work, the images were captured by a RPA at a maximum distance of 10 meters from the target objects. Thus, risers and i-tubes have a large size within the images.

[3] Boxes of various sizes and aspect ratios that serve as object candidates.

applying distortions to the images, changes in brightness, saturation, rotations, etc. (Liu et al., 2018) (LeCun et al., 2015).

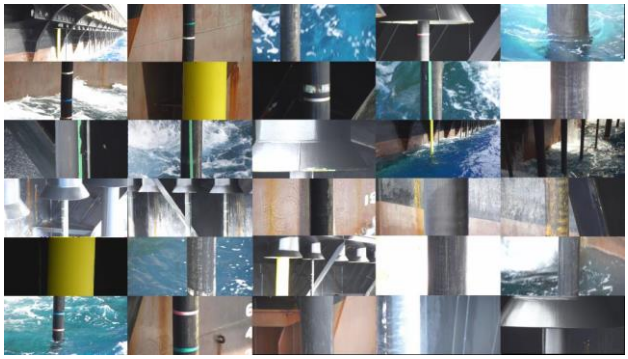


Figure 4. Real dataset for risers inspection process.

3.2.2 Virtual Dataset: One of the biggest drawbacks of the real dataset used to perform the research, is that the images were acquired only for visual inspection without considering photogrammetry requisites, as network design, image overlapping, spatial resolution, type of sensor and object texture (Thomas Luhman, Stuart Robson, Stephen Kyle, 2011) (Atkinson, 1996) (Marcellino et al., 2019),. And as mentioned earlier, this influences the result of 3D reconstructions. To overcome this, a ROS/Gazebo simulation environment (Salazar et al., 2019) was used for acquire different types of images. The virtual scenario, as illustrated in Figure 5, represents a real offshore platform and includes a DJI M210 RPA system, riser balcony, variation of sea surface, illumination and different camera models with parameters that can be modified. The risers used in the experiments were modelled using CAD, with some artificial defects and artifacts. Thus, during the CAD (GT) versus measurement comparison, it is possible to evaluate the quality of the reconstructed object.

Figure 5. Virtual environment containing the M210 RPA, riser balcony and the serpentine trajectory (in red) for inspection process.

In order to perform experiments and analyses, two types of cameras were used for virtual image acquisition: a DJI Z30 (2MP resolution) used in the real offshore platform for real dataset acquisition, and a FLIR Blackfly S equipped with a 40 mm focal length lens at the resolution of 4093 x 3000 (12 MP). The main specifications of the cameras used in simulation are shown in Table 1.

Based on the experimental results of (Salazar et al., 2019), for virtual images acquisition, important factors were considered to obtain good results using photogrammetry. In that context, the RPA performed a serpentine trajectory (a combination of vertical and horizontal displacements and yaw rotation) maintaining constant distance to the riser, and for better points correspondence between the images, 80% image overlap and sequence acquisitions were performed. Figure 6 shows some examples of the virtual dataset obtained. Finally, the virtual dataset is composed of 600 images.

Item \ Camera	DJI Z30	FLIR Blackfly S
Sensor type	1/2.8”	1.1”
Sensor size	5.3 mm x 3.0 mm	14.1 mm x 10.4 mm
Resolution [px]	1920 x 1080 (2 MP)	4096 x 3000 (12 MP)
f (focal length)	15 mm	40 mm
f (for 35 mm eq.)	102 mm	102 mm
AoV [°]	20.0° x 11.3°	20.0° x 14.7°
FoV @ 5 m	1.8 m x 1.0 m	1.8 m x 1.3 m

Table 1. Cameras specifications during simulation. Where: Angle of view (AoV) and field of view (FoV). Both focal lengths were adjusted to result in a 20º horizontal AoV and 1.8 mm FoV @ 5 m.

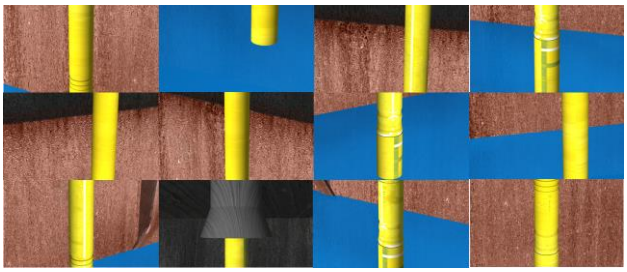


Figure 6. Example of the virtual dataset for riser inspection.

3.3 Performance evaluation

To evaluate the performance of the CNN-based model, the mAP metric was used. It has become the most used metric in detection competitions (Zhao et al., 2019) (Liu et al., 2018). The mAP is calculated by computing the AP (Average Precision) for different classes and averaging them. The AP is based on two underlying metrics: precision and recall. They are defined as follows.

$$precision\ (P) = \frac{TP}{TP + FP} \tag{1}$$

$$recall\ (R) = \frac{TP}{TP + FN} \tag{2}$$

In equation (1) and (2), TP (True Positive), FP (False Positive) and FN (False Negative) correspond to right detections, wrong detections and missed target, respectively. Thus, P represents the percentage of right risers and i-tubes detections among all those identified as risers and i-tubes. R refers to the correct rate of detections among all the GT in the dataset. Finally, the AP is approximate to the area under the precision-recall curve.

4. EXPERIMENTAL ANALYSIS AND RESULTS

The experiments and analysis were divided into two parts. The first part, evaluates the CNN-based approach (YOLOv2) for risers and i-tubes detection, using the real and virtual datasets. The second part focuses on the evaluation of the influence of parameters such as network design, sensor resolution, relative movements (dynamic photogrammetry), overlap and sequence image acquisition, through extensive experiments considering geometric evaluations and object segmentation results.

4.1 Object detection experimental setup

The YOLOv2 performance evaluation was realized employing the real and virtual datasets. Thus, the total of 3000 images were divided into training and test sets according to the ratio 9:1. To avoid overfitting, a simple data augmentation is performed randomly in the training dataset. In this context, considering the constant vertical positions of risers within the capture images, random images were pre-processed in terms of brightness, zoom and vertical flip as shown in Figure 7.

a) b) c) d)

Figure 7. Data augmentation: a) original image; b) Zoom in; c) horizontal mirror; and d) brightness transformation.

Due to the fact that CNNs require a lot of training data before achieve a good performance, it is common to pre-train a CNN in benchmark datasets like COCO (Lin et al., 2014) or VOC (Everingham et al., 2010), and then use the CNN as an initialization or a fixed feature extractor for the task of interest. The low-level and high-level features learned by the CNN in a source domain (the large datasets) can be transferred or fine-tuned in a different, but related, target domain (in this case, Risers and i-tubes), usually by training only the last few layers. This is known as transfer learning (Yosinski et al., 2014) and is usually enough to obtain good performance in the new domain as long as it does not differ drastically from the original. In this study, the Darknet19 CNN pre-train was used in the COCO dataset. The weights to carry out transfer learning are available on the YOLO website[4].

The YOLOv2 CNN-based detector containing the Darknet19 CNN was modified using the darkflow framework (Thtrieu, 2017). It is an implementation based in tensorflow (Abadi et al., 2016) and python, and it was implemented on a laptop[5]. Given the memory constraints of the CPU, the batch size was set to 8. The number of epochs and initial learning rate were set to 150, and 10^{-4}, respectively. Parameters such as momentum and weight decay refer to the original parameters in the YOLOv2 paper.

As previously mentioned, the test set to verify the performance of the model consists of 300 virtual and real RGB images with 1920x1080 and 4096x3000 resolution. They were randomly selected and contain a total of 330 risers and 78 i-tubes, generating a total of 408 object instances, or GTs. The GT bounding boxes of these images are represented in the Figure 8 (a). Note that the risers distribution is greater than the i-tubes distribution, this is mainly due to the fact that the main purpose of the RPA system is the inspection of risers on the offshore platform. True-false positive predictions for each of the classes are illustrated in Figure 8 (b). The few number of FP represents a high precision of the model detecting risers and i-tubes. Thus, the number of risers and i-tubes detected using an IoU of 0.5 is 321 and 75, respectively.

The Figure 8 (c-d) shows the area under the curve measured for AP calculation of individual class, 96.84% for risers and 96.04% for i-tubes. Note that YOLOv2 maintains 100% precision over a fairly wide range of recall, demonstrating the few number of FN in the predictions. Finally, the detector was able to recognize risers and i-tubes objects in the dataset with 96.44% mAP, which confirms the excellent generalization capability of the model to detect those two types of classes. This high performance is due to aspects such as: the data augmentation in the training dataset; the use of transfer learning with a darknet CNN pre-trained on COCO dataset by learning low level features in the first layers such as corners, line, shapes and the fact of the geometrical shapes of risers and i-tubes; and the relation of large dimension objects with the image by performing the inspection at a close distance from the target.

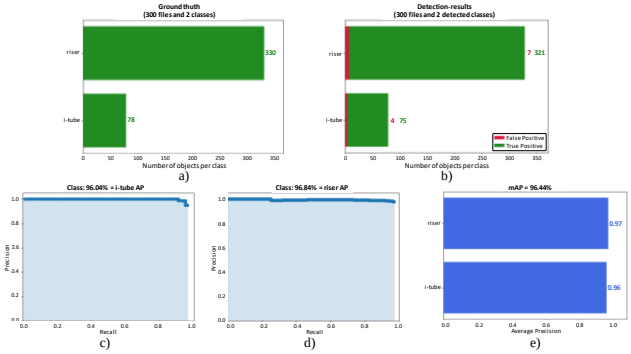


Figure 8. mAP evaluation[6].

As shown in Figure 9, partial occlusion and overlapping situations can occur. In this particular case, the model is able to detect all the target objects in the image considering a min IoU value of 0.5. The IoU values of the large objects (in the image foreground) are greater than 80%. For i-tubes in the background of the image, the value of IoU decreases.

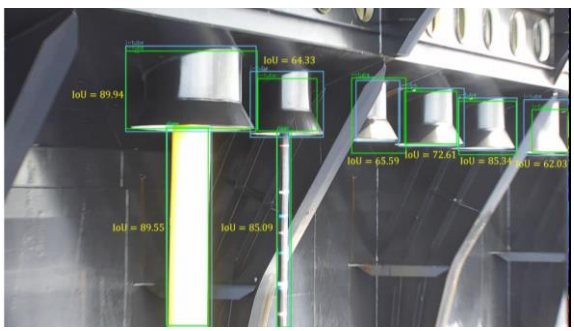


Figure 9. Quantitative results detection.

Figure 10 summarizes the results for several virtual and real testing images. The model detects risers (the principal object for inspection) and i-tubes under different lighting and contrast conditions, although the relative movements (a problem in dynamic photogrammetry) of sea and risers. Note the background variation through the different frames and how the model can detect the target objects. Objects with remarkably similar texture can also appear in the images (Figure 10 e-f). In these cases the model also maintains a good performance. Detecting this type of objects using classical image processing techniques such as, color space transformation and line detection with Hough transform (Lin and Otobe, 2001), is a very complex task, consumes a large processing and generates many false positives. On the other hand, using techniques based in deep learning, it is possible to make complex detections due to specialization of the

[4] https://pjreddie.com/darknet/yolo/

[5] Computer specifications: intel i5-8300H processor at 2.3 GHz with 16 GB RAM and a GTX 1050 Ti GPU with 6 GB memory.

[6] Results generated with: https://github.com/Cartucho/mAP

system (without generating overfitting), by learning complex features through the convolutional layers.

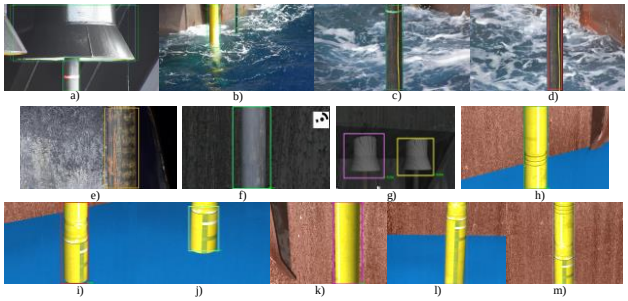


Figure 10. Detection results of risers and i-tubes from real and virtual images.

4.2 Photogrammetric analysis

To analyze the influence of factors such as: relative movements between camera, riser, platform and sea; sensors resolution and network design in the riser inspection process, several photogrammetric reconstructions were performed using Agisoft Metashape (Agisoft, 2020) and analyzed using GOM Inspect (GOM, 2018).
Using a sequence of real images from the training dataset, a 3D reconstruction of the riser inspection scene was obtained. The result is shown in Figure 11. Note that the software is unable to reconstruct the scene. This is influenced by aspects such as: the lack of overlap between the images acquisition; bad network design configuration; zoom in acquisitions and presence of relative movements between the sea and the riser.

Figure 11. Photogrammetric 3D dense cloud for the real images.

Based on the images captured in the virtual scenario, another 3D reconstruction of the scene containing a riser was performed. In this case, acquisition strategies presented in (Salazar et al., 2019) were used, as: the use of RPA serpentine trajectory and 80% overlapping between acquisitions to improve point matching. Moreover, relative movements between the sea and the risers were not considered in the simulation. The camera configuration used was the DJI Z30, which replicates the especifications of the real images acquisition. As shown in Figure 12, a succesful reconstruction was obteined, demonstrating the influence of the acquisition strategies in the reconstruction.

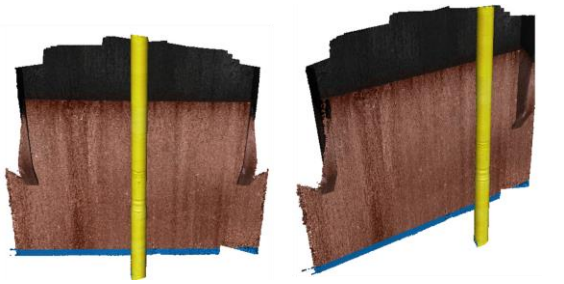
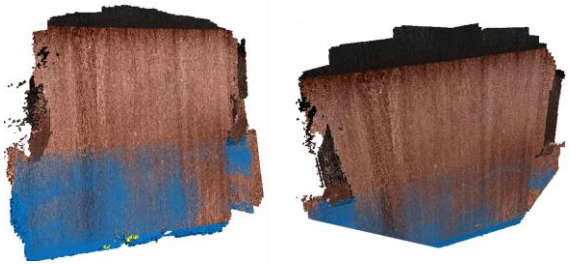


Figure 12. Photogrammetric 3D dense cloud for the virtual images acquired with DJI Z30 camera parameters.

In order to demostrate only the influence of relative movements in the quality of reconstruction, virtual images acquisitions were performed using configurations of Z30 (res. 1920x1080) and Blackfly S (res. 4096x3000) cameras within the virtual scenario, as well as the same acquisition estrategies used in the reconstruction shown in Figure 12. However, movements at sea and ±50 mm horizontal lateral movements (parallel to platform wall) of the riser were added, emulating real capture conditions. Figure 13 shows the result of the 3D reconstructions. It is possible to observe that the software was unable to reconstruct the inspected riser for both resolutions, evidencing the influence of the relative movements in the reconstruction.



a) DJI Z30. b) FLIR Blackfly.

Figure 13. Unsatisfactory 3D reconstruction from simulated acquisition of an inspection scene (with riser and sea movement) without background removal. The riser was not reconstructed.

To solve this problem, the YOLOv2 detector was used in the acquired images. Figure 14 illustrates an example of the segmentation process of the inspected riser within the capture scenario. First, the riser is detected through YOLOv2. With this result, a binaryized mask is created that represents the region of interest (the riser). Thus, the inputs for the reconstruction software are the original image and the mask that represents the object to be reconstructed (the riser), this way, eliminating the impact of the relative movements by the riser and the sea within the scenario capture. The main parameters for the two reconstructions are listed in the Table 2 and the result is illustrated in Figure 15. Here, it is possible to see that the reconstructions were successful, both for acquisitions with the Z30 camera (lower resolution) and for Blackfly (higher resolution). This result allows to perform a geometrical analysis of the reconstruction.

a) Input image b) Detection result c) Obtained mask

Figure 14. Developed CNN-based segmentation process.

Item \ Camera	DJI Z30	FLIR Blackfly S
# of images	232	220
# of points (millions)	0.91	4.7
Spatial resolution [mm/px]	0.95	0.46
Reconstruction time (hours)	0:38	2:46

Table 2. 3D reconstructions result for segmented virtual images.

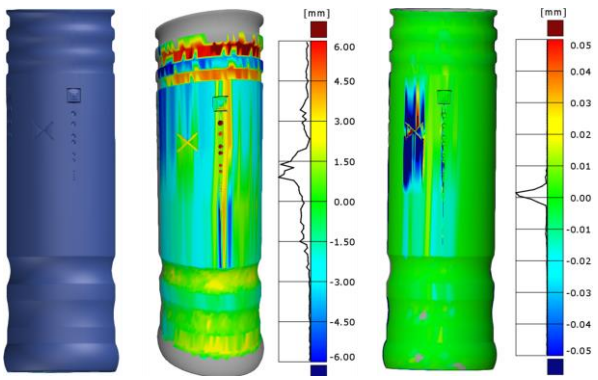
a) DJI Z30.

b) FLIR Blackfly.

Figure 15. Horizontal view of the successful reconstructed 3D textured meshes of the riser from simulated acquisition of an inspection scene (with riser and sea movement) using the developed CNN-based strategy for riser background removal.

Once the objective of eliminating relative movement has been achieved, it is possible to make an evaluation of the impact of the used cameras resolution. For this, it was performed a surface comparison between the reconstructed riser and the GT. This evaluation is performed by a point-to-point mesh comparison. Figure 16 shows the deviation map generated by the GOM Inspect software. It is possible to find a low (green), negative (blue) or positive (red) deviation of the reconstructed mesh in relation to the GT.

As a quality parameter, the standard deviation for the deviations between surfaces was found, being 3.4 mm for the lowest resolution camera and 0.15 mm for the highest resolution camera. These values show the relationship between the quality of the reconstruction and camera resolution. The lower error is mainly due to the better spatial resolution [mm/px] obtained using a higher resolution camera for a similar FoV.



a) GT CAD b) Result for Z30 c) Result for Blackfly

Figure 16. Deviation maps of the reconstructed riser to GT.

5. CONCLUSION

In this paper, a 3D optical riser inspection using RPAs and photogrammetry was improved by incorporating the CNN-based detector YOLOv2 in the 3D photogrammetric reconstruction process. The CNN-model detected risers and i-tubes in a scenario that presents relative movements and different environment conditions with 96.44% mAP in a dataset containing real and virtual images. Real images were acquired with a M210 RPAs equipped with a DJI Z30 camera, however, without considering photogrammetric strategies. To obtain virtual images, a ROS/Gazebo simulation environment was used considering acquisition strategies such as: serpentine trajectory, 80% overlap, sequence acquisition and DJI Z30 and FLIR Blackfly cameras parametrization. It was not possible to obtain a good reconstruction result using the real images samples, mainly because they were acquired only for visual inspection and no photogrammetry procedure were followed. The 3D reconstructions were successful for acquisition in the virtual environment with no relative movement (static sea and risers). However, with an inspection scene including relative movements, it was not possible to obtain a dense point cloud. With the developed detection model, the relative movements were eliminated by removing the background using a mask created from the detected riser in the inspected scenario, resulting in a successful 3D reconstruction. Finally, geometric evaluations for the 3D reconstructions were performed showing that the standard deviation error for the Blackfly camera (0.15 mm) is less than the error of the Z30 camera (3.37 mm), mainly due to different spatial resolution between the simulated cameras.

Future work will focus on testing the detection model in real images acquired with photogrammetric acquisition strategies. In addition, an onboard model detection implementation for the RPA will be implemented.

ACKNOWLEDGMENT

The authors would like to thank Petrobras/CENPES for funding the VANT3D research project. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., others, 2016. Tensorflow: a system for large-scale machine learning., in: OSDI. pp. 265–283.

Agisoft, 2020. Photogrammetry software - Agisoft Metashape 1.5.5.

Atkinson, K., 1996. Close Range Photogrammetry and Machine Vision, 1st ed. Whittles Publishing, Bristol.

Carrio, A., Sampedro, C., Rodriguez-Ramos, A., Campoy, P., 2017. A review of deep learning methods and applications for unmanned aerial vehicles. J. Sensors 2017. https://doi.org/10.1155/2017/3296874

Chang, C., Lin, C., 2013. LIBSVM : A Library for Support Vector Machines. ACM Trans. Intell. Syst. Technol. 2, 1–39. https://doi.org/10.1145/1961189.1961199

Chen, B., Miao, X., 2020. Distribution Line Pole Detection and Counting Based on YOLO Using UAV Inspection Line Video. J. Electr. Eng. Technol. https://doi.org/10.1007/s42835-019-00230-w

DJI, 2019a. Matrice 200 series v2 [WWW Document]. URL https://www.dji.com/matrice-200-series-v2 (accessed 2.5.20).

DJI, 2019b. DJI Zenmuse Z30 [WWW Document]. URL https://www.dji.com/br/zenmuse-z30

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The pascal visual object classes (VOC) challenge. Int. J. Comput. Vis. https://doi.org/10.1007/s11263-009-0275-4

Girshick, R., 2015. Fast R-CNN, in: Proceedings of the IEEE International Conference on Computer Vision. https://doi.org/10.1109/ICCV.2015.169

Girshick, R., Donahue, J., Darrell, T., Berkeley, U.C., Malik, J., 2014. R-CNN. 1311.2524v5. https://doi.org/10.1109/CVPR.2014.81

GOM, 2018. GOM Inspect 2018 Hotfix 3 [WWW Document]. URL https://www.gom.com/3d-software/gom-inspect.html

Hossain, S., Lee, D.J., 2019. Deep learning-based real-time multiple-object detection and tracking from aerial imagery via a flying robot with GPU-based embedded devices. Sensors (Switzerland). https://doi.org/10.3390/s19153371

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. pp. 448–456.

Jordan, S., Moore, J., Hovet, S., Box, J., Perry, J., Kirsche, K., Lewis, D., Tse, Z.T.H., 2018. State-of-the-art technologies for UAV inspections. IET Radar, Sonar Navig. 12, 151–164. https://doi.org/10.1049/iet-rsn.2017.0251

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. AlexNet. Adv. Neural Inf. Process. Syst. https://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep Learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

Li, K., Wan, G., Cheng, G., Meng, L., Han, J., 2020. Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS J. Photogramm. Remote Sens. https://doi.org/10.1016/j.isprsjprs.2019.11.023

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P., 2017. Focal Loss for Dense Object Detection, in: Proceedings of the IEEE International Conference on Computer Vision. https://doi.org/10.1109/ICCV.2017.324

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: Common objects in context, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-319-10602-1_48

Lin, X., Otobe, K., 2001. Hough transform algorithm for real-time pattern recognition using an artificial retina camera. Opt. Express 8, 503–508. https://doi.org/10.1364/OE.8.000503

Liu, L., Ouyang, W., Wang, X., Fieguth, P.W., Chen, J., Liu, X., Pietikäinen, M., 2018. Deep Learning for Generic Object Detection: {A} Survey. CoRR abs/1809.0.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016. SSD: Single shot multibox detector, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-319-46448-0_2

Marcellino, G.C., Fonseca de Oliveira, B.C., Borges, V., Figaro da Costa Pinto, T.L., 2019. A conceptual study of infrared and visible-light image fusion methods for three-dimensional object reconstruction 100. https://doi.org/10.1117/12.2527428

Marinho, C.A., Souza, C. De, Motomura, T., Silva, A.G. da S., 2012. In-service flares inspection by unmanned aerial vehicles (UAVs), in: 18th World Conference on Nondestructive Testing. Durban, pp. 16–20.

Mercuri, S.M., Fisicaro, A., A, V.T.E.S., 2015. UAV THE IMPACT & INFLUENCE IN THE O&G. Offshore Mediterr. Conf. Exhib. 1–8.

Nguyen, V.N., Jenssen, R., Roverso, D., 2018. Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning. Int. J. Electr. Power Energy Syst. https://doi.org/10.1016/j.ijepes.2017.12.016

Opromolla, R., Inchingolo, G., Fasano, G., 2019. Airborne visual detection and tracking of cooperative UAVs exploiting deep learning. Sensors (Switzerland). https://doi.org/10.3390/s19194332

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem, 779–788. https://doi.org/10.1109/CVPR.2016.91

Redmon, J., Farhadi, A., 2017. YOLO9000: Better, faster, stronger, in: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. https://doi.org/10.1109/CVPR.2017.690

Ren, S., He, K., Girshick, R.B., Sun, J., 2015. Faster {R-CNN:} Towards Real-Time Object Detection with Region Proposal Networks. CoRR abs/1506.0.

Rothe, R., Guillaumin, M., van Gool, L., 2015. Non-maximum suppression for object detection by passing messages between windows, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-319-16865-4_19

Sadykova, D., Pernebayeva, D., Bagheri, M., James, A., 2019. IN-YOLO: Real-time Detection of Outdoor High Voltage Insulators using UAV Imaging. IEEE Trans. Power Deliv. https://doi.org/10.1109/tpwrd.2019.2944741

Salazar, A., Regner, D., Oliveira, D., Marcellino, G., Buschinelli, P. de D.V., Tiago Loureiro Figaro da Costa Pinto, Santos, J.M. dos, Stemmer, M., 2019. Desenvolvimento de um ambiente de simulação ROS/Gazebo para inspeção fotogrametrica 3D de risers com RPAs, in: Anais Do 14º Simpósio Brasileiro de Automação Inteligente. Galoá, Ouro Preto, Brazil.

Thomas Luhman, Stuart Robson, Stephen Kyle, I.H., 2011. Close range photogrammetry: Principles, techniques and applications. Gardners Books, Luhmann, Thomas., Robson, Stuart,Whittles Publishing.

Thtrieu, 2017. darkflow [WWW Document]. URL https://github.com/thtrieu/darkflow (accessed 2.1.20).

Wang, C., Shankar, K., Morozov, E. V., 2016. Tailored design of top-tensioned composite risers for deep-water applications using three different approaches. Adv. Mech. Eng. 9, 1–18. https://doi.org/10.1177/1687814016684271

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks?, in: Advances in Neural Information Processing Systems.

Zhao, Z.Q., Zheng, P., Xu, S.T., Wu, X., 2019. Object Detection With Deep Learning: A Review. IEEE Trans. Neural Networks Learn. Syst. https://doi.org/10.1109/TNNLS.2018.2876865