POLISH MARITIME RESEARCH 1 (113) 2022 Vol. 29; pp. 63-75
10.2478/pomr-2022-0007

AUTOMATED MOTION HEATMAP GENERATION FOR BRIDGE NAVIGATION WATCH MONITORING SYSTEM

Veysel Gokcek *[1]
Gazi Kocak [1]
Yakup Genc [2]
[1] Istanbul Technical University, Tuzla, Istanbul, Turkey
[2] Gebze Technical University, Gebze, Kocaeli, Turkey

* Corresponding author: *gokcekv@itu.edu.tr (V.Gokcek)*

ABSTRACT

Most ship collisions and grounding accidents are due to errors made by watchkeeping personnel (WP) on the bridge. International Maritime Organization (IMO) adopts the resolution on the Bridge Navigation Watch Alarm System (BNWAS) detecting operator disability to avert these accidents. The defined system in the resolution is very basic and vulnerable to abuse. There is a need for a more advanced system of monitoring the behaviour of WP to mitigate watchkeeping errors. In this research, a Bridge Navigation Watch Monitoring System (BNWMS) is suggested to achieve this task. Architecture is proposed to train a model for BNWMS. The literature reveals that vision-based sensors can produce relevant input data required for model training. 2D body poses belonging to the same person are estimated from multiple camera views by using a deep learning-based pose estimation algorithm. Estimated 2D poses are projected into 3D space with a maximum 8 mm error by utilising multiple view computer vision techniques. Finally, the obtained 3D poses are plotted on a bird's-eye view bridge plan to calculate a heatmap of body motions capturing temporal, as well as spatial, information. The results show that motion heatmaps present significant information about the behaviour of WP within a defined time interval. This automated motion heatmap generation is a novel approach that provides input data for the suggested BNWMS.

Keywords: Safety of Navigation, Navigation Watch, Deep Learning, 3D Body Pose

INTRODUCTION

Many studies have shown that maritime accidents are often the result of human error. Along with improved ship design and technology, there has been noticeable progress on accident analysis [1, 2] and risk management [3, 4]. This progress contributed to a 50% drop in reported shipping losses in 2020, compared to 2011 [5]. Besides this, maritime accidents have declined globally since 2013 [6]. However, the numbers are still large enough to threaten marine ecosystems, the environment and local economies when considering the catastrophic consequences of such incidents [7]. In particular, collisions and groundings have the potential to cause catastrophic results. Analyses of collisions and groundings show that 96.5% of errors occur on the bridge, where the main actor involved has been the officer in charge of the navigation watch (OOW) [8]. Thus, proper watchkeeping during ship navigation is of great importance, to prevent pollution of the marine environment and loss of both life and property.

The master of the ship shall ensure that watchkeeping arrangements on the ship are adequate for safe navigation [9]. The most important issue in arranging watchkeeping is the competence and fitness of OOW and the lookout. OOW should understand individual and team roles and responsibilities during a navigational watch, and an OOW should be familiar with all navigational installations and equipment. The lookout must know how to keep a continuous and proper look-out. Watchkeeping personnel (WP), both the OOW and the lookout, should not deal with any other duties or actions other than their responsibilities related to navigational watch.

Proper watchkeeping and the responsibilities of WPs are well defined in the International Convention on Standards of Training, Certification, and Watchkeeping for Seafarers (STCW) Chapter VIII – Standards Regarding Watchkeeping [9]. There are routines to fulfil those watchkeeping standards, such as fixing the ship's position frequently, maintaining visual look-out, checking the track, monitoring the navigational hazards, verifying compass input and steering, etc. [10]. Deviation from these routines means that the watch is below standard and a situation may occur, forming the basis for the occurrence of a maritime accident. Poor look-out and insufficient use of navigation equipment are among the watchkeeping routines that are the root causes of groundings and collisions [11]. To prevent such non-conformities, the master of the ship controls WPs during navigational watch, by randomly visiting and checking them. However, the master cannot control all navigation watches during the whole voyage. Thus, International Maritime Organization (IMO) adopts the Resolution MSC.128(75) Performance Standards for a Bridge Navigation Watch Alarm System (BNWAS), to detect operator disability which could lead to marine accidents [12]. The system monitors the absence of watchkeeping on the bridge and automatically alerts the Master or the backup officer or, even, all of the crew. However, the defined system in the resolution is very basic and vulnerable to abuse.

BNWAS has a reset mechanism which is a combination of push-buttons and motion detectors on the bridge, as well as event listeners on the electronic navigation equipment. If it is not reset within the manually defined period (between 3 and 12 minutes), alarm stages start from the bridge to all necessary locations. While this approach can detect the absence or disability of WP on the bridge, it is not sufficient to evaluate whether there is proper watchkeeping. Resetting BNWAS at every period does not mean that WPs follow their watchkeeping responsibilities based on STCW, it is just proof that there is a WP on the bridge, even when the WP is drowsy or affected by fatigue. Along with BNWAS, there is a need for a more advanced evaluation system to enhance the safety of navigation.

An artificial intelligence-based automated system that continuously monitors the behaviour of WPs improves ship navigation safety. This system detects nonconformities and gives feedback to improve the behaviour of WPs by monitoring their watchkeeping performance. Also, WPs keep a proper watch if they know that they are continuously being monitored. In this respect, an architecture is proposed to train a model for a Bridge Navigation Watch Monitoring System (BNWMS). In this architecture, vision-based sensors are suitable for the bridge environment to collect training data.

This study focuses on the automated motion heatmap generation of WP during navigation watch. A multi-video camera system is established on the actual bridge. Multiple cameras are calibrated to enable 3D projection. 2D body poses belonging to the same person from multiple camera views are estimated by using a deep learning-based pose estimation algorithm. The backward projection method, with camera parameters, is utilised to construct a 3D body pose from estimated 2D poses. An error function is defined to eliminate incorrectly calculated 3D body poses, while maximum and minimum lengths for body parts are assigned to validate the results. A heatmap of body motions is generated by plotting validated 3D body poses in the bird's-eye view (2D) bridge plan. This automated motion heatmap generation, presenting both temporal and spatial information, is a novel approach that enables the training of a deep learning-based model monitoring the behaviour of WP during navigation.

The article is organised as follows. Section 2 gives information about the study's background and related literature. Section 3 outlines the methodology used in the analysis, while Section 4 details the results of a case study. Section 5 discusses the research findings. The final section remarks on the implications and concludes the article.

BACKGROUND

Although there is no research on the behaviour of seafarers, based on machine learning or deep learning techniques, many different methods are proposed for the analysis of human behaviour, in terms of human activity recognition in many areas. Human activity recognition determines body posture, movements, and actions using multimodal data from various sensors. Previous studies on the recognition of human activities can be categorised, broadly based on the sensors used. These categories include vision-based sensors, wearable sensors, mobile phone sensors, and social network sensors [13]. Vision-based sensors produce images that enable the recognition of many different activities [14]. Mobile phones, smartwatches, and other wearable sensors calculate cardiovascular parameters or inertial data to monitor health conditions and sports activities [15]. Social network sensors enable an understanding of users' behaviour and interests, if they actively use social media sites [16].

Both visual and wearable sensors can extract relevant data for the behaviour analysis of WP during navigation. While examining restricted areas, like a bridge, many types of research use vision sensor technologies, such as RGB cameras, for event monitoring and recognition, rather than relying on wearable sensors [17]. Besides, the use of wearable sensors can distract WP, and reduce awareness and performance. Thus, the use of image data would be more appropriate for the bridge environment. Models have been developed which detect and identify emotions [18], gestures [19], mouth and flexion movements [20], eye movements [21] skeletal structures [22], and physical activities [23], using only image data with a deep learning approach. For flexion, eye movement, and emotion recognition, the face of the person should be constantly monitored with a high-resolution camera. These models are mostly trained to assess drivers and pilots. However, WP on the bridge does not sit in a fixed position like a driver or pilot but walks in a wide area during a watch. Therefore, it is difficult to monitor the permanent face of WP on the watch. Besides, the area that WPs occupy during a navigation watch is important to understanding watchkeeping behaviour. Thus, constructing 3D body pose features from vision sensors and plotting them onto a 2D bridge plan is an appropriate approach for the bridge environment.

The simplified architecture shown in Fig. 1 is proposed to train a model for BNWMS on the behaviour analysis of seafarers. In this model, the input is the tracked 3D body-pose features (i.e. x_{ij} is the j^{th} pose parameter at the i^{th} frame) while the output is the label for the WP's behaviour in that time (i.e. y_k is the class label). Classification labels can easily be obtained by expert evaluation (for example, an expert can label the actions of the WP seen in the video). However, the body pose features within a defined time interval constitute a high dimensional space. Those features should be converted into simpler input data, useful for behaviour analysis. It is assumed that knowing the areas that WPs occupy on the bridge is an essential input to comprehend what they are doing during a navigational watch. Thus, we focus on motion heatmaps of the 3D body-poses of WPs.

Firstly, there should be multiple camera views to extract 3D body-pose features. When using multiple cameras to analyse the same scene, camera calibration is a necessary step. In the context of multiple view geometry in computer vision, camera calibration is used to find the camera's intrinsic and extrinsic parameters. While intrinsic parameters refer to optical characteristics and internal camera geometry, extrinsic parameters are the 3D position and orientation of each camera frame relative to a world coordinate system [24]. Those parameters enable mapping between 3D world coordinates and 2D image coordinates. There are various camera calibration algorithms and they can be classified as linear, nonlinear, and multi-step techniques [25, 26]. Multi-step techniques are more accurate than linear methods and are faster than nonlinear methods. The four-step camera calibration developed by Heikkila & Silven [27] and the bias-corrected version of this algorithm developed by Heikkila [28], are well-known and most-used in the multi-step category. In this study, the latest version of Heikkila's camera calibration is utilised.

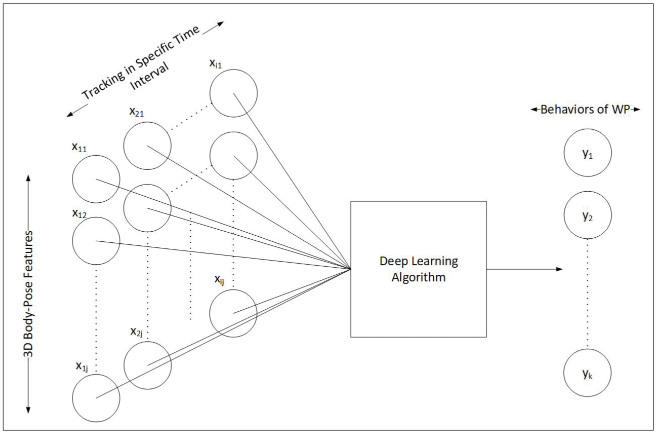


Fig. 1. Architecture for BNWMS

Secondly, a 2D pose estimation algorithm is required to estimate body parts from each camera. When a body pose is detected on at least two camera views, a 3D pose can be constructed by using camera parameters. The pose estimation algorithm must work in real-time and detect multiple persons in the scene. There are top-down and bottom-up approaches to detect people and their poses in the image. Top-down methods [29, 31] detect people first and then estimate their body parts on each detected region; this is followed by calculation of the relevant pose of each person. On the contrary, bottom-up methods detect all joints first, then associate them to create a possible pose for each person [22, 32]. Each method has its advantages, however, for this study, fast multi-person pose estimation is necessary to assess real-time behaviours of WP during a watch.

Top-down methods apply a person detector to detect people and use single-person pose estimation (SPPE) for each detected person. Since those methods need to detect each person independently, they show less accuracy in overlapping situations. Also, running the SPPE model for every person consumes time, depending on the number of detected people. On the other hand, bottom-up methods find all joints and combine them properly for each possible person in the image. They take the most time to pair corresponding joints. However, grouping the detected joints is less costly than repeating the SPPE for each detected person. Bottom-up methods can extract body parts correctly, even if there are overlapped people. The performances of the state-of-the-art bottom-up method OpenPose [22] and the top-down method HRNet [31] are compared. Each model was run on the same 4 hour navigation watch. OpenPose is approximately 24 times faster than HRNet, while HRNet seems more accurate than OpenPose. Since it is not possible to maintain real-time analysis by HRNet, we considered using OpenPose despite its slightly lower accuracy.

The final step is to construct 3D poses and define a plotting procedure to generate the motion heatmap of those poses. The next section gives more detailed information about the methodology.

METHODOLOGY

The research methodology to generate a motion heatmap consists of a deep learning-based pose estimation algorithm and multiple view computer vision techniques. 2D body poses belonging to the same person are estimated by the OpenPose algorithm from multiple camera views. The camera parameters are estimated by camera calibration, to project 2D body poses into 3D space. The constructed 3D body poses are then plotted on the 2D bridge plan, to generate a heatmap of body motion.

OpenPose is the real-time, multi-person 2D pose estimation algorithm based on a multi-stage Convolution Neural Network (CNN). The first stage of CNN predicts confidence maps of body part locations, while the second stage (called part affinity fields - PAF) encodes the degree of association between parts. The grouping of keypoint instances is carried out by using confidence maps and the PAFs together, in order to output the 2D keypoints for all people in the image. It predicts 25 keypoints for each person, as shown in Fig. 4. The detailed methodology of OpenPose is presented in [22].

Camera calibration estimates unknown parameters of the camera model. Heikkila’s geometric camera calibration model [28], based on the perspective projection, is utilised to find unknown camera parameters. Unknown parameters can be divided into intrinsic and extrinsic parameters. Extrinsic parameters enable the transformation of world coordinates (X, Y, Z) to camera coordinates (x, y, z), while intrinsic parameters provide the mapping of camera coordinates to pixel coordinates (u, v). This operation is the forward projection shown in Fig 2.

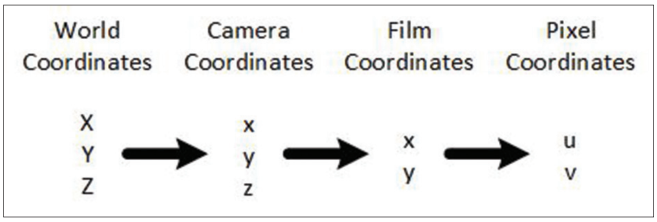


Fig. 2. Forward projection

In the backward projection, there is a reverse application of the forward projection shown in Fig. 3. By using one camera output, there are always two equations with three unknowns.

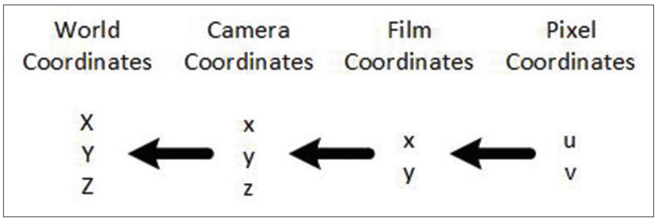


Fig. 3. Backward projection

If the Z dimension in the 3D plane is known, in addition to the single-camera image points, the other two world coordinates can be calculated. However, since the information of the Z dimension is not available, it is not possible to calculate the 3D coordinates using the single-camera data. Instead, if the pixel coordinates in the same spot are known from a second camera, all of the 3D coordinates, along with the Z dimension, can be calculated. Since two different sets of pixel data from two cameras belong to the same point in the world coordinates, a series of equations emerge:

$$p_{c_i} = \begin{bmatrix} mu_{c_i} \\ mv_{c_i} \\ m \end{bmatrix} = IE_i \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

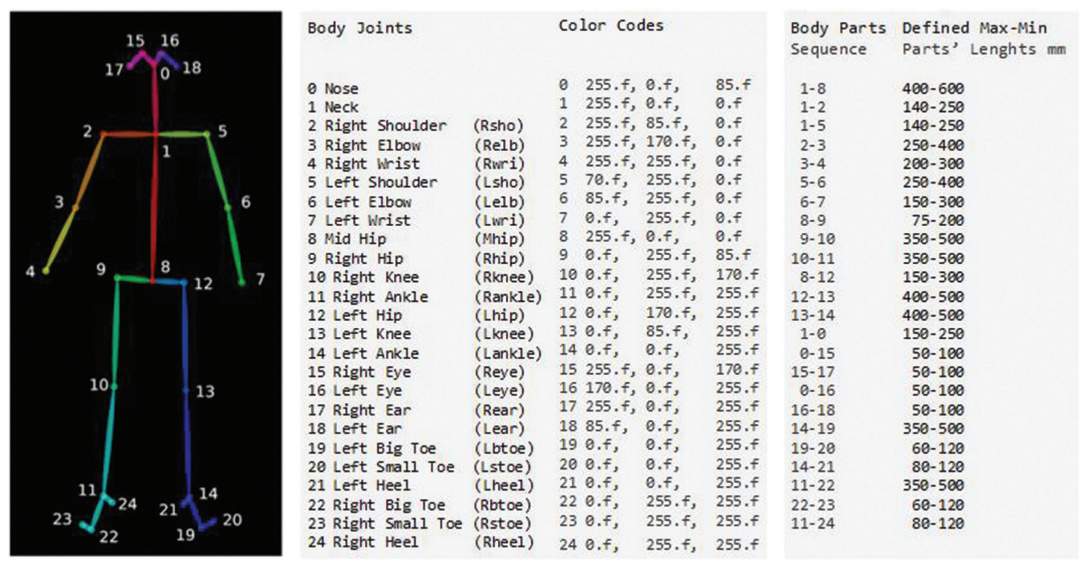
$$eqn1_i = (p_{c_i}(1{:}1,:)/\ p_{c_i}(3{:}3,:) == p_{c_i}(1{:}1,:));$$

$$eqn2_i = (p_{c_i}(2{:}2,:)/\ p_{c_i}(3{:}3,:) == p_{c_i}(2{:}2,:)); \qquad (1)$$

where I is the intrinsic parameters matrix, E_i is a combination of extrinsic parameters of the th camera, p_{c_i} are the calculated pixel points of the ith camera from unknown parameters of X, Y, Z coordinates, and p_i is a real observed point of the ith camera. The triangulation function transforms p_i into X, Y, Z [9]. There is an error function to verify the solution and this reprojects X, Y, Z into each of the th camera coordinates, to re-calculate pixel points (rp_{c_i}). The comparison between rp_{c_i} and original p_i is made over a threshold value. This value is equal to two times the camera calibration error because there are two calculation processes with calibration parameters including X, Y, Z and rp_{c_i}. The performance of the 3D pose construction depends on the difference between rp_{c_i} and p_i. To be acceptable, this difference should be smaller than the threshold value.

Since there is always more than one WP on the bridge, the OpenPose algorithm finds multiple poses. At that point, the problem of correct matching 2D poses from different views arises. If OpenPose finds two people on two camera views, the backward projection algorithm would create four possible 3D poses. Although the defined threshold value for backward projection eliminates the wrong 3D poses, max-min length values for detected body parts are also assigned. These length limits validate the result of the backward projection algorithm. Fig. 4 shows the pose format of OpenPose and defined max-min lengths for detected body parts.

Due to occlusions on the bridge, backward projection may not be able to construct a complete 3D pose. So, only essential


15
16
17
0
18
2
5
1
3
6
4
9
8
12
7
10
13
11
24
21
14
23
22
19
20

Body Joints		Color Codes				Body Parts Sequence	Defined Max-Min Parts’ Lenghts mm
0 Nose		0	255.f,	0.f,	85.f	1-8	400-600
1 Neck		1	255.f,	0.f,	0.f	1-2	140-250
2 Right Shoulder	(Rsho)	2	255.f,	85.f,	0.f	1-5	140-250
3 Right Elbow	(Relb)	3	255.f,	170.f,	0.f	2-3	250-400
4 Right Wrist	(Rwri)	4	255.f,	255.f,	0.f	3-4	200-300
5 Left Shoulder	(Lsho)	5	70.f,	255.f,	0.f	5-6	250-400
6 Left Elbow	(Lelb)	6	85.f,	255.f,	0.f	6-7	150-300
7 Left Wrist	(Lwri)	7	0.f,	255.f,	0.f	8-9	75-200
8 Mid Hip	(Mhip)	8	255.f,	0.f,	0.f	9-10	350-500
9 Right Hip	(Rhip)	9	0.f,	255.f,	85.f	10-11	350-500
10 Right Knee	(Rknee)	10	0.f,	255.f,	170.f	8-12	150-300
11 Right Ankle	(Rankle)	11	0.f,	255.f,	255.f	12-13	400-500
12 Left Hip	(Lhip)	12	0.f,	170.f,	255.f	13-14	400-500
13 Left Knee	(Lknee)	13	0.f,	85.f,	255.f	1-0	150-250
14 Left Ankle	(Lankle)	14	0.f,	0.f,	255.f	0-15	50-100
15 Right Eye	(Reye)	15	255.f,	0.f,	170.f	15-17	50-100
16 Left Eye	(Leye)	16	170.f,	0.f,	255.f	0-16	50-100
17 Right Ear	(Rear)	17	255.f,	0.f,	255.f	16-18	50-100
18 Left Ear	(Lear)	18	85.f,	0.f,	255.f	14-19	350-500
19 Left Big Toe	(Lbtoe)	19	0.f,	0.f,	255.f	19-20	60-120
20 Left Small Toe	(Lstoe)	20	0.f,	0.f,	255.f	14-21	80-120
21 Left Heel	(Lheel)	21	0.f,	0.f,	255.f	11-22	350-500
22 Right Big Toe	(Rbtoe)	22	0.f,	255.f,	255.f	22-23	60-120
23 Right Small Toe	(Rstoe)	23	0.f,	255.f,	255.f	11-24	80-120
24 Right Heel	(Rheel)	24	0.f,	255.f,	255.f		

Fig. 4. OpenPose body joints, colour codes, body parts sequence [22], and defined limits

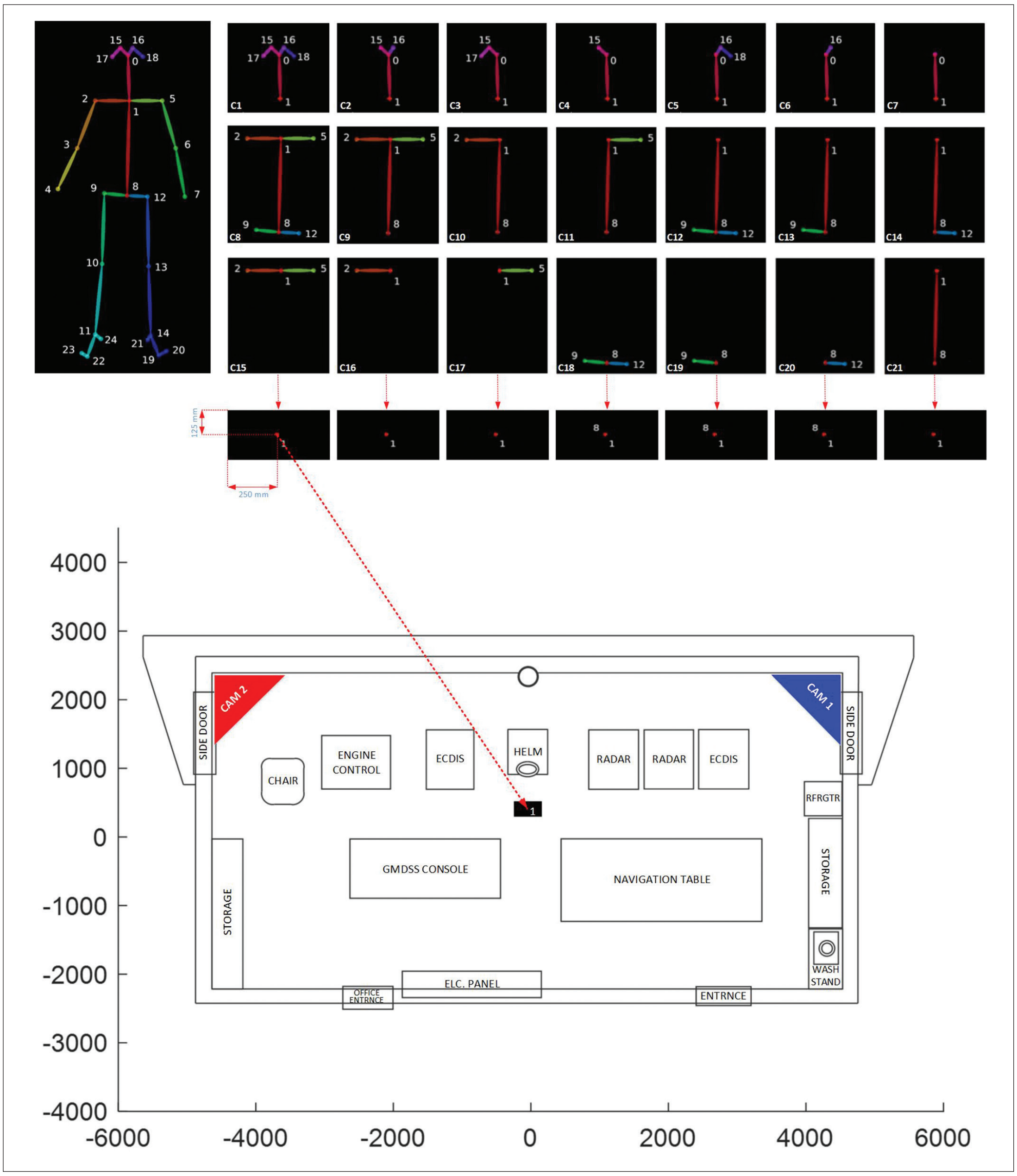


Fig. 5. Plotting 3D poses on the 2D bridge plan

body parts are defined. Fig. 5 shows the required body parts of the 3D pose and its plotting on the 2D bridge plan. Our algorithm seeks those multiple parts in a given sequence, *C1* to *C21*, based on those detected first. If any of the combinations are found within validated 3D poses, the algorithm plots it on the 2D bridge plan. While plotting, Neck or Mid Hip is adjusted as a centre of rectangular dimensions 250 mm x 500 mm (height x width).

Plotting all of the validated 3D poses of WPs on the 2D bridge plan, within a defined time interval, creates a map. A heatmap of those plotted poses shows us the motions of WP during a watch. The more heated areas on the map represent the more occupied locations during a watch. Since the location of the navigation equipment is known, the assessment of watches can be inferred based on which locations the WP occupied.

CASE STUDY

During the case study, a night watch was used because vision-based pose estimation during the nighttime is more challenging than in the daytime. Data was collected from a real bridge environment. A camera system was established on the bridge of a bulk carrier. The installed video cameras had clear night vision that would not disturb the WPs. Each video camera location and angle of view was adjusted to maximise the field of view and minimise the blind spots. The resolution of the recorded camera views was 2560 x 1944 pixels, which was enough to identify body pose. Three months of video recording data were collected from multiple camera views. The proposed methodology was applied to generate motion heatmaps and the results are given in the following subsections.

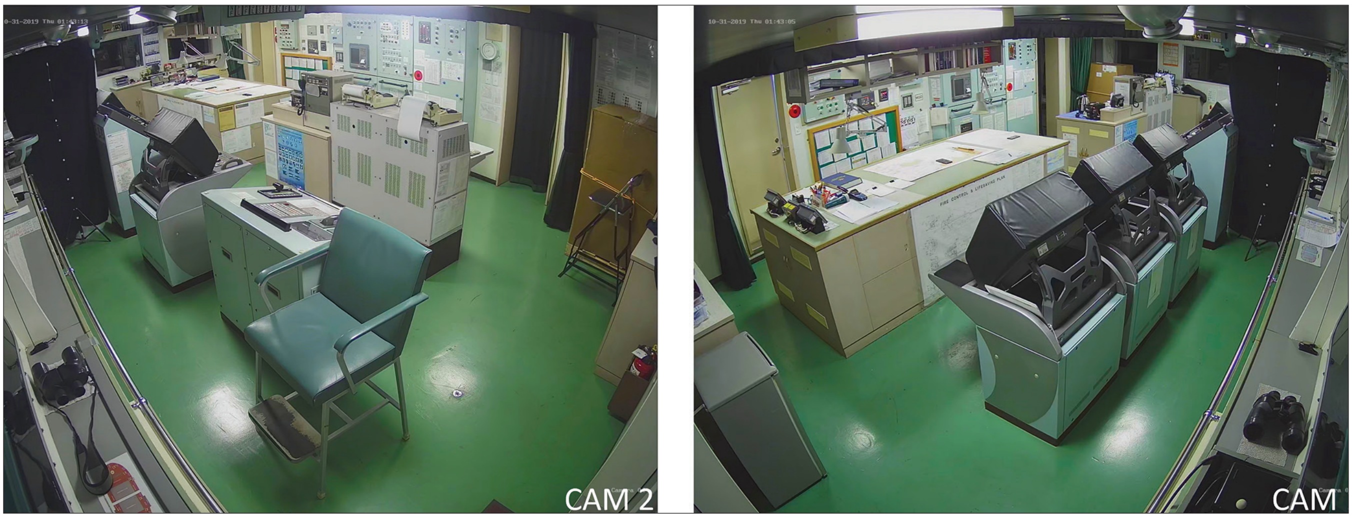


Fig. 6. Bridge camera views with the calibration object

CAMERA CALIBRATION

The calibration procedure required the 3D coordinates of the control points and the corresponding pixel coordinates for those control points in each camera view. Our data comprised video recordings from two video cameras on the bridge. Each camera was positioned differently, with a special orientation to reduce blind spots. Sample images of the camera views and calibration objects are shown in Fig. 6. 380 mutual control points were defined. *X, Y, Z* coordinates of each control point and their observed *u, v* points were recorded for each camera. The recorded input data was used to conduct camera calibration by the Heikkila method [28].

During calibration optimisation, the intrinsic and extrinsic parameters of the cameras were calculated in 20 iterations. Fig. 7 shows the extrinsic parameters for two of the cameras, along with the control points used; Table 1 shows the recovered intrinsic parameters (assumed to be the same for all cameras) with average pixel error.

Tab.1. Camera Intrinsic Parameters

Focal Length	[1757.47 1761.57] +/- [5.27 5.24]
Principal point	[1254.80 956.48] +/- [11.52 7.11]
Skew	[0.00] +/- [0.00]
Distortion	[-0.32 -0.08 -0.00003 -0.00004 0.00] +/- [0.007 0.007 0.00 0.00 0.00]
Pixel error	[2.97 3.00]

An error of 2.97-3.00 pixels in (*u, v*) coordinates corresponds to a maximum 8 mm error in the world coordinate system. These error values are acceptable for detecting and tracking human body parts in the bridge environment. The threshold values for reprojection errors were set to 5.94-6.00 pixelswhich is two times the camera calibration error.

3D POSE CONSTRUCTION

By using calibrated multiple camera views, the body pose of WP could be mapped and tracked in 3D space. However, there

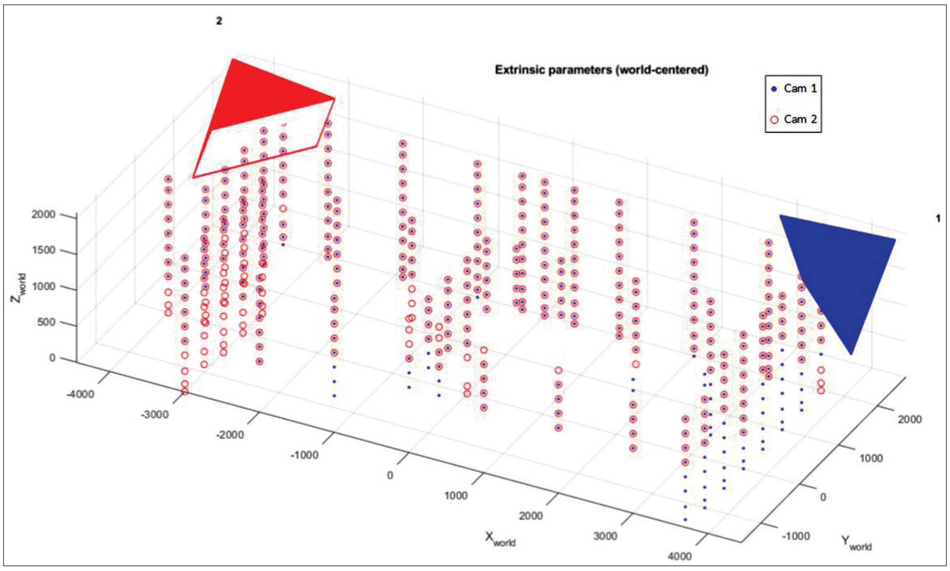


Fig. 7. Camera extrinsic parameters with used control points

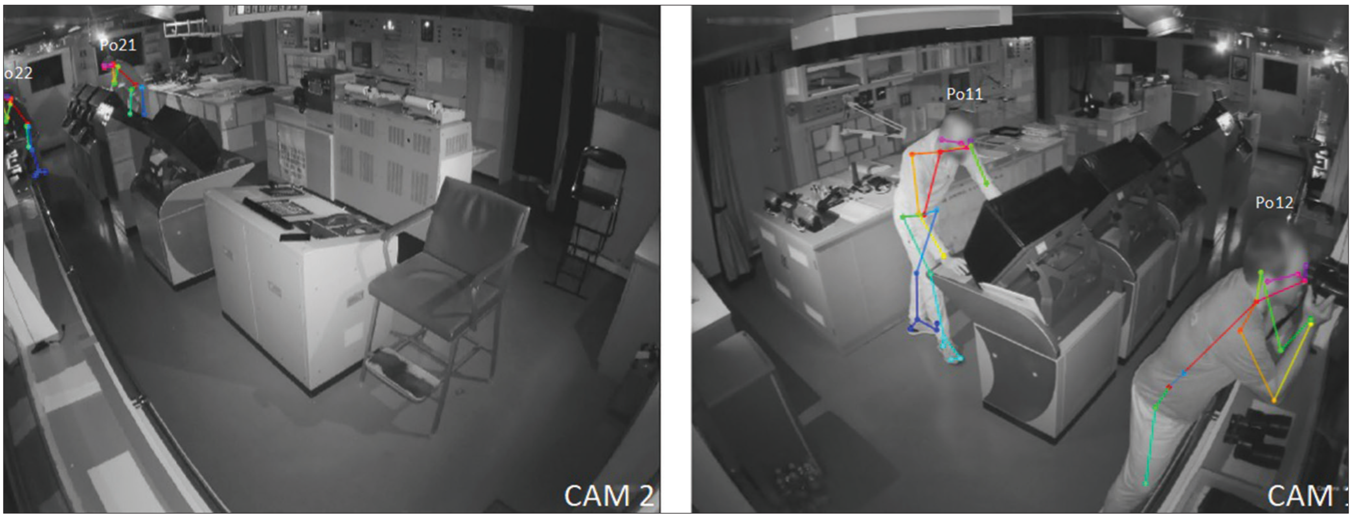


Fig. 8. Results of pose estimation algorithm on each camera image

should be 2D poses from each camera, to construct the 3D poses. Therefore, a real-time multi-person 2D pose estimation algorithm – OpenPose (developed by Cao et al. [22]) – was utilised to estimate body poses from multiple views.

The pose detection algorithm seeks the pixel data of each joint on each camera image, then combines them properly for each possible person in the image. As shown in Fig. 8, poses for two people were detected on both camera views.

Openpose estimated Po_{11} and Po_{12} on camera CAM1 and Po_{21} and Po_{22} on camera CAM2. This means that the backward

Tab. 2. Main particulars of the barge model

BODY JOINTS		Pixel Coordinates of Detected Body Joints (u-v)				World Coordinates Of Possible Body Joints (X-Y-Z)			
		CAM1		CAM2		Pe_1	Pe_2	Pe_3	Pe_4
		Po_{11}	Po_{12}	Po_{21}	Po_{22}				
0	Nose	1048-546	2336-1054	443-242	48-358	2783-812-1450	-	4292-2401-1670	3235-2531-1450
1	Neck	947-562	2151-1134	480-232	84-367	2826-645-1434	3664-1594-1431	4168-2230-1597	3270-2392-1355
2	Rsho	839-572	2089-1245	476-229	87-360	2993-612-1448	3725-1574-1397	4207-2215-1611	3443-2377-1362
3	Relb	865-801	2218-1511	480-294	73-433	3021-635-1142	-	4175-2272-1413	3336-2525-1079
4	Rwri	960-960	2358-1219	-	52-376	-	-	4209-2358-1562	3297-2559-1357
5	Lsho	1060-538	2168-1022	495-239	85-381	2592-644-1411	3432-1588-1421	4106-2228-1563	3040-2384-1327
6	Lelb	1123-683	2244-1320	482-302	68-447	2567-734-1177	3967-1677-1249	4070-2305-1400	3058-2525-1073
7	Lwri	-	2358-1202	-	46-385	-	-	-	3237-2575-1338
8	Mhip	885-802	1818-1460	564-310	151-457	2764-352-1000	3432-1314-888	3886-1920-1221	3231-2168-879
9	Rhip	804-811	1766-1539	546-327	146-465	2917-361-977	3444-1324-791	3967-1941-1213	3325-2177-847
10	Rknee	909-1029	1729-1828	541-418	154-544	2849-465-592	3627-1393-358	3805-1923-920	3385-2224-476
11	Rankle	960-1270	-	-	-	-	-	-	-
12	Lhip	934-784	1875-1406	587-319	154-468	2610-302-962	3397-1314-890	3835-1913-1193	3092-2175-847
13	Lknee	856-1025	-	595-420	168-557	2801-247-518	-	3814-1865-869	-
14	Lankle	859-1202	-	-	194-637	-	-	3717-1782-536	-
15	Reye	1026-530	2308-1029	440-233	53-347	2824-812-1486	-	4287-2378-1692	3284-2502-1485
16	Leye	1061-520	2339-997	445-235	55-351	2750-808-1480	-	4265-2369-1679	3180-2509-1469
17	Rear	952-516	2192-1058	470-232	-	2826-673-1473	3612-1629-1465	-	-
18	Lear	-	-	-	75-354	-	-	-	-
19	Lbtoe	933-1242	-	-	185-645	-	-	3652-1831-503	-
20	Lstoe	934-1217	-	-	179-655	-	-	3667-1863-507	-
21	Lheel	833-1239	-	-	217-640	-	-	3689-1682-461	-
22	Rbtoe	1024-1350	-	-	-	-	-	-	-
23	Rstoe	987-1355	-	-	-	-	-	-	-
24	Rheel	957-1303	-	-	-	-	-	-	-

Tab. 2. Pixel points of each detected joint and possible world coordinates of those joints

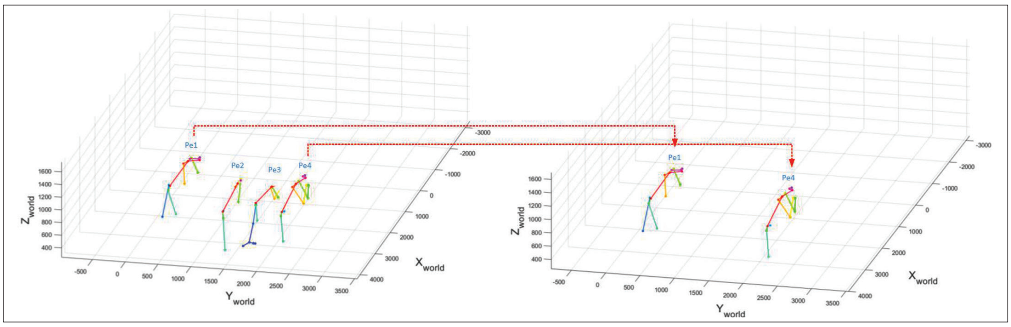


Fig. 9. 3D pose of WPs

projection algorithm produced four possible 3D poses. The pixel points of the joints were detected by OpenPose and the 3D coordinates of those joints were calculated by backward projection and are listed in Table 2. Pe_1, Pe_2, Pe_3, and Pe_4 are possible 3D poses created by the combination of Po_{11}, Po_{12}, Po_{21} and Po_{22}. When obtaining the 3D poses, only the joints detected on both camera views were used by the backward projection algorithm.

Table 3 shows reprojection errors for the calculated 3D coordinates of joints and the lengths of body parts obtained from those joints. While both Pe_1 and Pe_4 have acceptable reprojection errors and part lengths, Pe_2 and Pe_3 have large reprojection errors and unacceptable lengths for more than 50% of detected body parts.

3D plotting of all the estimated 3D poses and validated 3D poses, after elimination, is shown in Fig. 9. It can be seen that validated 3D poses of WPs explain the real poses and locations of WPs shown in Fig. 8. Since some joints are not detected on both cameras, some body parts are missing in the constructed 3D poses. However, essential body part combination *C2* is detected for both Pe1 and Pe_4. This means that validated 3D poses can be mapped on the 2D bridge plan.

Tab. 3. Reprojection errors and lengths of body parts

BODY JOINTS		Reprojection errors of calculated 3D Body Joints (<6 pixels)				Body Parts Sequence	Lengths of Body Parts (mm)				
		Pe1	Pe2	Pe3	Pe4		Filters	Pe1	Pe2	Pe3	Pe4
0	Nose	-	240-818	-	3-5	1-8	400-600	528	**655**	563	528
1	Neck	2-8	163-591	696-109	5-3	1-2	140-250	171	**72**	**44**	174
2	Rsho	1-2	160-597	696-136	1-0	1-5	140-250	235	233	**71**	231
3	Relb	5-11	252-647	-	4-3	2-3	250-400	308	-	**209**	336
4	Rwri	-	288-549	-	7-1	3-4	200-300	-	-	**176**	283
5	Lsho	2-6	158-580	614-102	14-10	5-6	250-400	252	**569**	**183**	291
6	Lelb	1-2	226-649	848-60	6-2	6-7	150-300	-	-	-	323
7	Lwri	-	-	-	4-1	8-9	75-200	154	98	83	100
8	Mhip	7-7	149-420	521-142	2-1	9-10	350-500	405	475	**336**	378
9	Rhip	4-8	161-467	482-145	9-7	10-11	350-500	-	-	-	-
10	Rknee	2-4	189-416	473-105	4-2	8-12	150-300	166	**35**	**59**	142
11	Rankle	-	-	-	-	12-13	400-500	487	-	**328**	-
12	Lhip	7-5	146-421	550-120	1-1	13-14	400-500	-	-	**357**	-
13	Lknee	1-1	184-426	-	-	1-0	150-250	173	-	224	172
14	Lankle	-	192-389	-	-	0-15	50-100	54	-	**32**	66
15	Reye	0-1	219-775	-	2-6	15-17	50-100	140	-	-	-
16	Leye	-	209-753	-	3-1	0-16	50-100	44	-	43	63
17	Rear	2-7	-	676-96	-	16-18	50-100	-	-	-	-
18	Lear	-	-	-	-	14-19	350-500	-	-	**87**	-
19	Lbtoe	-	196-366	-	-	19-20	60-120	-	-	**35**	-
20	Lstoe	-	203-394	-	-	14-21	80-120	-	-	**128**	-
21	Lheel	-	178-355	-	-	11-22	350-500	-	-	-	-
22	Rbtoe	-	-	-	-	22-23	60-120	-	-	-	-
23	Rstoe	-	-	-	-	11-24	80-120	-	-	-	-
24	Rheel	-	-	-	-	Detected Parts		13	7	17	13
Avg error		3.5	**364.6**	**365.2**	3.9	Acceptable Parts		13	**3**	**4**	13

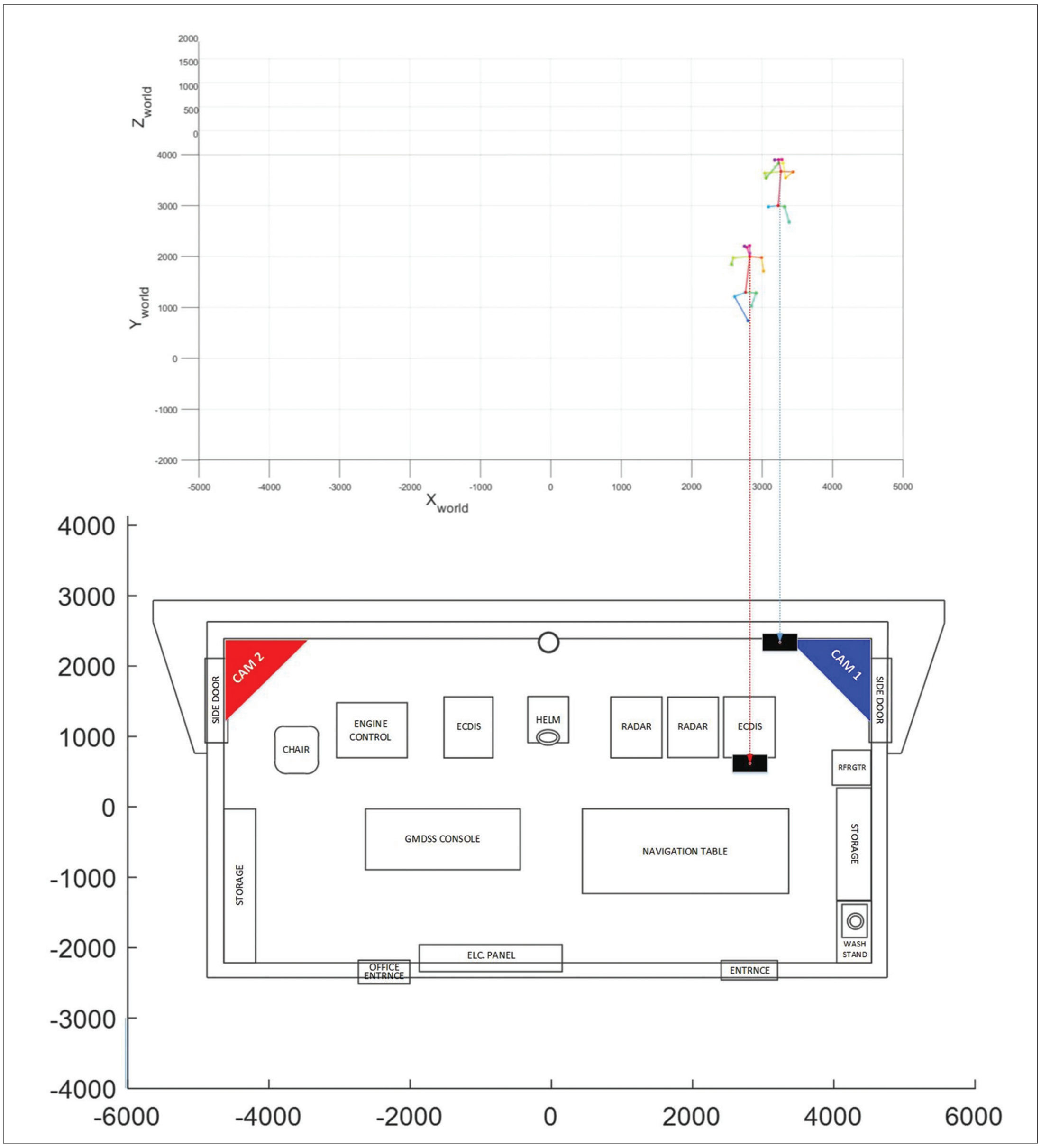


Fig. 10. Plotting 3D poses on the 2D bridge plan

MOTION HEATMAP GENERATION

Estimated 3D poses of WPs are plotted onto the 2D bridge plan shown in Fig. 10. The position of the Neck is adjusted as the centre of the rectangle and the dimensions are 250 mm x 500 mm (height x width).

As shown in Fig. 10, one WP is standing on the Electronic Chart Display and Information System (ECDIS), which is the navigating software program, while another is in the look-out area, which is the infront of the electronic navigation equipment. This may mean that the officer WP is checking the position of the vessel on the ECDIS, while the lookout WP is continuing as a look-out to detect collision situations. However, a single screenshot of the watch is not enough for that assessment. The evaluation should be based on the behaviour of WPs within a time interval. Heatmap plotting of estimated 3D poses on the 2D bridge plan, within defined time intervals, is assumed to fulfil evaluation criteria. Although there should be a continuous and proper look-out at all times, during dense marine traffic, more attention should be given by both WPs. As well as a look-out, proper use of other electronic navigation equipment is also essential to avoid collisions. A ship's position should be checked periodically, depending on the proximity to shallow waters and how dense the marine traffic is. Broadly speaking, the importance of watchkeeping can be classified depending on navigation areas, such as shallow waters, coastal waters, and open seas. Heatmaps should be adjustable, based on those navigation areas. The defined period for BNWAS is 3 to 12 minutes [12]. The same intervals can be assigned for our heatmaps as follows: 3 minutes for shallow waters, 6 minutes for coastal waters, and 12 minutes for open seas. In this study, the motion heatmaps generated within 12 minute intervals are presented in Fig.11.

Visual checks show that, during Case 1, the lookout stays and walks on the look-out area, while the OOW uses the

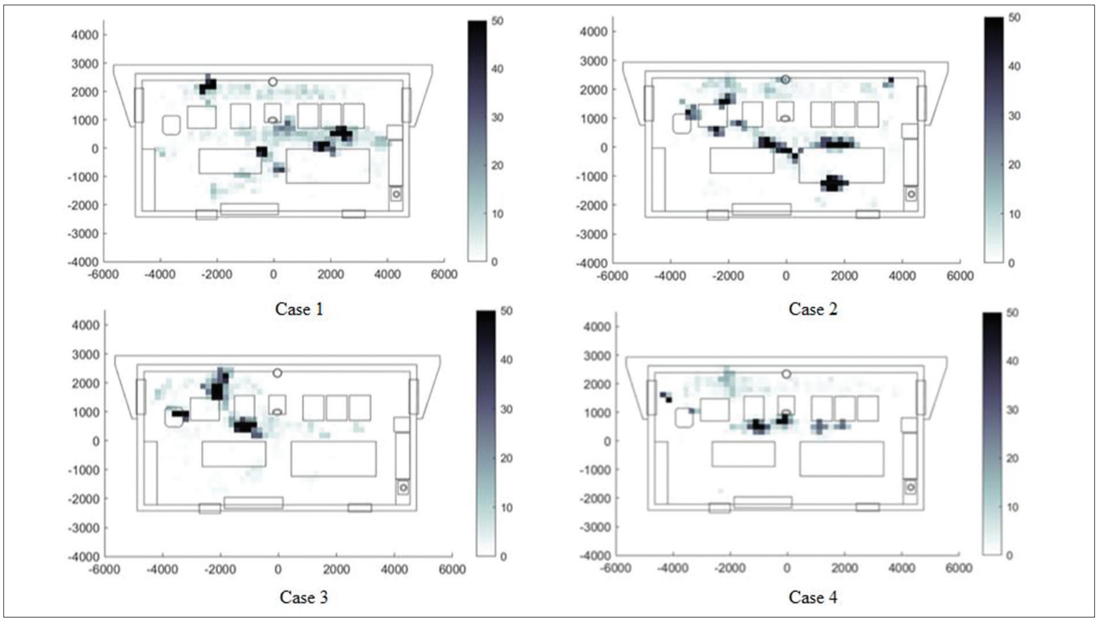


Fig. 11. Motion heatmaps of 3D body poses belonging to 12 minutes period of a night navigation watch

ECDIS, Radio Detection and Ranging (RADAR), Auto Pilot-Helm, Global Maritime Distress and Safety System (GMDSS) console, and navigation table. During Case 2, the lookout still uses the look-out area, however, the OOW largely uses the navigation table. In Case 3, the lookout just stays in the same place and the OOW stays on the ECDIS and sits on the chair. In the final case, the lookout walks in the look-out area, while the OOW uses all of the navigation equipment, except the GMDSS console. These cases show the brief behaviour of WPs for the defined period.

Since the performance of WPs is affected by how they follow their responsibilities, heatmaps provide simple and significant information for understanding what they are doing during their watch. In that respect, motion heatmaps explain the behaviour of WPs, providing suitable input data for training a model on behavioural analysis of WPs.

DISCUSSION ON FINDINGS

Camera calibration results show that pixel coordinates can be converted to world coordinates by using the backward projection algorithm with a maximum 8 mm error (~3.00 pixel). This error can be reduced to a sub-pixel value [33], [34] by using a more accurate calibration object with precise localisation on the bridge. However, acceptance of the error value depends on the task to be performed. The plotting algorithm for motion heatmaps creates a rectangle with the dimensions 250 mm x 500 mm. 8 mm error in both X and Y axes is equal to a 0.05% predicted plotting area. In other words, camera calibration has 99.95% accuracy in motion heatmap generation.

The accuracy of the 3D pose construction depends on the accuracy of both the 2D pose estimation algorithm and camera calibration. The heatmap generation algorithm seeks a combination of neck, shoulders and hip joints and plots ‘neck’ or ‘mid-hip’ on the map. The accuracy of OpenPose on those key joints is ~85% on the MPII human multi-person dataset [35]. Since the camera calibration has too small an error, compared to Openpose, 3D pose construction accuracy is equal to the 2D pose estimation algorithm. There is a trade-off between accuracy and speed, when choosing the right pose estimation algorithm. Collected high-resolution video recordings have 12 frames per second (fps) and there are always multiple persons on the bridge. The pose estimation algorithm should work with at least 24 fps, to estimate multiple poses from two video cameras in real-time. Only the OpenPose running in NVIDIA GeForce GTX-1080 Ti GPU and i7-6850K CPU is satisfying that requirement for this research. Fps can be lowered, to enable more accurate algorithms running real-time, such as AlphaPose [29] or METU [36]. Nevertheless, lower fps means that fewer frames will be used to produce motion heatmaps. This will lead to a non-smooth tracking with pose jumps. So, Openpose is the most appropriate algorithm for the time being, however, if a new pose estimation algorithm which would have more accurate results with more fps is developed in the future, it can be easily adapted to the system developed in this research.

Direct tracking of 3D pose features that present full information about the physical activities of WP can be proposed as an input. Only the uninterrupted, complete 3D pose construction can make this possible. Navigation equipment obstruct constructing complete 3D poses. Rather than tracking completed 3D poses, another input producing method should be suggested. Monitoring the watchkeeping behaviour of WP is mainly based on whether the WP is following the routines of the navigation watch. The location where WPs stand and the time they spend at that location are proof that they are following those routines. While the motion heatmap generation developed in our research works with incomplete 3D pose information, it captures temporal as well as spatial information. The motion heatmaps shown in

Fig. 11, present the behaviour of WPs within the defined time interval. In that respect, heatmaps of body motions is a novel approach to generate input for training a deep learning-based behaviour analysis model.

CONCLUSIONS

IMO recognises that many operational bridge-related marine accidents are caused due to the lack of a system detecting the incapacity of the OOW. IMO makes the BNWAS mandatory for the ships defined in Resolution MSC.282(86) to avert those accidents. However, the defined system is very basic and vulnerable to abuse. There is a need for a more advanced system to mitigate watchkeeping errors and improve the safety of navigation. In this study, BNWMS, which continuously monitors the behaviour of WP autonomously, is suggested to fill this gap. Automated motion heatmap generation is developed to provide input data for BNWMS.

A multi-video camera system was established to obtain data from an actual bridge. A real-time multi-person 2D pose estimation algorithm was run on each camera view to estimate 2D body poses. The backward projection method constructed 3D body poses from binary 2D poses. Although the defined error function in backward projection eliminates wrong 3D poses, a filtering algorithm was developed to validate the results. Validated 3D body pose features were plotted on the 2D bridge plan to generate motion heatmaps of WPs within a defined time interval.

The results show that it is possible to obtain motion heatmaps that give important information about watchkeeping behaviour. Automated motion heatmap generation is a novel approach to produce input data for behaviour analysis of WPs. An expert can make visual evaluations with the heatmaps of certain periods constructed in this study. However, the final goal is to automate the whole process by establishing BNWMS. Therefore, training a deep learning model using motion heatmaps to make the evaluation process by machine forms part of our future work.

This is the first study of vision-based behaviour analysis on a ship's bridge. It is thought that this study, which is the first in its field, will be the basis for a series of other studies. In addition, the approach in this study will pave the way for behaviour analysis in environments other than ships (such as factories) that require working in a large area.

FINANCIAL SUPPORT

Funding: This study was supported by Istanbul Technical University, Scientific Research Projects (BAP), Project ID: MDK-2019-41955.

REFERENCES

1. W. Qiao, Y. Liu, X. Ma, and Y. Liu, "A methodology to evaluate human factors contributed to maritime accident by mapping fuzzy FT into ANN based on HFACS," *Ocean Eng.*, vol. 197, p. 106892, 2020.

2. S. Fan, J. Zhang, E. Blanco-Davis, Z. Yang, and X. Yan, "Maritime accident prevention strategy formulation from a human factor perspective using Bayesian Networks and TOPSIS," *Ocean Eng.*, vol. 210, p. 107544, 2020.

3. K. Kulkarni, F. Goerlandt, J. Li, O. V. Banda, and P. Kujala, "Preventing shipping accidents: Past, present, and future of waterway risk management with Baltic Sea focus," *Saf. Sci.*, vol. 129, p. 104798, 2020.

4. V. Laine, F. Goerlandt, O. V. Banda, M. Baldauf, Y. Koldenhof, and J. Rytkönen, "A risk management framework for maritime Pollution Preparedness and Response: Concepts, processes and tools," *Mar. Pollut. Bull.*, vol. 171, p. 112724, 2021, doi: https://doi.org/10.1016/j.marpolbul.2021.112724.

5. AGCS, "Safety and Shipping Review 2021," Allianz Global Corporate and Speciality, 2021. https://www.agcs.allianz.com/content/dam/onemarketing/agcs/agcs/reports/AGCS-Safety-Shipping-Review-2021.pdf (accessed Sep. 09, 2021).

6. Y. Zhang, X. Sun, J. Chen, and C. Cheng, "Spatial patterns and characteristics of global maritime accidents," Reliab. Eng. *Syst. Saf.*, vol. 206, p. 107310, 2021.

7. K. Liu, Q. Yu, Z. Yuan, Z. Yang, and Y. Shu, "A systematic analysis for maritime accidents causation in Chinese coastal waters using machine learning approaches," *Ocean Coast. Manag.*, vol. 213, p. 105859, 2021.

8. A. Graziano, A. P. Teixeira, and C. G. Soares, "Classification of human errors in grounding and collision accidents using the TRACEr taxonomy," *Saf. Sci.*, vol. 86, pp. 245–257, 2016.

9. IMO, *STCW including 2010 Manila Amendments (ID938E)*, 2017th ed. London: International Maritime Organization, 2017.

10. M. Bull, *Bridge Watchkeeping: A Practical Guide – 3rd Edition.* The Nautical Institute, 2021.

11. M. Kaptan, Ö. Uğurlu, and J. Wang, "The effect of nonconformities encountered in the use of technology on the occurrence of collision, contact and grounding accidents," *Reliab. Eng. Syst. Saf.*, vol. 215, p. 107886, 2021.

12. IMO, *IMO RESOLUTION MSC.128(75), Performance Standards for a Bridge Navigational Watch Alarm System (BNWAS)*, no. May. International Maritime Organization, 2002.

13. H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Syst. Appl.*, 2018.

14. L. Onofri, P. Soda, M. Pechenizkiy, and G. Iannello, "A survey on using domain and contextual knowledge for human activity recognition in video streams," *Expert Syst. Appl.*, vol. 63, pp. 97–111, 2016.

15. S. Bhattacharya and N. D. Lane, "From smart to deep: Robust activity recognition on smartwatches using deep learning," in *Pervasive Computing and Communication Workshops (PerCom Workshops), 2016 IEEE International Conference on*, 2016, pp. 1–6.

16. Y. Jia, X. Song, J. Zhou, L. Liu, L. Nie, and D. S. Rosenblum, "Fusing Social Networks with Deep Learning for Volunteerism Tendency Prediction," in *AAAI*, 2016, pp. 165–171.

17. A. Jalal, Y.-H. Kim, Y.-J. Kim, S. Kamal, and D. Kim, "Robust human activity recognition from depth video using spatiotemporal multi-fused features," *Pattern Recognit.*, vol. 61, pp. 295–308, 2017.

18. Y. Fan, J. C. K. Lam, and V. O. K. Li, "Video-based Emotion Recognition Using Deeply-Supervised Neural Networks," in *Proceedings of the 2018 on International Conference on Multimodal Interaction*, 2018, pp. 584–588.

19. N. Neverova, C. Wolf, G. W. Taylor, and F. Nebout, "Multi-scale deep learning for gesture detection and localization," in *Workshop at the European conference on computer vision*, 2014, pp. 474–490.

20. W. Zhang, Y. L. Murphey, T. Wang, and Q. Xu, "Driver yawning detection based on deep convolutional neural learning and robust nose tracking," in *Neural Networks (IJCNN), 2015 International Joint Conference on*, 2015, pp. 1–8.

21. Y.-J. Han, W. Kim, and J.-S. Park, "Efficient Eye-Blinking Detection on Smartphones: A Hybrid Approach Based on Deep Learning," *Mob. Inf. Syst.*, vol. 2018, 2018.

22. Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, 2019.

23. D. Wu, N. Sharma, and M. Blumenstein, "Recent advances in video-based human action recognition using deep learning: a review," in *Neural Networks (IJCNN), 2017 International Joint Conference on*, 2017, pp. 2865–2872.

24. R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE J. Robot. Autom.*, vol. 3, no. 4, pp. 323–344, 1987.

25. P. F. Sturm and S. J. Maybank, "On plane-based camera calibration: A general algorithm, singularities, applications," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*, 1999, vol. 1, pp. 432–437.

26. R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

27. J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, 1997, pp. 1106–1112.

28. J. Heikkila, "Geometric camera calibration using circular control points," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1066–1077, 2000.

29. H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "Rmpe: Regional multi-person pose estimation," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2334–2343.

30. K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

31. J. Wang et al., "Deep high-resolution representation learning for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020.

32. L. Pishchulin et al., "Deepcut: Joint subset partition and labeling for multi person pose estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4929–4937.

33. H. Hirschmuller and S. Gehrig, "Stereo matching in the presence of sub-pixel calibration errors," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 437–444.

34. T. Yang, Q. Zhao, X. Wang, and Q. Zhou, "Sub-Pixel Chessboard Corner Localization for Camera Calibration and Pose Estimation," *Applied Sciences* , vol. 8, no. 11. 2018, doi: 10.3390/app8112118.

35. M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2d human pose estimation: New benchmark and state of the art analysis," in *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, 2014, pp. 3686–3693.

36. M. Kocabas, S. Karagoz, and E. Akbas, "Multiposenet: Fast multi-person pose estimation using pose residual network," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 417–433.

CONTACT WITH THE AUTHORS

Veysel Gokcek
e-mail: gokcekv@itu.edu.tr

Gazi Kocak
e-mail: kocakga@itu.edu.tr

Istanbul Technical University
Tuzla, 34940 Istanbul
Turkey

Yakup Genc
e-mail: kocakga@itu.edu.tr

Gebze Technical University
Gebze, 41400 Kocaeli
Turkey