A System for Automatic Reporting of Technician Movements During Offshore Maintenance Trips

Issam Boukhennoufa*, Xiaojun Zhai*, Amit Kumar Singh*, Bob Cooper†, Tom Mehew†
*School of Computer Science and Electronic Engineering, University of Essex, Colchester, United Kingdom
†Njord Offshore Ltd, Tendring, United Kingdom
Emails: ib20472@essex.ac.uk, xzhai@essex.ac.uk, a.k.singh@essex.ac.uk, bcooper@njordoffshore.com, tmehew@njordoffshore.com,

Abstract—**Technician transfer reporting for Offshore Wind Turbines (OWTs) is essential for tracking movement and productivity, but traditional manual methods are laborious, error-prone, inconsistent, time-consuming, and lack real-time updates. This paper presents the first automated, vision-based system to detect, track, and report technician transfers between crew transfer vessels (CTVs) and OWTs. Using low-resolution cameras on the CTV, the system employs deep learning for object detection, investigating the Single Shot MultiBox Detector (SSD) (with MobileNetV2/InceptionNet backbones), combined with the efficient Centroid Tracker (CT) for robust tracking. Processing occurs onboard via a specialised System-on-Module using NVIDIA Jetson Nano GPU capabilities. The chosen configuration achieves real-time performance (greater than 37 FPS) with minimal data storage. The system automatically logs key events such as the push-on of the vessel, personnel counts, precise transfer times, and work durations, offering the granularity of the data absent in manual reporting. Developed and validated on a unique dataset from diverse real-world offshore conditions, the system is deployed and operational on three vessels from a leading UK CTV operator's fleet (Njord offshore ltd) since February 2024. This automation improves reporting accuracy, enhances operational efficiency, and reduces administrative burden on crew.**

Index Terms—**Object Tracking, Single Shot MultiBox Detector, Real-time, Crew Transfer Vessels, Offshore wind-farm**

I. INTRODUCTION

Effective Offshore Wind Turbine (OWT) maintenance is fundamental to the cost efficiency, operational success, and long-term viability of Offshore Wind Farm (OWF) projects [1]. Ensuring these complex machines achieve peak performance is paramount for maximising energy generation [2], maintaining grid stability, and securing profitability [3] [4]. Consequently, well-planned maintenance scheduling is essential to prevent costly unscheduled downtime or premature component failures and uphold rigorous safety standards [5]. This necessitates specialised technicians undertaking regular, complex journeys offshore [6].

Tracking these essential personnel movements – vital not only for monitoring progress against schedules, verifying technician numbers, calculating time on site, and enabling accurate billing [6], but also for safety compliance and overall personnel accountability – still overwhelmingly relies on traditional manual recording methods [6]. Crew Transfer Vessels (CTVs), often specialised high-speed catamarans, are indispensable for transporting these technicians safely and efficiently across often considerable distances offshore [7]. Purpose-built for demanding maritime conditions including rough seas, high winds, and variable visibility, these challenging environments inevitably impact vessel operations, potentially compromising safe personnel transfers between vessel and turbine, and affecting adherence to strict maintenance schedules [8].

Despite wider technological advances within the OWF sector, the critical process of recording these personnel transfers remains predominantly manual, typically reliant on handwritten logs or basic spreadsheets maintained by the vessel crew. This traditional, time-consuming method is inherently prone to significant human error. Common issues include inaccurate or estimated timings, entirely missed entries during busy operational periods, and inconsistent documentation formats, all of which seriously hinder accurate data retrieval, subsequent analysis, and can impact operational reporting and safety management procedures. All of which seriously hinder accurate data retrieval, subsequent analysis, and can impact operational reporting and safety management procedures, a challenge that has driven the adoption of automated monitoring in other industrial sectors [9].

To address these persistent operational inefficiencies and data integrity issues, the adoption of modern, automated tracking systems is becoming increasingly necessary. Building upon this need, we propose a novel, vision-based system specifically designed to automate the reporting of technician movements during offshore maintenance campaigns, significantly reducing reliance on fallible manual logging procedures.

Our innovative system utilises strategically mounted cameras, potentially located on the vessel's bridge or near the transfer deck area, coupled with state-of-the-art object detection deep learning algorithms for automatic, unobtrusive personnel tracking. It operates efficiently using low-resolution video feeds, thereby minimising onboard data storage requirements – a key practical consideration on space-constrained vessels. Impressively, it achieves real-time processing performance (exceeding 37 frames per second) even when deployed on power-efficient edge computing hardware, such as the NVIDIA Jetson Nano platform, demonstrating its suitability for onboard implementation. By automating this critical transfer tracking function, the system effectively removes the administrative burden and inherent potential inaccuracies associated with manual

This work is supported by UK Research and Innovation funding on KTP number 10049632.

Fig. 1: Hardware setup for processing the CTV automatic reporting

reporting. This directly frees up valuable CTV crew time and cognitive load, allowing them to concentrate fully on other safety-critical tasks, including navigation, precise vessel handling during turbine approaches and transfers, and maintaining crucial situational awareness in dynamic offshore conditions. The system's practicality and effectiveness are not merely theoretical; it has undergone rigorous real-world testing and has been successfully deployed and validated within the operational fleet of one of the United Kingdom's foremost and highly experienced CTV operators. This successful implementation demonstrates its tangible real-world viability and highlights the significant potential for such technology to enhance operational efficiency and safety standards across the wider offshore wind industry. The main contributions are the design of the first automated vision-based system for this task, its validation on a unique real-world dataset, and its successful deployment on three commercial vessels, where it achieves real-time performance on low-power onboard hardware.

The remainder of this paper is organised as follows. Section II details the methodology, including system architecture, functionalities, hardware implementation, data collection, dataset preparation, and the core algorithm pipeline. Section III presents experimental results, covering performance evaluation of different configurations, justification for the chosen setup, and accuracy considerations. Finally, Section IV concludes the paper, summarising the key findings and contributions.

II. Methodology

This work introduces an automated tool specifically developed for the detection, tracking, and subsequent reporting of technician movements during transfers between CTVs and OWTs. The system's workflow is sequential: an algorithm automatically triggers video recording upon turbine approach, the feed is processed in real-time by the pipeline detailed in Figure 2 to log all events, and recording ceases on departure. The core functionalities of this tool include:

1) Counting Personnel: It automatically quantifies the number of technician members who undertake the transfer, specifically registering each instance of an individual ascending onto the turbine structure and each instance of an individual descending back to the vessel.
2) Recording Timings: The system captures and logs essential time-based data points. This includes recording the specific times when transfers take place and also documenting the overall work durations spent by technicians on the turbine.
3) Tracking Key Events: Beyond general timings, the tool actively tracks and records the precise moments of several specific key events that define the transfer process sequence. The specific events monitored are:
   - The exact time of the vessel's initial 'Push onto' the turbine structure.
   - The duration of time measured from that initial 'push-on' until the first person is visible on the vessel's deck area designated for transfer.
   - The duration of time measured from when that first person appears on the deck until the first actual transfer action (ascending the ladder) commences.
   - The specific time when the last person involved in the transfer operation is visible on the vessel's deck.

The practical implementation of this automatic tool, illustrated in Figure 1, relies on efficient edge computing performed directly onboard the vessel. All the necessary processing to achieve these functionalities is handled by a specialised System-on-Module (SoM) positioned under the vessel's bridge. This computing unit utilises the powerful GPU capabilities derived from the NVIDIA Jetson Nano architecture for its core processing. It is augmented by additional peripherals not natively part of the standard Jetson Nano platform, specifically CAN bus and RS485 connections for interfacing with vessel systems, along with extra onboard storage. These additions are crucial for the system's operational requirements and integration needs, including providing data inputs used by an independent algorithm to automatically trigger the video recording and subsequent detection/tracking processes when the vessel approaches a turbine. For development, testing, and validation purposes, this system was deployed across three operational vessels drawn from the fleet of Njord offshore LTD

The automatic tool operates through several distinct steps within its pipeline. At the outset, video footage is captured from either a standard USB camera or an existing CCTV camera, provided it is mounted appropriately on the CTV's bridge. Crucially, the camera must be positioned so that its field of view points directly towards the designated transfer area located at the front of the vessel's deck. To optimise data capture and manage storage efficiently, the cameras are programmed to begin recording automatically only when the vessel approaches a wind turbine structure and to cease filming once the vessel moves away from the turbine's immediate vicinity.

Once the relevant video feed is being captured, the system employs established computer vision algorithms for analysis in real-time. It utilises the Single Shot MultiBox Detector (SSD) [10] algorithm, coupled with Non-maximum suppression (NMS) techniques, to perform robust detection of technicians appearing within the video frames. Following detection, the well-regarded Centroid Tracker (CT) [11] algorithm is applied. This algorithm tracks the detected technicians' movements consistently across consecutive video frames, which enables the system to log their paths accurately and register the transfer actions described earlier. The complete processing pipeline, visually outlining these steps from video capture through to tracking, is presented in Figure 2.

Next, we describe our techniques for data collection/preparation and crew transfer counter pipeline.

A. Dataset Collection and Preparation

This automated tool was developed using an extensive dataset meticulously gathered during prolonged real-world operations aboard the Njord Forseti, Magni, and Thor offshore support vessels, utilising strategically positioned CCTV and USB cameras on the bridge.

A critical component is the automated recording function, engineered through iterative refinement to capture only pertinent operational periods. This system, refined through operational testing, is triggered by a dedicated algorithm running on the onboard SoM. To accurately determine when the vessel is approaching or leaving a wind turbine, this algorithm processes a fusion of real-time data streams. These streams include inputs from various onboard sensors connected via standard maritime communication protocols like RS485 and CAN bus, integrated with precise positional data (e.g., GPS/GNSS) and motion/orientation readings from an Inertial Measurement Unit (IMU). This sophisticated approach ensures recordings focus exclusively on relevant operational phases near the turbines.

Optimising video parameters demanded a particularly labour-intensive phase of exhaustive experimentation, driven by the challenging constraints of limited onboard storage and the computational demands of edge processing. Testing numerous configurations involving a wide spectrum of resolutions and frame rates under operational conditions was essential. Rigorous analysis confirmed that higher settings severely taxed system resources, while lower frame rates risked missing critical transient events during transfers. Ultimately, through this demanding process, a resolution of 640×360 pixels and a frame rate of 10 frames per second (FPS) was validated as providing the optimal equilibrium between a manageable data footprint and the processing efficiency required for reliable real-time performance on the deployed hardware.

The final curated dataset, representing the culmination of this extensive data collection campaign, comprises two highly relevant hours of focused video footage. Crucially, this curated dataset was distilled from a much larger corpus, representing well over 800 hours of raw recordings captured across numerous distinct operational deployments and transfer scenarios at OWTs. A primary objective during the acquisition phase was the deliberate capture of footage across the widest possible spectrum of challenging environmental conditions. This involved actively recording during diverse weather patterns (from clear skies to inclement weather) and highly variable lighting situations (spanning bright daylight, heavy overcast conditions, dawn/dusk transitions, and operations under artificial nighttime illumination), ensuring the dataset mirrors the complex operational realities and promotes robust system performance.

From this rich pool of curated footage, requiring the review of a vast number of initial frames, those containing visible technicians—even partially visible ones within the transfer zone—were painstakingly extracted. These selected frames then underwent meticulous, frame-by-frame annotation, standardised according to the widely-adopted Pascal VOC format. To establish a rigorous training and validation protocol aimed at maximising model generalisation, this comprehensive annotated dataset was strategically partitioned: 80% was allocated for training the detection algorithms, while the remaining 20% was strictly sequestered as a distinct validation set, guaranteeing an unbiased assessment of the final system's real-world applicability. Figure 3 shows a sample annotated frame representative of the dataset's quality

B. Crew Transfer Counter Pipeline

In this work, the SSD was investigated [10], with two different backbones: MobileNetV2 and InceptionNet. SSD was chosen because it is best suited for real-time applications as it processes faster than traditional detection models. Furthermore, SSD allows for greater choice in backbone selection: MobileNetV2 is lightweight and ideal for applications with limited computational resources, while InceptionNet gives improved accuracy at the cost of more processing power.

1) Detection and Tracking Pipeline: The resulting predictions are then subjected to NMS to eliminate duplicate detections of the same object, ensuring that each object is only retained with the highest confidence. Afterwards, different tracking techniques were considered, and the Centroid Tracker was chosen for its performance in keeping the identification of objects between frames.

The Centroid Tracker is an object-tracking algorithm that links objects across frames by comparing the centroids of detected objects. It matches objects by calculating the distance between centroids and linking them if this falls below a given threshold. In special cases when few objects move consistently with no appreciable overlap, it is a very efficient approach and thus fits real-time applications. Tracking occurs within a region of interest (ROI) at the ladder portion of the wind turbine with a confidence of 35%. The ROI is used to track only crew objects and identify as few irrelevant objects as possible (due to the low confidence used). The centroid tracker is used at

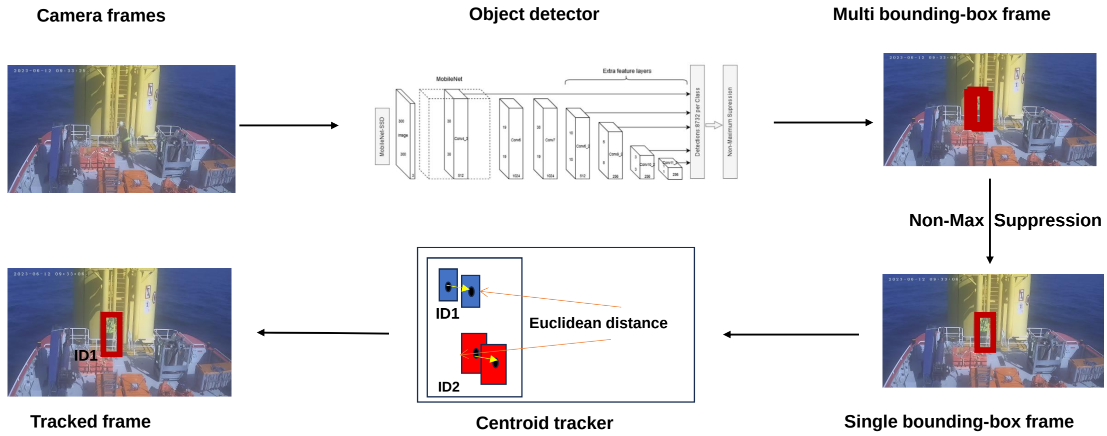

Fig. 2: Technician transfer report pipeline.

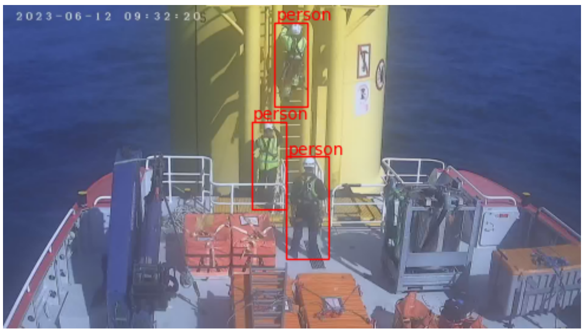

Fig. 3: Annotated frame from the captured dataset

maxDisappeared 1.5 FPS and MaxDistance 0.15 frame height and the NMS filter was set to 25% overlap. MaxDistance is the maximum allowable distance between the centroids of recognised objects that can be linked across frames while maxDisappeared specifies the maximum number of frames an object can disappear before being deemed lost. After that, when an item moves from bottom to top and crosses a predetermined coordinate, the "counter up" is incremented; when it moves from top to bottom and crosses another point, the "counter down" increases. Furthermore, the turbine and wind farm names, as well as the times of first and last detection, are overlayed on the frame.

For the core task of identifying technicians within the video feed, this work investigated the SSD [10]. SSD is a well-established, single-stage object detection algorithm, meaning it performs object localisation and classification in a single forward pass of the neural network. This architectural design makes it inherently faster than two-stage detectors (like Faster R-CNN), rendering it particularly suitable for applications demanding real-time processing. We explored implementations using two different convolutional neural network backbones to extract features: MobileNetV2 and InceptionNet.

The choice of SSD was driven primarily by its real-time efficiency and its flexibility in backbone selection. This flexibility allows for a crucial trade-off between processing speed and detection accuracy, essential when targeting specific hardware. The MobileNetV2 backbone is notably lightweight, architecturally designed for computational efficiency on mobile and embedded devices. This characteristic makes it highly appropriate for resource-constrained scenarios, such as deployment on the onboard SoM which utilises the Jetson Nano GPU's processing capabilities. In contrast, the InceptionNet backbone, while offering potentially higher detection accuracy, demands significantly greater computational resources. Although more recent state-of-the-art object detectors (e.g., various iterations of YOLO, EfficientDet) might provide further accuracy improvements on high-end hardware, their computational complexity often exceeds the practical processing capacity of the target Nano-based SoM if real-time frame rates (like the targeted 10 FPS or higher) are to be consistently achieved. Therefore, SSD, particularly when paired with the MobileNetV2 backbone, represents a pragmatic and effective choice, balancing performance and accuracy for this specific edge computing application.

Following the initial detection phase by SSD, NMS was applied to the raw bounding box predictions. NMS is an essential post-processing technique universally employed in object detection pipelines. Its function is to resolve situations where multiple bounding boxes are predicted around the same object by suppressing redundant, overlapping boxes based on their confidence scores. This ensures that, ideally, only the single bounding box with the highest confidence score is retained for each distinct technician detected, leading to cleaner and more accurate final detections for subsequent tracking.

For the subsequent task of object tracking – that is, maintaining the identity of detected technicians as they move across consecutive video frames – various techniques were considered. The CT [11] was ultimately selected for implementation. CT is a computationally lightweight and relatively simple tracking algorithm. Its core principle involves calculating the Euclidean distance between the centroids (geometric centres) of detected object bounding boxes in one frame and the centroids of detected objects in the next frame. It then associates objects across frames by matching pairs of centroids that exhibit the minimum distance, provided this distance is below a predefined threshold.

The primary advantages of CT are its very low computational overhead and ease of implementation, making it eminently suitable for real-time operation on computationally limited hardware like the target Nano-based SoM. While more sophisticated tracking algorithms exist (such as DeepSORT, which incorporates appearance features, or trackers employing Kalman filters with complex motion models), these invariably demand substantially more processing power and memory, potentially hindering real-time performance on the edge device. The CT approach, despite its simplicity, works reliably well in scenarios where objects exhibit relatively smooth and predictable motion with minimal bounding box overlap between distinct individuals – conditions frequently met in the specific context of technicians transferring via an access ladder, making it a well-justified, resource-efficient choice for this application.

Tracking was limited specifically to a ROI defined around the OWT ladder area. This focuses computational effort on the critical zone and helps manage potential false detections. A confidence threshold of 35% was used for initial SSD detection within this ROI; this relatively low threshold allows for detection even in challenging lighting or partial visibility conditions, while the ROI constraint helps mitigate the risk of spurious detections outside the transfer area being processed by the tracker. The CT parameters were carefully tuned through experimentation: a maxDisappeared value of 1.5 FPS was set (representing the maximum time, relative to the frame rate, an object can remain undetected before its track is considered lost) and a MaxDistance value of 0.15 times the frame height (specifying the maximum allowed pixel distance between centroids across frames for them to be linked as the same object). For the NMS post-processing step, the overlap threshold was set to 25%.

Fig. 4: Frame with tracking information overlaid.

Within this defined ROI, the system implements the core counting logic.

2) Region of Interest and Counting Mechanism: Tracked objects (technicians identified by the CT) crossing predefined horizontal coordinates trigger the relevant counters: movement from bottom-to-top increments the "counter up," while movement from top-to-bottom increments the "counter down." For enhanced reporting and operational context, key information such as the specific turbine and wind farm names, along with the precise times of the first and last technician detection during that particular visit, were also overlaid directly onto the output video frames. Crucially, the time of each individual transfer event (specifically, the moment of crossing the up or down counting line) is also logged into a corresponding data file for permanent record-keeping and later analysis. Figure 4 illustrates a typical output frame produced by the application of this complete pipeline, showing tracked individuals within the ROI and the overlaid contextual information.

To clarify, the system's output is quantitative, not a qualitative text summary. It generates a structured log file of discrete transfer events with precise timestamps, alongside an annotated video for real-time awareness. Consequently, its reporting capability is evaluated on the objective accuracy of these logged counts and times.

III. Experimental results

To determine the optimal configuration for the technician detection component, balancing accuracy with the real-time processing constraints of the onboard SoM platform, a series of experiments were conducted. These experiments focused on evaluating the performance of the selected Single Shot SSD [10], specifically comparing the two chosen network backbones: MobileNetV2 (referred to as MobileNet ssd) and InceptionNet (referred to as Inception ssd). The comparison was performed across various input resolutions fed into the detection model.

It is important to clarify that 'resolution' in the context of these experiments (and as presented in Table I) refers to the square dimensions (e.g., 300×300, 600×600) to which the input video frames – originally captured at 640×360 pixels – were resized using bilinear interpolation. This resizing step ensures that the input data dimensions precisely match the requirements of the SSD model's input layer. The detailed performance metrics comparing the backbones across these different input resolutions are presented in Table I.

The results summarised in Table I revealed distinct performance characteristics for each backbone. The configurations utilising the InceptionNet ssd backbone generally exhibited a slightly lower overall loss (defined as the sum of the model's regression loss and classification loss). This suggests that InceptionNet potentially offered marginally higher raw detection accuracy. However, configurations using the MobileNet ssd backbone consistently demonstrated significantly faster processing speeds, achieving considerably higher FPS throughput on the target hardware.

Considering the paramount requirement for real-time performance in this operational application, a careful trade-off between detection accuracy and processing speed was necessary. Based on the results in Table I, the configuration employing the MobileNet ssd backbone with input frames resized to 600×600 pixels was selected as the optimal choice for deployment. This specific configuration achieved a respectable combined loss of 2.585 while delivering a processing speed of 37 FPS. This performance was deemed the most favourable balance, providing a sufficient level of detection accuracy while comfortably exceeding typical real-time requirements (e.g., >10 FPS), ensuring smooth and effective operation on the edge computing platform.

It is important to contextualise these results. The selection of SSD with MobileNetV2 was a deliberate trade-off, prioritising real-time performance on the resource-constrained SoM over the marginal accuracy gains of more computationally demanding models which were unsuitable for this hardware. While formal end-to-end counting metrics were not part of this study, the system's operational accuracy was validated through its successful deployment since February 2024. During this period, its automated logs consistently proved more reliable and less error-prone than the traditional manual reporting methods it replaced

Alongside the selection of the core detector configuration, the various parameters controlling the tracking process were also refined through empirical evaluation. As detailed in the previous section (Section II-B in the original reference, describing the Detection and Tracking Algorithms), the specific values for the CT parameters (maxDisappeared, MaxDistance), the NMS overlap threshold, and the initial detection confidence threshold (35%) were chosen based on iterative testing. These values were selected because they yielded the most accurate and robust tracking performance across the diverse range of conditions present in the collected video dataset.

Detector loss values (such as those in Table I) indicate model fit but do not directly represent the overall system's operational accuracy (covering detection, tracking, and counting). A trade-off was made (MobileNetV2 chosen over the lower-loss InceptionNet) prioritising real-time speed. End-to-end accuracy depends on the entire pipeline; errors can stem from imperfect detection (missed or false detections) or tracking failures (ID switches, lost tracks). Therefore, empirically tuning the tracker (CT) and post-processing (NMS) parameters was crucial for optimising practical accuracy and mitigating errors cascading from the detector stage. Whilst the system's deployment suggests operational viability, the study did not include a specific quantitative measure of end-to-end transfer counting accuracy against a ground-truth dataset.

Finally, regarding the practical application of the system's output, the calculation of total work duration on a turbine was defined. In operational scenarios, vessels frequently visit the same turbine multiple times within a single offshore trip (e.g., for drop-off and later collection of technicians). Therefore, the system calculates the total work duration for a specific asset by logging the precise times of each transfer up ('counter up' event) and the corresponding transfer down ('counter down' event). It then sums the durations of these individual work periods (time between ascending and descending) for that turbine across the entire duration of the vessel's trip, providing an accurate aggregate measure of time spent working on the asset.

Metric	**Model**	**300**	**400**	**600**	**800**
Loss	Mobilenet_ssd	4.523	3.012	2.585	2.491
	Inception_ssd	3.921	2.710	2.323	2.201
Performance (FPS)	Mobilenet_ssd	68	60	37	28
	Inception_ssd	52	45	24	17

TABLE I: Comparison of Mobilenet_ssd and Inception_ssd across different resolutions.

IV. CONCLUSION

This study successfully demonstrates a novel, real-time automated system for tracking technician transfers between CTVs and OWTs, addressing the limitations of traditional manual reporting. Processing low-resolution video onboard using efficient deep learning (SSD with MobileNetV2, CT) on a low-power SoM (using Jetson Nano GPU capabilities), the system improves accuracy and consistency over manual logs. Real-time performance (greater than 37 FPS) is achieved while automatically documenting key transfer events with minimal storage requirements. Successful deployment since February 2024 on three vessels of a leading UK CTV operator confirms its real-world applicability. This automation enhances operational accuracy and efficiency, reduces manual reporting burden, and generates reliable data. The generated data enables improved operational analysis, planning optimisation, and potential safety enhancements, indicating promise for wider adoption in the offshore wind industry.

REFERENCES

[1] X. Wu, Y. Hu, Y. Li, J. Yang, L. Duan, T. Wang, T. Adcock, Z. Jiang, Z. Gao, Z. Lin *et al.*, "Foundations of offshore wind turbines: A review," *Renewable and sustainable energy reviews*, vol. 104, pp. 379–393, 2019.

[2] C. Jung, L. Sander, and D. Schindler, "Future global offshore wind energy under climate change and advanced wind turbine technology," *Energy Conversion and Management*, vol. 321, p. 119075, 2024.

[3] S. Y. Oh, C. Joung, S. Lee, Y.-B. Shim, D. Lee, G.-E. Cho, J. Jang, I. Y. Lee, and Y.-B. Park, "Condition-based maintenance of wind turbine structures: A state-of-the-art review," *Renewable and Sustainable Energy Reviews*, vol. 204, p. 114799, 2024.

[4] H. Ali, X. Zhai, U. U. Tariq, and L. Liu, "Energy efficient heuristic algorithm for task mapping on shared-memory heterogeneous mpsocs," in *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, 2018, pp. 1099–1104.

[5] Z. Ren, A. S. Verma, Y. Li, J. J. Teuwen, and Z. Jiang, "Offshore wind turbine operations and maintenance: A state-of-the-art review," *Renewable and Sustainable Energy Reviews*, vol. 144, p. 110886, 2021.

[6] F. Earle, J. Huddlestone, T. Williams, C. Stock-Williams, H. van der Mijle-Meijer, L. de Vries, H. van Heemst, E. Hoogerwerf, L. Koomen, E.-J. de Ridder *et al.*, "Spowtt: Improving the safety and productivity of offshore wind technician transit," *Wind Energy*, vol. 25, no. 1, pp. 34–51, 2022.

[7] O. Gudmestad and S. Viddal, "Personnel-transfer from vessel to offshore floating wind turbine," *Trends in Renewable Energies Offshore*, pp. 517–523, 2022.

[8] N. Holliday and L. D. Otero, "A simulation model for the optimization of crew transport vessels to service offshore wind farms," in *2023 IEEE International Systems Conference (SysCon)*. IEEE, 2023, pp. 1–6.

[9] J. Li, Q. Miao, Z. Zou, H. Gao, L. Zhang, Z. Li, and N. Wang, "A review of computer vision-based monitoring approaches for construction workers' work-related behaviors," *IEEE Access*, vol. 12, pp. 7134–7155, 2024.

[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.

[11] F. Anderson, "Real time, video image centroid tracker," in *Acquisition, Tracking, and Pointing IV*, vol. 1304. SPIE, 1990, p. 82.