Visual Attention Assessment for Expert-in-the-loop Training in a Maritime Operation Simulator

Guoyuan Li, *Member, IEEE,* Runze Mao, Hans Petter Hildre, and Houxiang Zhang, *Senior Member, IEEE*

***Abstract*—Improving the training programs for maritime operations is beneficial to enhance maritime safety in practice. In this paper, we propose a novel approach to the assessment of visual attention in a maritime operation so as to support an expert-in-the-loop training program. Experts' knowledge of maritime operation and experiences in the simulator are incorporated into the training program in three ways. First, through a questionnaire, information about task division, identification of critical operation, and definitions of areas of interest (AOIs) is incorporated as prior knowledge for modeling visual attention. Second, a weight scale factor that emphasizes the high importance of visual focus in critical operations is utilized to generate an operations-dependent attention map. Third, based on an expert's attention map and visual switch between AOIs, a similarity metric is designed as a comprehensive evaluation between saliency and visual transition. A case study of heavy lifting operation is performed by two groups of trainees who have received different briefings about "critical operation". The assessment result shows the group with more detailed briefing obtains a 6% similarity score higher than the other group, which is consistent with the debriefing result of a superior performance of that group. The proposed approach is thus verified effective in assessing visual attention for the expert-in-the-loop training program.**

***Index Terms*—Area of interest, Eye tracking, Saliency map, Visual attention, Maritime operation.**

I. Introduction

RECENT years have seen an increase in maritime activities in challenging oceanic regions [1]. Undertaking demanding maritime operations, such as offshore petroleum extraction, wind turbine installation, and subsea pipeline deployment in the complicated environment, requires substantial experience and skills. The increasing complexity of maritime operations has led to a growing risk of maritime accidents. Human error is reported the prevailing factor that accounts for 80% of those occurring worldwide [2]. Significant investments in training have been made for maritime safety [3]. It is expected that effective training in a simulated scenario, such as in a simulator, could provide experience and reduce human error. Identifying the problems that lead to human error via visual attention in a training program is vital to accident prevention. Thus understanding trainees' visual focus with greater accuracy in a simulator is necessary.

With the development of sensing technology, eye tracking has been applied as a training tool in various domains, including surgery, sports, driving, and aircraft inspection [4]–[8]. Gaze points are tracked by means of high-speed cameras either integrated into wearable glasses or mounted on desktop systems. These tools generally identify two primary eye movements: fixation and saccades [9]. Based on the two types of eye movements, various eye-tracking metrics such as fixation frequency and dwell time, hit count in AOIs, attention map, and scanpath can be calculated to build meaningful representations for visual attention analysis [10]. Considering these features, it seems likely that applying eye trackers to maritime operation training will not only offer a means of perceiving operator's situation awareness, but also provide a possible solution to improve the training program in terms of maritime safety.

Two trends have emerged in research on visual attention. On the one hand, researchers from psychology have focused on investigating the correlation between behavior and visual attention, such as analyzing search patterns of workers for hazard identification[11] and comparing visual focus in flight simulators and in-flight [12]. Researchers who study computer vision, on the other hand, have made efforts to model visual attention, such as saliency maps and scanpath [13], [14], as well as visualization [15].

Both steams of research are crucial to training using a maritime operation simulator. We can improve the training program from two aspects. First, experts and novices use different search strategies during operations [11], [16]. The training program should focus on incorporating expertise so that trainees can learn what to focus on during demanding operations and gain hazard awareness. Second, modeling experts' visual focus in a simulator could facilitate the assessment of trainees' visual attention. However, given the different operating habits among experts in operation, especially when performing a critical operation, it is difficult to use ready-made models (see [13] and references therein) directly.

In this paper, we propose a novel approach that leverages both experts' expertise and experiences in maritime operation to assess trainees' visual attention in a simulator. The contributions of this paper include:

- A novel framework that brings experts into the loop in the whole process of creating the training procedure, from briefing, to training, to debriefing;
- Design of a weight scale factor to prioritize importance of visual focus in critical operations;

The research is supported partially by a grant from the Knowledge-Building Project for Industry "Digital Twins for Vessel Life Cycle Service" (Project nr: 280703), and partially by a grant from the Research-based Innovation "SFI Marine Operation in Virtual Environment (SFI-MOVE)" (Project nr: 237929) in Norway.

The authors are with the Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology, Ålesund, 6025, Norway (Corresponding author: Guoyuan Li). Emails: guoyuan.li@ntnu.no, maor@stud.ntnu.no, hans.p.hildre@ntnu.no, hozh@ntnu.no.

Manuscript received Month xx, 2xxx; revised Month xx, xxxx; accepted Month x, xxxx.

- A similarity metric based on the generation of experts' attention models to assess trainee's visual attention.

II. Related Work

A. Visual Attention Comparison between Expert and Novice

Experience and expertise are considered the main factors that separate experts' and novices' visual attention. Attempts have been made to investigate the visual focus in simulators among operators with different level of skills, as well as different operating abilities [11], [16], [17]. In most of the comparative studies, visual attention is assessed by comparing fixation related data, such as fixation location, number, duration and percentage in specific AOIs. There is also research work that makes use of scanpath patterns to distinguish experts from novices [18]. These studies suggest that experienced operators can perform more efficient fixation in operation. For example, increased expertise requires less fixation to identify pathology in a surgical assessment [4]; experienced drivers perform a wider horizontal scan for safety [17]; and experienced workers exhibit fewer fixations when identifying hazards [11]. Based on this, the project outlined here prioritizes experts' experience and expertise to design a training program for maritime operation.

B. Saliency-Based Similarity Metrics

Assessment of visual attention by means of saliency-based metrics has been a hot topic in the past few decades [13], [14], [19]. Dozens of similarity metrics have been developed to evaluate saliency-based attention. Receiver operating characteristic (ROC) analysis is one of the most popular methods [20]. A constant threshold for the salient level of the ground truth map and a variable threshold for the predicted map are used. The ROC method compares each pixel's saliency value in the map with the variable threshold and classifies it as either fixated or non-fixated. An ROC curve depicting the relationship between the false positive rate and true positive rate of the pixels can be drawn; the area under the curve represents the similarity between the two saliency maps. The linear correlation coefficient (CC) is another metric for comparison between the two saliency maps [21]. It is defined as the ratio between the covariance of the two saliency maps and the product of their standard deviations. An absolute value of CC close to 1 indicates a perfect correlation.

It is possible to use a probability distribution function to compare saliency similarity. In general, the two saliency maps need to be normalized as two corresponding probability distribution functions. Judd et al. designed a similarity metric that sums the minimum value of the two functions at each pixel of the map [22]. The similarity score has an upper bound of 1, which corresponds to identical distribution of the two saliency maps. In addition, some other similarity metrics using the probability distribution function like the Kullback-Leibler divergence metric and the earth movers distance metric have been applied to compare saliency maps [23], [24].

Instead of directly comparing two saliency maps, there are methods that evaluate eye fixation by comparing their corresponding saliency values on a benchmark saliency map.

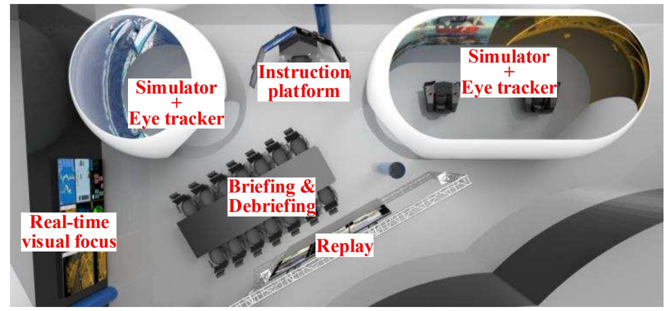

Fig. 1. Layout of training center for maritime operations.

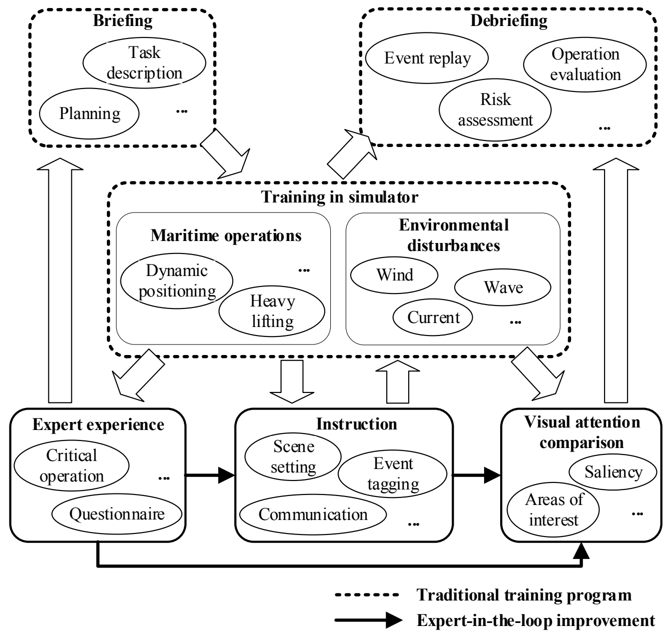

Fig. 2. Integration of experts' knowledge and their experiences in simulator into maritime operation training.

For example, the normalized scanpath saliency (NSS) method standardizes the saliency map with a zero mean and unit standard deviation and averages the saliency values of all the tested fixation locations [25]. An NSS value close to 1 represents a good correspondence between the fixation and the saliency map, whereas a negative NSS value indicates an opposite correspondence. Peters and Itti proposed a percentile metric to assess eye fixation on a saliency map [26]. For each fixation, it counts the number of pixels that have saliency values smaller than that of the fixation location and normalized the number by dividing the total number of pixels. The final score is the average of the normalized values of all the fixations.

The aforementioned similarity metrics are more focused on spatial comparison of visual attention. Considering the importance of temporal information in a critical maritime operation, a metric that accounts for spatial-temporal similarity is needed. We will present such a metric in Section IV.

III. Design of Expert-in-the-loop Training Program for Maritime Operations

Accurate understanding of trainees' behavior in maritime operations especially in terms of situational awareness, is the most important design criteria for the training program. However, illustrative techniques to visualize situational awareness in simulators are lacking until the advent of a new generation of eye trackers. Taking into account the ability of real-time tracking of visual focus, as well as the ability for post-analysis on an attention map, eye trackers are considered a suitable instrument to be integrated into the simulator.

Fig. 1 illustrates the arrangement of facilities that are used in our maritime operation training program. An open meeting mode is established that all related personnel can sit aside the simulators for briefing and debriefing. The behavior of the trainee who wears the eye tracker and operates in the simulator can be visualized on the video wall and replayed if necessary. The instruction platform plays a role in governing the training process, from which the instructor is able to adjust the simulation situation and record critical events.

Inspired by the work in [27], an improved brief-training-debriefing paradigm that integrates experience and expertise is proposed, as shown in Fig. 2. The following introduces the three phases with emphasis on how to apply experts' experience and their visual attentions to this procedure.

- *Briefing*: In this phase, trainees are provided with instructions, goals and rules about the maritime operation. Apart from these information, some of experts' assessments, such as the difficulty of the operation and where the visual focus should be, are summarized via questionnaire. The assessment may be based on either past work experience or from experience in simulators. The added information will enable the trainees to engage in the operation actively and effectively.
- *Training*: The training takes place in the left spherical dome, as shown in Fig. 1. A wearable eye tracker is utilized together with an overhead camera to monitor the user's behavior. The type of maritime operations and the simulation environment including wind, wave and sea current, are set through the instruction platform. In addition, operational expertise such as how to identify critical operations is applied to record corresponding events for analysis in the debriefing phase. Operational samples performed by experts in the simulator are collected prior to the training for comparison.
- *Debriefing*: As the core element of simulation-based training, debriefing aims to help trainees to explore and understand relationships among actions and events, and grasp operating insights. To better comprehend trainees' behavior in the course of operation, a comparison of visual attention between experts and trainees is conducted, and the result is used as one of the evaluation indicator for the operation.

Fig. 3 depicts the roles of experts and trainees, respectively, in assessment of visual attention. For each specific operation, a questionnaire is given to experts. They are asked to split the task into different stages based on operational inherent

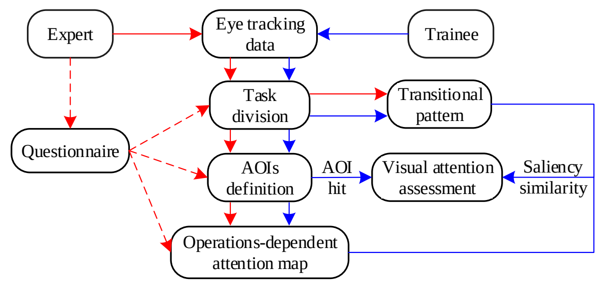

Fig. 3. Diagram of modeling and assessment of visual attention for the training program.

sequence [28]. The purpose is to simplify the comparison of visual attention stage by stage. The time period of critical operations in each sub-task is annotated, indicating more effort should be taken when entering or exiting it. This is closely related to the trigger of events during training, and thus can be used as a measurement in the time domain for the generation of an operations-dependent attention map. AOIs are another concern in the questionnaire. The unified AOIs will be used to extract the temporal characteristics of an expert's fixations, i.e., the transitional pattern.

When a trainee finishes the operation, his/her eye-tracking data is divided into segments that correspond to the stages defined by experts. Each segment of data will be used to evaluated the trainee's visual attention from two aspects. The first one is AOI hit, that is, the number of fixations within the AOIs. Here we neglect the visit time, length and angle among AOIs but focus on hit rate, as it is intuitive that a higher hit rate implies a potential good practice. The second evaluation target, on the premise of high hit rate, is saliency similarity. A similarity metric considering both temporal and spatial characteristics of the trainee and the expert is proposed. More details are introduced in Section IV.

IV. Operations-dependent Visual Attention Assessment

In this section, a pair-wise method is proposed to compare visual attention between experts and trainees. We use a ship maneuvering example to illustrate how to use a weight scale factor to generate attention map, and how to evaluate saliency similarity from both spatial and temporal perspectives.

A. Eye-tracking Data Modeling

The collected eye-tracking data contains various information, such as timestamp, eye movement type, and gaze position. For each type of maritime operation, an operational scene image I with dimensions $m \times n$ is applied to post analysis as a static visual stimulus used for uniformed comparison model establishment. Saccade and unknown type of eye movement data are filtered out as there are no corresponding gaze points mapped on the image. Only fixation data is remained and mapped to the local coordinate system of I.

A maritime operation due to its inherent sequence can be expressed in stages as $O := (o_1, o_2 ..., o_K)$, where K is the stage number. Instead of evaluating the entire operation

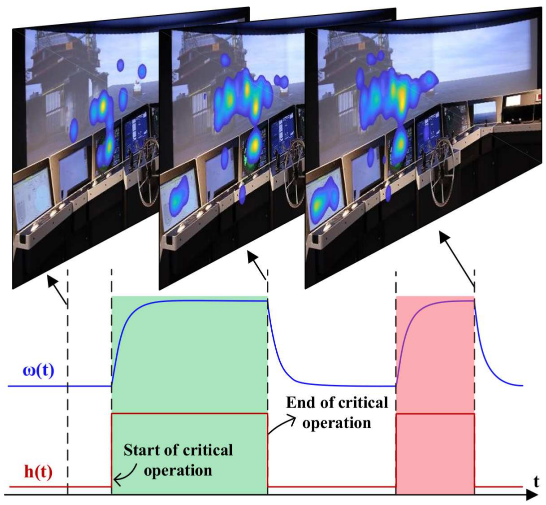

Fig. 4. An example of attention map generation. Green region: critical operation for approaching the rig. Red region: critical operation for orientation keeping in DP mode.

directly, the following assessment is designed for each stage of operation o_k. It is a kind of pair-wise comparison that the eye-tracking data from an expert e_j will be taken as a reference to evaluate the eye-tracking data from the trainee. Once the degree of saliency similarity $ss(e_j, o_k)$ in o_k based on e_j is computed, the overall evaluation can be obtained by:

$$ss = \frac{1}{K}\sum_{k=1}^{K}\max_{e_j \in E} ss(e_j, o_k) \tag{1}$$

where E is the expert set. Higher value of ss indicates more similarity of visual attention compared to that of experts and thus a better operational performance by the trainee.

B. Operations-dependent Attention Map

For the reference expert e_j in the stage of operation o_k, the eye movement is modeled as a sequence of gaze points $P := (p_1, p_2 ..., p_N)$ along the time line $T := (t_1, t_2 ..., t_N)$, where N is the number of gaze samples. Each gaze point p_i is within the image dimensions $m \times n$. Because the eye tracker has a dynamic sampling rate up to 50 HZ, the interval $\Delta T := (\Delta t_1, \Delta t_2, ...\Delta t_{N-1})$ is not a constant vector.

As mentioned in Section III, critical operation accounts for higher attention during the stage of operation. Here we propose a weight metric that classifies the temporal ordering of gaze points into two different levels. Given the event trigger time t_{in} and t_{out}, representing the operating time when starting and ending the critical operation, respectively, an event stimulus function $h(\cdot)$ is introduced:

$$h(t_i) = \begin{cases} 1 & \text{if } t_{in} \le t_i \le t_{out} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

To avoid abrupt change on the weights of the gaze points, a leaky integrator is utilized [29]:

$$\Delta\omega(t_i) = \{-\frac{\omega(t_{i-1})}{\tau} + rh(t_{i-1})\}\Delta t_{i-1} \tag{3}$$

where τ is the leaky rate and r denotes the scale factor of weight.

A spatial attention density for the reference expert is designed as the weighted mean of an overlay of his/her gaze positions on the static stimuli I:

$$\rho = \frac{\sum_{i=1}^{N}\omega(t_i)g(p_i)}{\sum_{j=1}^{N}\omega(t_j)} \tag{4}$$

Here $g(\cdot)$ is the bidimensional Gaussian function:

$$g(p_i) = \frac{1}{2\pi\sigma^2}e^{-\frac{\|p-p_i\|^2}{2\sigma^2}} \tag{5}$$

where p is the spatial coordinates of image I; σ is the Gaussian standard deviation, representing how wide the gaze points are affected over the image. In eye tracking community, σ is commonly accepted to be set in pixels corresponding to 1° of visual angle [14]. Therefore, σ is determined by the experimental setup, such as the viewing distance and the screen size.

An example of how to generate the operations-dependent attention map is shown in Fig. 4. It is a ship maneuvering task. The operator is asked to first steer the ship toward the rig and then start dynamic positioning (DP) when the ship is close enough to the rig. Since there is another vessel which is also approaching the rig during the operation, two critical operations are identified. The first one is the close-range maneuver to avoid collision; the other one is the ship orientation in DP mode. They are represented in the timeline in Fig. 4 as a green region and a red region, respectively. There is a clear shift of attention at the time of the completion of the two critical operations. This is consistent with Eq. (4), as the gaze points in these two time periods have more effect in generating the attention map.

C. Saliency Similarity Assessment

The proposed saliency comparison method is akin to the hybrid method in [14] but attempts to make a comprehensive evaluation regarding how similar the transitional patterns and the saliency in AOIs between the expert and the trainee. Suppose there are R AOIs defined by experts in the stage of operation, $A := (A_1, A_2, ..., A_R)$. The AOIs are treated as polygons and thus it is simple to extract an AOIs mask. The combination of the AOIs and the generated attention map enables a visualization of AOI hit with its attention density in time domain. On the other hand, the attention density in AOIs can be processed with different thresholds, forming a different density level of salient areas (SAs). As a result, the degree of saliency similarity can be evaluated in each spatial-temporal block. Fig. 5 is an example of the evaluation process for a trainee in the ship maneuvering task. Note the trainee has his/her own weight profile. The degree of similarity varies with AOI hit number, switch between AOIs, and distribution in SAs. The following describes the similarity metric for the whole stage of operation. The similarity computation in each spatial-temporal block can be deduced accordingly.

The switch between AOIs is a valid indicator used for scan-path comparison [30]. Therefore, it is used here as a part of the

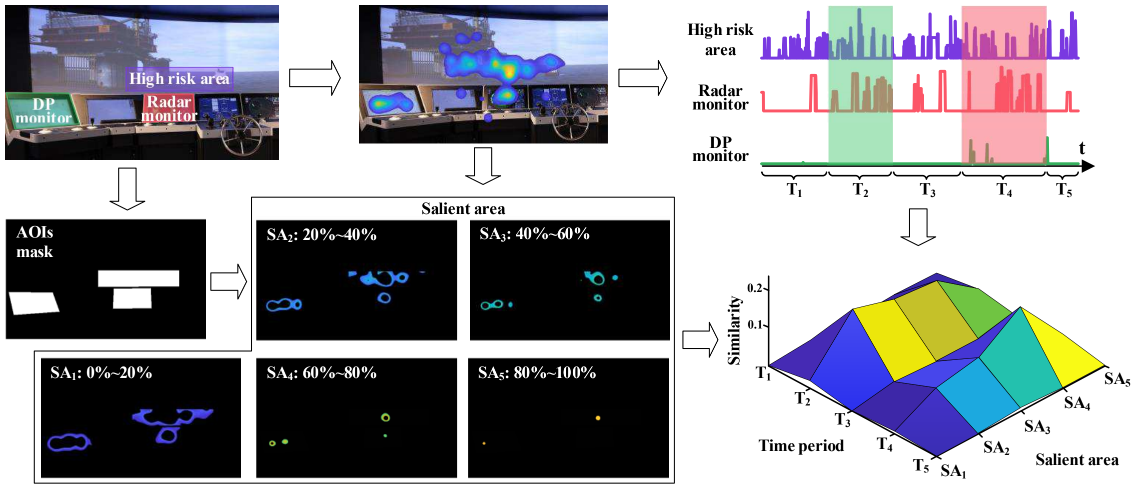

Fig. 5. An example of assessing saliency similarity by considering both temporal and spatial characteristics.

comparison metric. The transitional model β for the reference expert has a $(R+1)\times(R+1)$ tabular representation. The extra dimension is due to the gaze points falling outside the AOIs. Each element $\beta(i,j)$ is computed by accumulating the number of transitions from A_i to A_j and then regularized by dividing by the total number of visits. Assume the subscripts a and b denote the expert and the trainee, respectively. Given the transitional model β_a and ρ_a from the reference expert, and the gaze point sequence $Q := (q_0, q_1, q_2 ..., q_{N'})$ (q_0 is the duplication of q_1) and weight profile ω_b from the trainee, the similarity score for the trainee is designed as the weighted mean of convolution of β_a and ρ_a:

$$f_b = \frac{\sum_{i=1}^{i=N'} \omega_b(t_i)\, \beta_a(\gamma(q_{i-1}), \gamma(q_i))\, \rho_a(q_i)}{\sum_{j=1}^{j=N'} \omega_b(t_j)} \tag{6}$$

where $\gamma(\cdot)$ is the function to convert the coordinate point into the index of AOI where it falls. The metric can be interpreted from two aspects. First, from a spatial viewpoint, it evaluates how close the trainee's gaze points are compared to those of the expert by projecting them onto the expert's attention model, i.e., β_a and ρ_a. Second, from a temporal perspective, ω_b is closely coupled with the time period of completion of critical operations by the trainee, which results in a time-dependent metric. In order to eliminate the dependency, the metric is divided by an accumulated weight which corresponds to the operation time by the trainee.

Considering f_b is an absolute measurement of gaze points of the trainee applied to the expert's attention model, it has no explicit upper bound [14]. A feasible improvement is to make another measurement f_a to evaluate the expert's gaze points on his/her own attention model. This can be achieved by substituting gaze points Q with P, and ω_b with ω_a in Eq. 6. In this way, we can provide the saliency similarity metric as a relative measurement for the stage of operation (recall the denotation of the expert e_j and the stage of operation o_k in Section IV-A):

$$ss(e_j, o_k) = \frac{f_b}{f_a} \tag{7}$$

The metric has a theoretical upper bound of 1. An ss value close to 1 would indicate a good performance of visual attention by the trainee in the maritime operation.

V. Experiment

A. Heavy Lifting Operation Test

A case study of a heavy lifting operation was conducted in our maritime operation simulator (see the deployment in Fig. 1). Fig. 6a shows the operation scene for a crane driver in the simulation dome. The task, as illustrated in Fig. 6b, is to use a 250 tons knuckle boom crane to lift a 80 tons suction anchor with a height of 20 m and a diameter of 5.3 m from the deck of an offshore construction vessel to a 100 m deep seabed, and reverse the operation until the anchor is placed and secured in the fastening structure on the deck. We changed the environmental disturbance by gradually increasing the significant wave height from 0.5 m to 2 m during the operation. The operating challenge lies in the considerable swing of the anchor and the difficulty for the operator to set it back on deck.

There were two experts with about ten years of operational experience, and ten trainee participants who have four to six years of operational experience in the experiment. The trainees are considered having the same level of ability to accept, understand and utilize the information for training. They were divided into three groups: expert group $E = \{e_1, e_2\}$, group one $G_1 = \{s_1, s_2, ...s_5\}$ and group two $G_2 = \{s_6, s_7, ...s_{10}\}$. The difference between G_1 and G_2 is that in the briefing phase, trainees in G_1 obtained detailed information about "critical operations", such as the AOIs in Fig. 7, the risks in operation, and the visual focus to ensure safety; whereas for trainees in G_2, they were only verbally told the potential risks for the operation. The purpose is to test how the information would affect their visual attention.

Based on experts' survey responses, the task is divided into four stages $O = \{o_1, o_2, o_3, o_4\}$, as shown in Fig. 6c. In stage 1, the anchor is lifted and swung out over the side of the ship. Stage 2 includes descending the anchor until it is completely

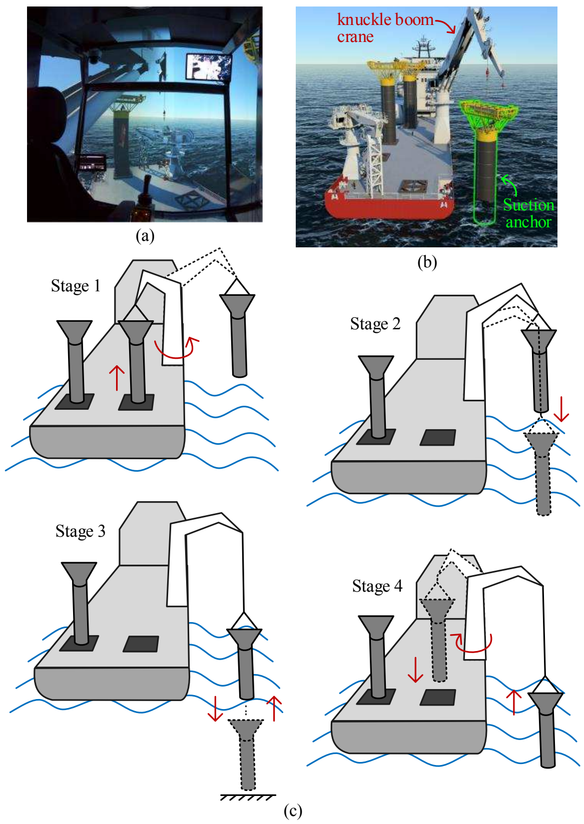

Fig. 6. Operation scene in dome (a), overview of the operation (b), and task division (c) for the heavy lifting operation experiment.

TABLE I
AOIS AND CRITICAL OPERATIONS IN THE HEAVY LIFTING OPERATION.

Stage	AOIs	Critical operation
o_1	$A_1 \sim A_4, A_6$	Initial lift of the anchor
o_2	A_1, A_3, A_5	Stabilizing swing when descending
o_3	$A_1 \sim A_3$	Slow speed when landing to seabed
o_4	$A_1 \sim A_4, A_6$	Stabilizing to place on deck

sunk. In stage 3, the anchor continues to descend to the seabed and then is lifted up to water surface, where it is followed by stage 4, i.e., stabilizing the anchor and placing it on the deck. In addition, Fig. 7 illustrates the AOIs defined by the experts. The AOIs in different stages, in conjunction with the corresponding critical operation, are depicted in Table I.

Data was collected by our developed software [31] and the gaze points were mapped to the image shown in Fig. 7 with a pixel size of 1210×1047. Table II lists the operating result for the three groups. Note that here "AOI hit rate" is the ratio of AOI hit between the critical operation and the entire operation. There is no significant difference for the entire operation time of trainees in G_1 and G_2. This reflects that the operation time is not an efficient indicator for the assessment. Trainees in G_2 used more time to cope with the critical operation; however, longer operation time does not increase their AOI hit rates. This implies that in G_2 there are a few improper visual focuses appeared in the critical operation. The following sections will analyze the visual attention of participants from the three groups for the heavy lifting operation.

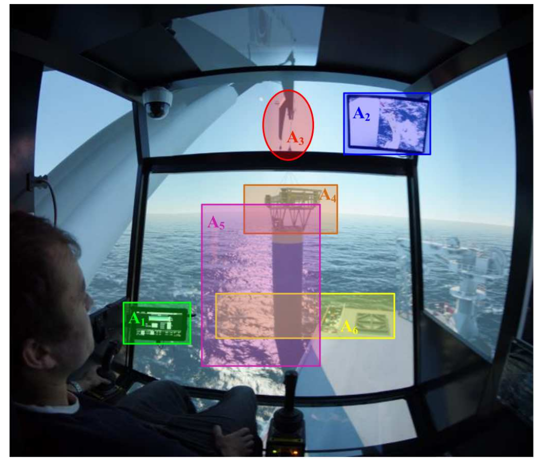

Fig. 7. Defined AOIs by experts.

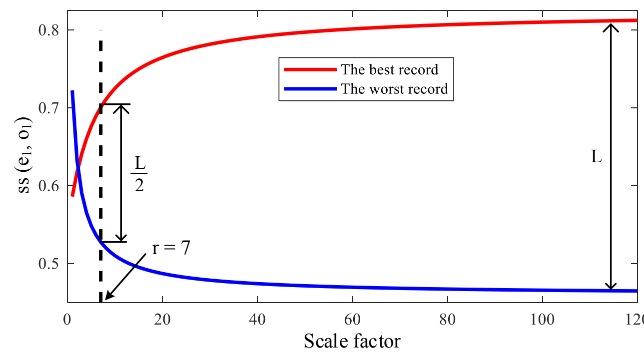

Fig. 8. Selection of weight scale factor for e_1 in o_1.

B. Generation of Experts' Attention Maps

From Section IV-B, experts' attention maps are associated with the leaky rate τ, the Gaussian standard deviation ρ and the scale factor of weight r. In this case study, τ is set to 1 and ρ is set to 20 pixels according to the visual angle at a distance of 1.5 m. Setting of r has a great influence on ss. A smaller weighting factor does not reflect the importance of visual attention of critical operations; but an excessive weighting factor leads to the convergence of ss, indicating an over-reliance on the visual attention of critical operations.

Here an empirical method is applied to determine r for each stage of operation. For trainees in o_k, the variation of $ss(e_j, o_k)$ with the growth of r is observed. Two records with the largest difference of convergence value are selected. Suppose the difference of convergence value is L. We increase r until the similarity difference of the two records equals half of L. Fig. 8 is an example of selecting r for e_1 in o_1. Note if there exists an intersection between the two records, the selected r should be after the intersection; otherwise, r should be chosen to make the difference of the two records equal to or close to 50%.

Fig. 9 illustrates the snapshots of the heavy lifting operation from the experts and the trainees. According to Eq. (2)-(5), experts' attention maps are generated, as shown in Fig. 10, together with the scale factors for the four stages of the operation. In o_1, both experts paid attention to the crane tip, the

TABLE II
STATISTICS OF OPERATIONAL RESULT FOR EXPERTS AND TRAINEES (M $\pm$ SD).

Stage	E			G_1			G_2		
	Total op. time [s]	Critical op. time [s]	AOI hit rate [%]	Total op. time [s]	Critical op. time [s]	AOI hit rate [%]	Total op. time [s]	Critical op. time [s]	AOI hit rate [%]
o_1	70.0 ± 14.1	21.5 ± 4.9	37.3 ± 6.6	76.6 ± 10.1	24.8 ± 7.8	36.5 ± 6.8	84.0 ± 12.1	29.6 ± 11.8	32.2 ± 7.4
o_2	108.0 ± 17.3	22.0 ± 2.8	25.2 ± 2.4	129.0 ± 26.2	26.4 ± 6.7	19.0 ± 3.2	117.4 ± 33.5	31.1 ± 8.5	16.6 ± 3.9
o_3	968.0 ± 135.8	19.5 ± 0.7	2.8 ± 0.3	1142.8 ± 55.0	20.8 ± 3.0	2.1 ± 0.6	1070.4 ± 83.3	21.7 ± 5.2	2.0 ± 1.1
o_4	429.5 ± 70.0	178.0 ± 9.9	53.6 ± 11.9	501.2 ± 86.7	188.0 ± 16.6	46.1 ± 7.3	522.4 ± 63.5	212.1 ± 19.5	44.5 ± 12.7

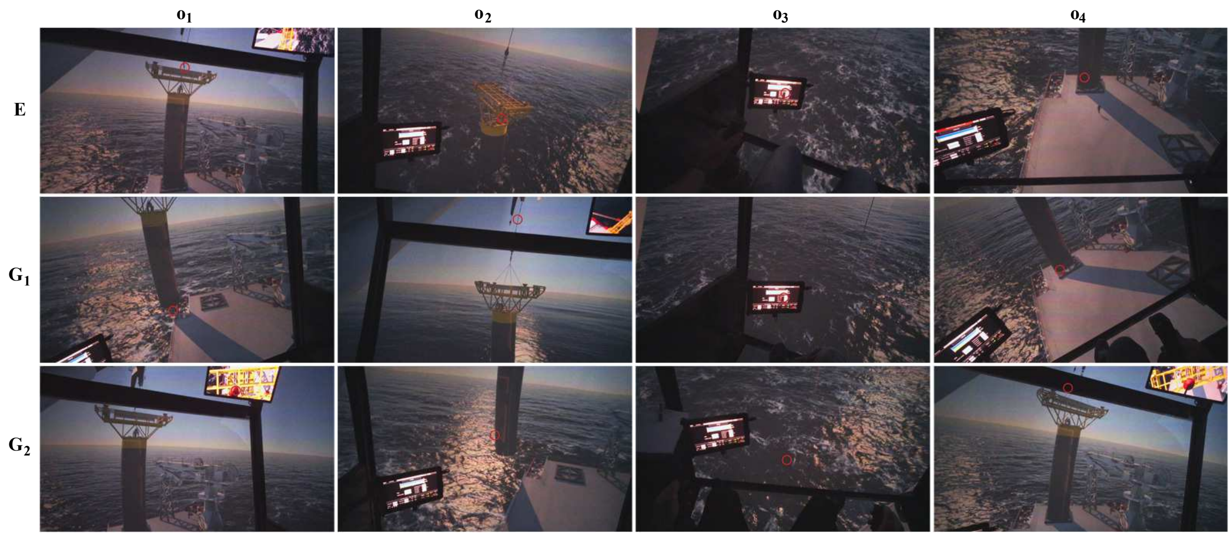

Fig. 9. Snapshots of the four stages of operation from participants in the three groups.

monitor in the lower left corner and the swing of the anchor. There is a small visual attention difference in o_2 in that one expert cared more about the crane tip while the other focused more on the immersion of the anchor. The monitor in the lower left corner was the main concern for both experts in o_3. Only a few fixations were put on the crane tip by one of the experts. Due to environmental disturbance, visual attention was more focused on placing the anchor in the fastening structure on deck in o_4. The result reveals that experts' attention maps are almost consistent with the AOIs they defined in Fig. 7.

C. Saliency Similarity Comparison

We compared saliency similarity of G_1 and G_2 in e_1's and e_2's attention model for the four stages of operation respectively, and averaged the degree of similarity by Eq. (1). The mean and the standard derivation (SD) of ss for G_1 and G_2 are illustrated in Fig. 11.

It is noted that for both G_1 and G_2, the mean of ss in e_2's model is higher than that of ss in e_1's model in o_1. The difference is due to the similarity score collected in A_6. In particular, when the AOI hits are located near the fastening structure on deck, fewer similarity scores will be collected in e_1's model by comparing the attention maps of o_1 shown in Fig. 10. The same situation occurred to A_5 in o_2, which indicates trainees preferred to focus on the swinging anchor in A_5, paying less attention to A_3, i.e., the crane tip.

In o_3, both G_1 and G_2 obtained the highest mean of ss among the four stages of operation. This is because this stage of operation is relative simple, even though it accounts for 60% of the entire operation time. The visual focus has no apparent difference between critical and noncritical operation, as the lower left monitor is the most noteworthy area for receiving underwater sensor information. Nevertheless, note that higher mean of ss is obtained in e_1's model. The difference stems from the difference of attention maps in A_3 between the two experts.

The most difficult part of the task is in o_4, in which up to 40% of the time was used for critical operation. A higher mean of ss in e_2's model, which is similar to the case in o_1 and o_2, is observed. This is attributed to higher AOI hits in A_6 than in A_1 or A_3, as shown in Fig. 10.

In addition, the difference of ss between G_1 and G_2 is obvious in Fig. 11. Trainees in G_1 obtained a higher mean of ss in each stage of operation and therefore achieved a better result. By contrast, trainees in G_2 obtained a higher SD of ss, which implies they did not know what to focus on and hence more free viewing was performed during operation. In fact, Table II already roughly reveals the inferior performance of visual focus, i.e., they obtained a lower AOI hit rate although they used more time in critical operation. Nevertheless, the quantified saliency similarity metric ss in Fig. 11 provides a more comprehensive comparison rather than AOI hit rate.

Besides the overall comparison in Fig. 11, it is also interesting to gain insight into individual visual attention if necessary, e.g., to investigate why the trainee in G_2 obtained the lowest ss in e_1's model in o_2. Fig. 12 depicts the AOI hit with attention value for this record. The trainee used about 29s for critical

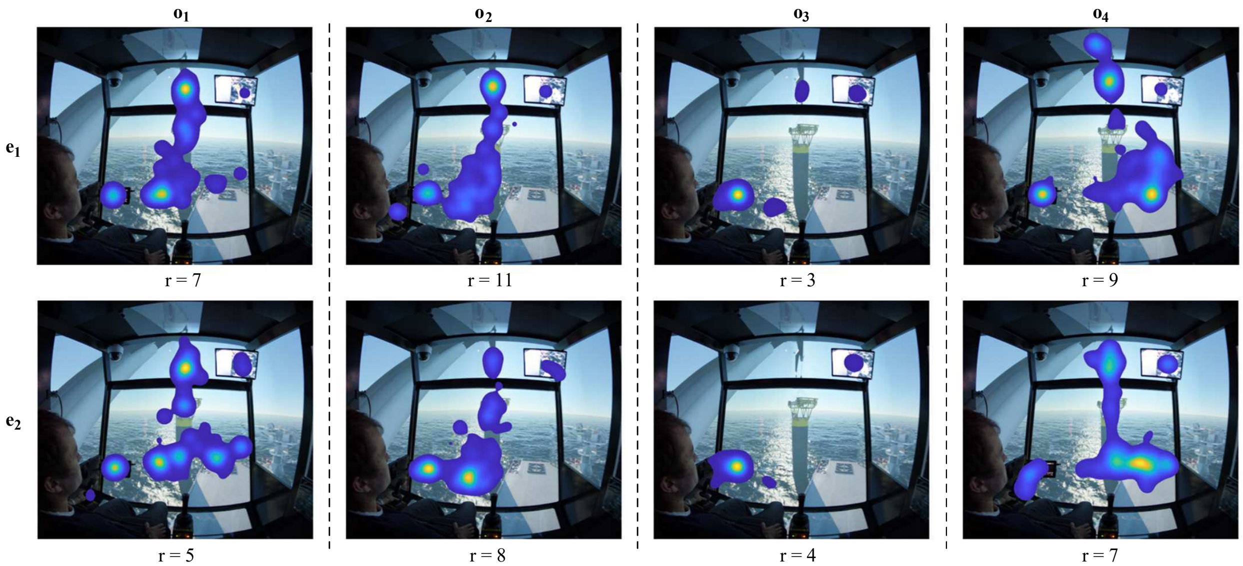

Fig. 10. Experts' attention maps in the four stages of operation.

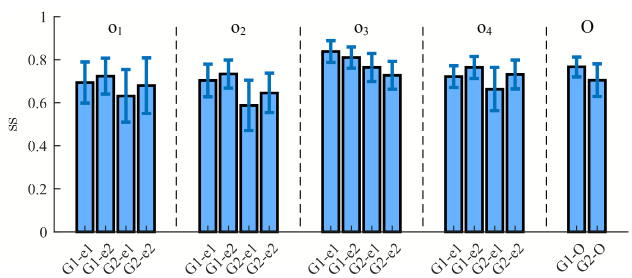

Fig. 11. Statistical result of saliency similarity For G_1 and G_2.

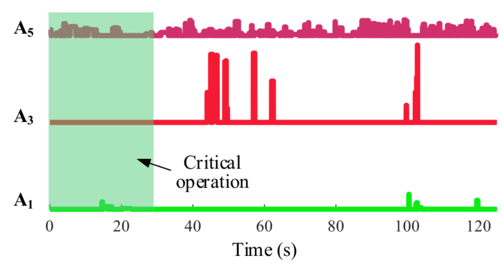

Fig. 12. AIO hit of the record with the lowest ss between G_2–e_1 in o_2.

operation. There are plenty of AOI hits in A_5 but sparse AOI hits in A_1 and A_3. From Fig. 7 and e_1's attention map in o_2 shown in Fig. 10, A_5 is full of low attention density, and A_3 contains a relatively high attention density. This indicates AOI hits in A_5 have limited contribution to ss. Table III sums up the AOI hit rate and ss values of critical/noncritical operation in five different SAs for this record. It is observed that 78% of AOI hits are located in 0∼20% of SAs in noncritical operation but they contributes only about 30% of ss. Due to the weight scale factor, AOI hits in this SA in critical operation (most are located in A_5) contributes almost all of the rest of ss. The key reason of low ss for this record is the lack of AOI hits in high value of SAs in critical operation, e.g., as Fig. 12 suggests, the trainee did not pay attention to A_3 at all during the critical operation.

TABLE III
DISTRIBUTION OF AOI HIT RATE AND ss FOR THE SAMPLE WITH LOWEST ss BETWEEN G_2–e_1 IN o_2.

SAs	AOI hit rate		ss	
	Critical op.	Noncritical op.	Critical op.	Noncritical op.
0 ∼ 20%	0.175	0.780	0.260	0.129
20% ∼ 40%	0.002	0.018	0.008	0.009
40% ∼ 60%	0.000	0.003	0.000	0.002
60% ∼ 80%	0.000	0.001	0.000	0.001
80% ∼ 100%	0.000	0.021	0.000	0.012

D. Discussion

Although performing free viewing in the operation may obtain a high ss (e.g., the trainee in G_2 with the highest overall ss of 0.78), the case study reveals that emphasizing "critical operation" in a briefing phase can attract more visual focus into the proper AOIs and thus increase the overall ss value (e.g., trainees in G_1 and G_2 with an overall ss of 0.77 and 0.71, respectively). This is consistent with the conclusion in debriefing that trainees in G_1 performed better than trainees in G_2. Note the visual attention assessment here is not the final result for the operation. As illustrated in Fig 2, it could be used in the debriefing phase as one of the indicators for comprehensive evaluation.

The case study verifies the effectiveness of the visual attention assessment in the expert-in-the-loop training framework. However, there are some subjective factors in the assessment procedure that may affect the results. For example, determining the trigger time of critical operation (t_{in} and t_{out} in Eq. (2)) is crucial to ss but difficult to identify precisely. Efforts can be made towards the refinement of the questionnaire, as well as a further confirmation of the trigger time in a specific operation with the help of experts, to minimize the impact on the generation of the attention map. Another factor is the AOIs defined by experts. Their importance may be not fully reflected in experts' attention models, e.g., Fig. 10 shows both experts focused little attention on A_2 in o_3.

In addition, the assessment procedure will be time consuming if the number of experts increases, as it is a pair-wise comparison method. To avoid this situation, it is possible to establish a mixed attention model by combining the attention model from each expert and optimizing the mixed model via evaluation by the experts themselves.

To sum up, if a well-tuned mixed attention model is created and the subjective factors are thoroughly considered, the expert-in-the-loop framework for training personnel for maritime operation can be highly effective.

VI. Conclusion

In this paper, we make use of expertise and experience of maritime operations to model and assess visual attention in an expert-in-the-loop training program. As a fundamental of the training program, expertise is utilized to divide the task, identify critical operation, and define AOIs. Each expert's visual attention is modeled from both spatial and temporal perspectives, forming a weighted attention map and a transitional pattern between AOIs. A saliency similarity metric is designed that accounts for trainees' fixation under the transitional pattern and its attention density in the weighted attention map. Assessment of visual attention in a heavy lifting operation is carried out. From the results using two groups of trainees, we conclude the proposed method is valid to assist the training program of maritime operations. Future work will be focused on the optimization of weight scale factor using more objective criteria, as well as the development of a mixed attention model according to experts' experience in the simulator.

Acknowledgment

The authors would like to thank Mr. Jo Even Tomren for inviting participants to the experiment and Mr. Bjørn Aase Dimmen from Offshore Simulator Centre AS for technical support in simulator.

References

[1] K. Johnson and G. Dalton, *Building industries at sea:'blue growth'and the new maritime economy*. River Publishers, 2018.

[2] M. Luo and S.-H. Shin, "Half-century research developments in maritime accidents: Future directions," *Accident Analysis & Prevention*, vol. 123, pp. 448 – 460, 2019.

[3] H. P. Berg, "Human factors and safety culture in maritime safety," *the International Journal on Marine Navigation and Safety of Sea Transportation*, vol. 107, 2013.

[4] T. Tien, P. H. Pucher, M. H. Sodergren, K. Sriskandarajah, G.-Z. Yang, and A. Darzi, "Eye tracking for skills assessment and training: a systematic review," *journal of surgical research*, vol. 191, no. 1, pp. 169–178, 2014.

[5] B. Law, M. S. Atkins, A. E. Kirkpatrick, and A. J. Lomax, "Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment," in *Proceedings of the 2004 symposium on Eye tracking research & applications*. ACM, 2004, pp. 41–48.

[6] A. Piras, R. Lobietti, and S. Squatrito, "A study of saccadic eye movement dynamics in volleyball: comparison between athletes and non-athletes," *Journal of Sports Medicine and Physical Fitness*, vol. 50, no. 1, p. 99, 2010.

[7] P. Konstantopoulos, P. Chapman, and D. Crundall, "Driver's visual attention as a function of driving experience and visibility. using a driving simulator to explore drivers eye movements in day, night and rain driving," *Accident Analysis & Prevention*, vol. 42, no. 3, pp. 827–834, 2010.

[8] S. R. Bowling, M. T. Khasawneh, S. Kaewkuekool, X. Jiang, and A. K. Gramopadhye, "Evaluating the effects of virtual training in an aircraft maintenance task," *The International Journal of Aviation Psychology*, vol. 18, no. 1, pp. 104–116, 2008.

[9] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the 2000 symposium on Eye tracking research & applications*. ACM, 2000, pp. 71–78.

[10] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011.

[11] R.-J. Dzeng, C.-T. Lin, and Y.-C. Fang, "Using eye-tracker to compare search patterns between experienced and novice workers for site hazard identification," *Safety science*, vol. 82, pp. 56–67, 2016.

[12] C. P. Ryffel, C. M. Muehlethaler, S. M. Huber, and A. Elfering, "Eye tracking as a debriefing tool in upset prevention and recovery training (uprt) for general aviation pilots," *Ergonomics*, no. just-accepted, pp. 1–29, 2018.

[13] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 185–207, 2013.

[14] O. Le Meur and T. Baccino, "Methods for comparing scanpaths and saliency maps: strengths and weaknesses," *Behavior research methods*, vol. 45, no. 1, pp. 251–266, 2013.

[15] K. Kurzhals, M. Hlawatsch, F. Heimerl, M. Burch, T. Ertl, and D. Weiskopf, "Gaze stripes: Image-based visualization of eye tracking data," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 1, pp. 1005–1014, 2016.

[16] T. Falkmer and N. P. Gregersen, "A comparison of eye movement behavior of inexperienced and experienced drivers in real traffic environments," *Optometry and vision science*, vol. 82, no. 8, pp. 732–739, 2005.

[17] A. K. Mackenzie and J. M. Harris, "A link between attentional function, effective eye movements, and driving ability." *Journal of experimental psychology: human perception and performance*, vol. 43, no. 2, p. 381, 2017.

[18] J. H. Yang, Q. Kennedy, J. Sullivan, and R. D. Fricker, "Pilot performance: assessing how scan patterns & navigational assessments vary by flight expertise," *Aviation, space, and environmental medicine*, vol. 84, no. 2, pp. 116–124, 2013.

[19] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara, "Predicting human eye fixations via an lstm-based saliency attentive model," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 5142–5154, 2018.

[20] D. M. Green, J. A. Swets *et al.*, *Signal detection theory and psychophysics*. Wiley New York, 1966, vol. 1.

[21] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 5, pp. 802–817, 2006.

[22] T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," MIT, Tech. Rep., 2012.

[23] S. Mathe and C. Sminchisescu, "Actions in the eye: Dynamic gaze datasets and learnt saliency models for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 7, pp. 1408–1424, 2015.

[24] L. Dempere-Marco, X.-P. Hu, S. M. Ellis, D. M. Hansell, and G.-Z. Yang, "Analysis of visual search patterns with emd metric in normalized anatomical space," *IEEE Transactions on Medical Imaging*, vol. 25, no. 8, pp. 1011–1021, 2006.

[25] R. J. Peters, A. Iyer, L. Itti, and C. Koch, "Components of bottom-up gaze allocation in natural images," *Vision research*, vol. 45, no. 18, pp. 2397–2416, 2005.

[26] R. J. Peters and L. Itti, "Applying computational tools to predict gaze direction in interactive visual environments," *ACM Transactions on Applied Perception (TAP)*, vol. 5, no. 2, p. 9, 2008.

[27] M. Kolbe, B. Grande, and D. R. Spahn, "Briefing and debriefing during simulation-based training and beyond: Content, structure, attitude and setting," *Best Practice & Research Clinical Anaesthesiology*, vol. 29, no. 1, pp. 87 – 96, 2015.

[28] R. M. Foerster and W. X. Schneider, "Functionally sequenced scanpath similarity method (funcsim): Comparing and evaluating scanpath similarity based on a tasks inherent sequence of functional (action) units," *Journal of Eye Movement Research*, vol. 6, no. 5, 2013.

[29] C. Eliasmith and C. H. Anderson, *Neural engineering: Computation, representation, and dynamics in neurobiological systems*. MIT press, 2004.
[30] B. De Smet, L. Lempereur, Z. Sharafi, Y.-G. Guéhéneuc, G. Antoniol, and N. Habra, "Taupe: Visualizing and analyzing eye-tracking data," *Science of Computer Programming*, vol. 79, pp. 260–278, 2014.
[31] G. Li, E. Homlong, H. Zhang, and B. A. Dimmen, "Integration of visual focus into marine operation simulator for behavior observation and analysis," in *OCEANS 2017-Aberdeen*. IEEE, 2017, pp. 1–7.

Guoyuan Li (M'14) received the Ph.D. degree from the Institute of Technical Aspects of Multimodal Systems (TAMS), Department of Informatics, University of Hamburg, Hamburg, Germany, in 2013. From 2014, he joined the Mechatronics Laboratory, Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology, Norway. From 2018, he become an associate professor in ship intelligence. He has extensive research interests including eye tracking analysis, modeling and simulation of ship motion, artificial intelligence, optimization algorithms and locomotion control of bio-inspired robots. In these areas, he has published over 40 papers.

Runze Mao received the B.Eng. from the School of Naval Architecture & Ocean Engineering, Jiangsu University of Science and Technology, Zhenjiang, Chian in 2015. He is currently a Master's student from Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology, Norway. His research interest includes maritime operation and eye tracking.

Hans Petter Hildre is professor and head of the Department of Ocean Operations and Civil Engineering at the Norwegian University of Science and Technology (NTNU). His area of interest is product design and system architecture design. Hans Petter is Centre Director for Centre for Research Driven Innovation (SFI-MOVE) within maritime operations. This is cooperation between NTNU, SINTEF, University Sao Paulo and 15 companies at the west coast of Norway. Professor Hildre is head of research in national program Global Centre of Expertise Blue Maritime, project leader in several research projects, member of the board in 5 companies, and has a number of patents.

Houxiang Zhang (M'04-SM'12) received Ph.D. degree in Mechanical and Electronic Engineering in 2003. From 2004, he worked as Postdoctoral Fellow at the Institute of Technical Aspects of Multimodal Systems (TAMS), Department of Informatics, Faculty of Mathematics, Informatics and Natural Sciences, University of Hamburg, Germany. In Feb. 2011, he finished the Habilitation on Informatics at University of Hamburg. Dr. Zhang joined the NTNU, Norway in April 2011 where he is a Professor on Robotics and Cybernetics. The focus of his research lies on two areas. One is on biological robots and modular robotics. The second focus is on virtual prototyping and maritime mechatronics. In these areas, he has published over 130 journal and conference papers and book chapters as author or co-author.