2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
September 24–28, 2017, Vancouver, BC, Canada

Measurement and Prediction of Situation Awareness in Human-Robot Interaction based on a Framework of Probabilistic Attention*

Amir Dini, Cornelia Murko, Saeed Yahyanejad, Ursula Augsdörfer, Michael Hofbaur, and Lucas Paletta

Abstract—**Human attention processes play a major role in the optimization of human-robot interaction (HRI) systems. This work describes a novel methodology to measure and predict situation awareness and from this overall performance from gaze features in real-time. The awareness about scene objects of interest is described by 3D gaze analysis using data from wearable eye tracking glasses and a precise optical tracking system. A probabilistic framework of uncertainty considers coping with measurement errors in eye and position estimation. Comprehensive experiments on HRI were conducted with typical tasks including handover in a lab based prototypical manufacturing environment. The methodology is proven to predict standard measures of situation awareness (SAGAT, SART) as well as performance in the HRI task in real-time and will open new opportunities for human factors based performance optimization in HRI applications.**

I. Introduction

Collaborative robotics has recently progressed to human-robot interaction in real manufacturing. Human factors are crucial as industrial robots are enabling human and robot workers to work side by side as collaborators and to assess the user's experience with a robot, while understanding how humans feel during their interaction with it [1]. Furthermore, human-related variables are essential for the evaluation of human-interaction metrics [2].

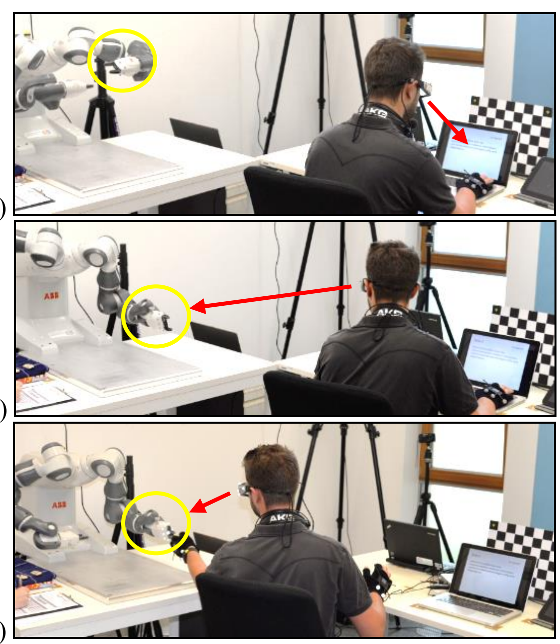
To work seamlessly and efficiently with their human counterparts, robots must similarly rely on predictions of the human worker's behavior, emotions, task specific actions and intent to plan their actions, such as, in anticipatory control with human-in-the-loop architecture [3] to enable robots to proactively perform task actions based on observed gaze patterns to anticipate actions of their human partners according to its predictions. However, measuring and modeling of situation awareness based on gaze triggered information recovery is mandatory for the understanding of immediate and delayed action planning.

This work elaborates a framework on human situation awareness in the manufacturing domain of human-robot interaction on the basis of concrete measures of eye movements towards production relevant processes that need to be observed and evaluated by the human.

Motivated by the theoretical work of [4] on situation awareness it specifically aims at dynamically estimating (i) the principal saliency and (ii) the perception of objects of relevance for the production process by the human observer and worker, respectively, as well as (iii) the semantics in eye movement behavior in the frame of process understanding and tracking. Gaze in the context of collaboration is analyzed in terms of - primarily, visual - affordances for collaboration.

A human worker has to learn to anticipate a randomly delayed handover event and is informed by gaze. Handover of relevant components in the collaborative robotics process requires coordination strategies that predict user states [5].

In this work we stress the relevance of considering eye movement features for a profound characterization of gaze behavior, with the purpose to optimize anticipatory behavior. The estimation of the specific status of situation awareness of the human worker can be crucial for the elaboration of performance analysis through measurement of executive functions, evaluation of interruption impact, as well as for the prediction of accidents.



(a) (b) (c)

Figure 1. Human-robot interaction and attention based situation awareness. (a) Focused on primary task (reading aloud), (b) head and eye gaze to get aware about secondary task (handover), (c) handover in time.

* This work has been supported by the Austrian Ministry for Transport, Innovation and Technology (BMVIT) within the project framework CollRob (Collaborative Robotics).

Amir Dini, Cornelia Murko, and Lucas Paletta (corresponding author with phone: +43-316-876-1769; fax: +43-316-876-91769; e-mail: lucas.paletta@joanneum.at) are with DIGITAL – Institute for Information and Communication Technologies, JOANNEUM RESEARCH Forschungsgesellschaft mbH, Graz, Austria.

Saeed Yahyanejad and Michael Hofbaur are with ROBOTICS – Institute for Robotics and Mechatronics, JOANNEUM RESEARCH Forschungsgesellschaft mbH, Klagenfurt, Austria.

Ursula Augsdörfer is with Graz University of Technology, Institute for Computer Graphics and Knowledge Visualisation, Graz, Austria.

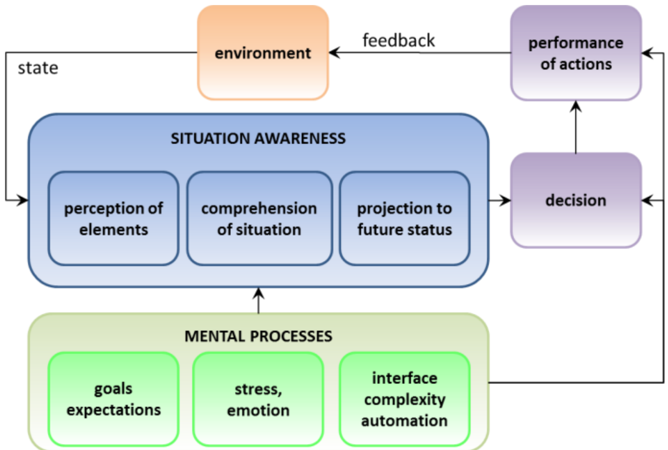


Figure 2. Schematic sketch of Endsley's [4] theory of situation awareness in dynamic situations and its interdependency on human decision making, action performance and mental processes.

Figure 1 sketches the relevance of situation awareness for performance in collaborative robotics: (a) The worker's attention is focused on the primary task and not aware about robot activities, (b) the worker turns head and fixates relevant elements of the robot process, (c) based on a successful model of the randomized robot action sequence, the worker is capable to anticipate the "handover" event and put the component into the gripper of the robot just in time.

This work presents an innovative methodology to measure situation awareness in real-time from gaze interaction with objects of interest using (i) eye tracking glasses and a (ii) motion capture system for 3D positioning. As conceptual contribution, real-time eye tracking based attention analysis is for the first time – according to the knowledge of the authors - related to human situation awareness, in particular, in the frame of process oriented monitoring and understanding. The novel probabilistic framework of uncertainty copes with measurement errors in eye and position tracking. The presented methodology is proven to predict HRI performance in real-time and will open new opportunities for human factors based performance optimization in numerous HRI application scenarios.

The presented work extends human factors measurements in HRI in the frame of situation awareness in the following dimensions: (i) Real-time measurement using mobile eye tracking glasses (Figure 3), (ii) gaze mapping in a framework of uncertainty (probability theory; Figure 4), (iii) situation awareness based on 3D point distribution processes (Figure 5) and (iv) application in the area of HRI based on recognizing and predicting human mental status which enables situation awareness processing with the human–in–the–loop (Figure 9).

II. Related Work

A. Psychological Measures of Situation Awareness

Based on the cognitive ability, flexibility and knowledge of human beings on the one hand and the power, efficiency and persistence of industrial robots on the other hand, collaboration between both elements is absolutely essential for flexible and dynamic systems like manufacturing [6]. Efficient human-robot collaboration requires a comprehensive perception of essential parts of the working environment of both sides and an alternately as well as clearly interaction in task design and task process. Human decision making is a substantial component of collaborative robotics under dynamic environment conditions, such as, within a working cell. Situation awareness is in this context crucial, in particular, to identify decisive parts of task execution.

Endsley [4] described three essential levels of situation awareness (Figure 2) which are understood as both inclusive and progressive components.

- The *first level* considers the perception of elements in a current situation which is understood as kind of registration of occurrence of objects in space.
- The *second level* represents an advanced state of situation awareness and concerns the comprehension of a current situation, understanding the context in which elements occur, such as, by following the trajectory of an object of relevance in space at the time where it is relevant, such as, for interaction.
- The *third level* concerns even more advanced awareness of the situation by being able to project to future states based on a predictive situation model.

In human factors, situation awareness is principally evaluated through questionnaires, such as, (i) the Situation Awareness Global Assessment Technique (SAGAT; [7]) or (ii) the Situational Awareness Rating Technique (SART; [8]). Psychological studies on situation awareness are drawn in several application areas, such as, in air traffic control, driver attention analysis, or military operations.

SAGAT is known to provide a more objective measurement methodology, however, for its measurement, the task must be stopped which requires application merely under the execution of a fully controllable simulation tool. In the contrary, the SART questionnaire is performed after the end of the simulation, on the basis of a subjective estimation of situation awareness of the operator herself [8]. Following [9] no studies yet reported significant correlation between scores of SAGAT and SART, respectively, concluding that the SAGAT and SART are assessing different aspects of SA. "*SAGAT, measures the extent to which a participant is aware of pre-defined elements in the environment and their understanding of these elements. SART, on the other hand, provides a measure of how generally aware participant's perceive themselves to be without referring to specific elements within the environment*" [10].

Kallus [12] emphasized the relevance of anticipatory processes for errors cooperation tasks and proposed to change the classical view on situation awareness into human-in-the-loop control. More recently, relevant environments were viewed as 'schemata' [13] or 'frames' [14] of relevant objects and their spatiotemporal relations. This work intends to follow this framework by enabling closed-loop situation awareness control from real-time eye movement measures.

B. Situation Awareness in Technical Systems

Due to the disadvantages of the questionnaire technologies of SART and SAGAT, more reliable and less invasive technologies were required, however, among others, eye tracking - representing a psycho-physiologically based, quantifiable and objective measurement technology - has been proven to be effective [11][13]. In several studies in the frame of situation awareness, eye movement features, such as dwell and fixation time, were found to be correlated with various measures of performance [15][16][17][18][19].

III. Situation awareness in a Framework of Probabilistic Attention

A. Recovery of 3D Gaze in Human-Robot Interaction

Localization of human gaze is essential for the localization of situation awareness with reference to relevant processes in the working cell. [22] firstly proposed 3D information recovery of human gaze with monocular eye tracking and triangulation of 2D gaze positions of subsequent key frames within the scene video of the eye tracking system. [23] proposed gaze estimation in 3D space with a special stereo rig that is required in addition to a commercial eye tracking device, and achieved indoor measurement accuracy of ≈ 3.6 cm at 2 m distance. [24] achieved even improved accuracies ≈ 1 cm with RGB-D based position tracking within a predefined 3D model of the environment.

In order to achieve the highest level of gaze estimation accuracy in a prototypical research setup, it is crucial to track a user's frustum and gaze behavior with respect to the worker's relevant environment. Solutions that realize this include vision-based motion capturing systems, such as, OptiTrack [25] (Figure 3).

However, while such approaches can achieve high tracking and gaze estimation accuracy (≈ 0.06 mm), the need to deploy them for every collaborative robot and work cell the user might want to interact with currently severely limits uptake and truly pervasive and spontaneous gaze-based interaction. For the purpose of feasible work cell equipment, vision based positioning by artificial [26][27] or natural [21] landmarks are recommended.

B. Probabilistic Situation Awareness from Gaze

In the presented framework, situation awareness (SA) is based on several levels following Endsley's seminal theoretical concept [4]. The following definitions were used in the analysis of the collaborative robotics study in reference to Endsley's three levels of situation awareness, as follows,

1. **Oriented perception** (SA level 1, SAL1): coverage of objects / volume of interest within the worker's view frustum. The object enters the horizon of *covert attention* [28] once it appears in the field of view. This work proposes SAL1 in terms of *fixation and frustum based coverage of object presence*, using probabilistic gaze and *turn rates* performed between objects of interest.
2. **Focused perception** (SAL2): coverage of objects within the worker's focus of attention through object related fixations. The object enters the area of highest resolution in the analysis, a process often caused by *overt attention* [29] under conscious control of eye movements. This work proposes SAL2 in terms of the *3D-NNI index* (Section III.C), i.e., by characterizing the distribution of fixation points with respect to objects of interest as well as by *average dwell time* [30] of fixations on those objects.
3. **Comprehension** (SAL3): this level would require a projection into the future. Anticipative behavior in terms of top-down action-oriented attention processes [31] should be evaluated towards action critical and decisive events in the tasks. In this work, SAL3 is modelled by *3D-NNI* in *conjunction with turn rates* in order to fit well estimates on SAL3 questionnaire evaluation.

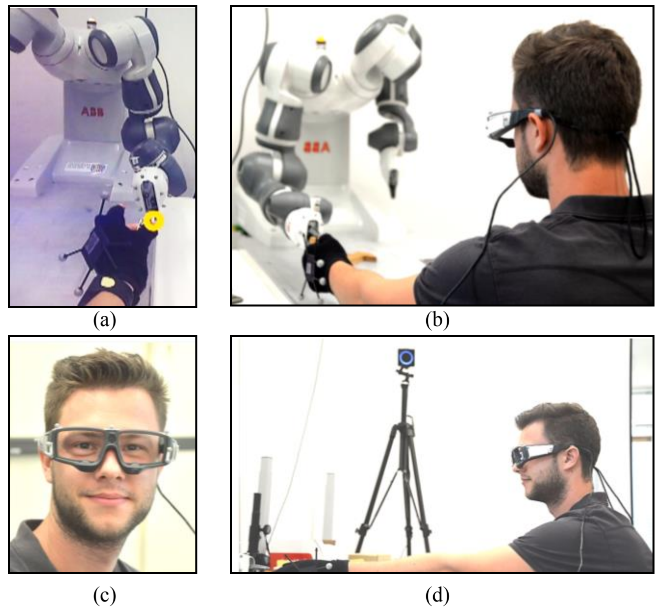


Figure 3. Human-robot interaction under monitoring and evaluation of attention based situation awareness. The human worker has to learn to anticipate a randomly delayed handover event and is informed by gaze. (a) Egocentric view of the worker represented by the scene camera's view from eye tracking glasses, (b) handover of a workpiece. IR markers for the position monitoring of (c) eye tracking glasses and (d) human moton capture (hands) by OptiTrack IR cameras (in the background).

Basic requirement for all proposed analysis is a *robust* estimate of the intersection of the gaze vector $\boldsymbol{v}_t$ at time t with objects $O_{i,t}$ in the scene. For this purpose, a probabilistic evaluation of situation awareness is outlined based on preliminary progress [32] as follows. Let distance $\chi_{i,t}$ of object $O_{i,t}$ define a measure of awareness about the presence of the object $O_{i,t}$ by means of the observer's current gaze $\boldsymbol{v}_t$: $\chi_{i,t}$ is then associated with the length of the vector that originates at object center $\boldsymbol{C}_{i,t}$ and that is conditioned to orthogonally intersect with gaze $\boldsymbol{v}_t$. Then we define the probability $P(\cdot)$ of intersecting object $O_{i,t}$ through gaze $\boldsymbol{v}_t$ by

$$P(O_i|\boldsymbol{v}_t) = F\, exp(-\chi_{i,t}/2\sigma^2), \qquad (1)$$

with normalizing factor F; σ defines the uncertainty in the eye tracking measurements.

Figure 4 visualizes situated relations in probabilistic object awareness from uncertain gaze: Human eye gaze ($\boldsymbol{v}_t$, red ray) is positioned within the viewing frustum (green, transparent pyramid) and intersects with object spheres (e.g., yellow sphere of laptop object) that represent a current uncertainty in object localization by eye gaze. Eye gaze would be orientated at laptop object center via object center ray (visualized by straight yellow line). Vectors from object center $\boldsymbol{C}_{i,t}$ are orientated towards orthogonal intersection with eye gaze ray (small, yellow non-transparent cube) and intersect with probabilistic object sphere (small, yellow non-transparent sphere). Note that for some objects, which are substantially out of sight, it is possible to orthogonally intersect only with negative gaze $\boldsymbol{v}_t$ (e.g., blue sphere / line).

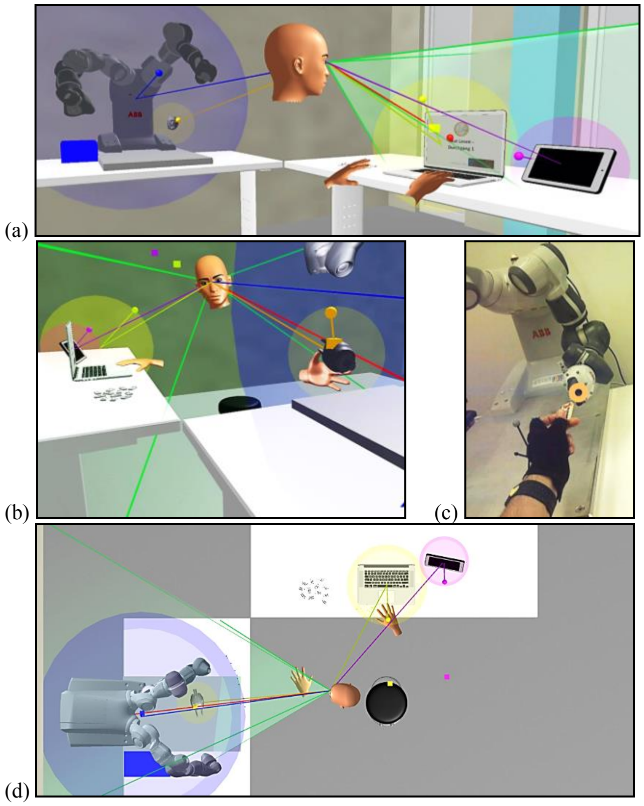


Figure 4. Probabilistic object awareness from uncertain gaze. (a) Full scene with relevant objects (robot / gripper / laptop / tablet / head / hands) each represented by colored and transparent spheres. Human attention with respect to these relevant objects is reconstructed based on relations between view frustum (green, transparent), view ray (red), and eye-object-center lines with orthogonal intersection on object view spheres. (b) Handover situation reconstructed and (c) from gaze video, with fixation on gripper. (d) Top orthographic projection about the collaborative task environment.

Distance χ is computed, as follows: $\boldsymbol{\tau}_t = \boldsymbol{v}_t \times \boldsymbol{\gamma}_t$, $\boldsymbol{\gamma}_t = \boldsymbol{C}_{i,t} - \boldsymbol{F}_t$; $\boldsymbol{C}_{i,t}$ is the vector to object center $\boldsymbol{C}_i$ and $\boldsymbol{F}_t$ is the camera focus vector; consequently $\boldsymbol{\eta}_t = \boldsymbol{\tau}_t \times \boldsymbol{v}_t$ results in a unit vector from $\boldsymbol{C}_{i,t}$ to an intersection with $\boldsymbol{v}_t$, with direction $\boldsymbol{\eta}_t \perp \boldsymbol{v}_t$, being orthogonal to the gaze. Gaze mapping in a framework of uncertainty (probability theory) enables probabilistic hit rates that are defined by means of exponentially decaying distance to the object center. Figure 4 visualizes these geometric relations with object view spheres, object center vectors, and intersections (small spheres) of orthogonal vectors $\boldsymbol{\eta}_t$ with the view sphere with $\chi_{i,t} = \sigma_i$.

C. *Novel Measure for 3D Fixation Distribution Analysis*

In human factors and ergonomics research, the analysis of eye movements enables to develop methods for investigating human operators' cognitive strategies and for reasoning about individual cognitive states [30]. Camilli et al. [20] have shown that distributions of eye fixations reflect variations in mental workload — *dispersed* when workload is high, and *clustered* when workload is low. Spatial statistics algorithms evaluate distribution types and can be further applied over fixations recorded during small epochs of time to assess online changes in the level of mental load experienced by the individuals. The NNI (Nearest Neighbor Index) has initially been developed as statistical index for scanning behavior of eye movements in 2D application [20], such as, on displays.

For the measurement of fixation distributions in 3D space, this work proposes a novel method - leading to the *3D-NNI* index - being based on the probabilistic measure for object awareness (Sec. III.0) and classical NNI methodology [20]. Computation of 3D-NNI is based on the requirement that the motion of observer view frustum should be comparably negligible compared to the angular dynamics of the viewing ray towards an object of interest, at least within the frame of small measurement durations, such as, between 10-20 seconds. Since this duration is typical for measuring the impact from psychophysiological effects, such as, mental workload, stress, emotion, concentration loss, etc., we can assume that 3D-NNI is capable to capture gaze interaction in the context of human factors dynamics. Following the methodology of [21] and [20], an actual density is measured by the distance d from each fixation to its nearest neighbor to generate the mean observed distance to be represented as

$$d_{act} = \Sigma_{(i,j=1..N)}\, d_{i,j} / N \ . \qquad (2)$$

In the presented framework, fixation points are on the view sphere around objects and distances are measured in arc length. For random distribution, the mean distance between fixation points was shown in [21] to have a value equal to

$$d_{rand} = 1/(2\sqrt{\rho}), \text{ and } I = d_{act} / d_{rand} \ . \qquad (3)$$

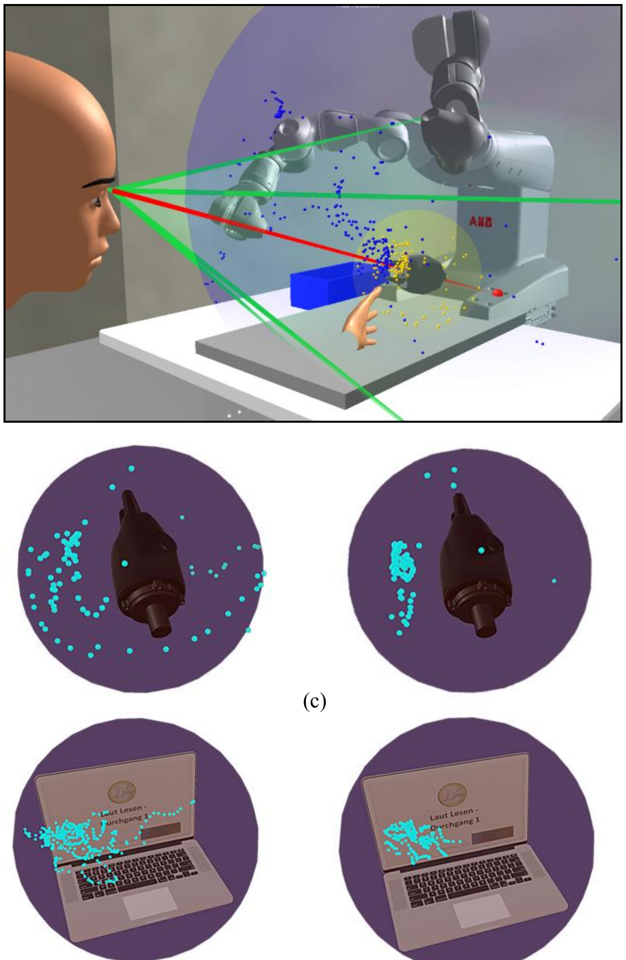
Hence I is used as a measure ("3D-NNI") of how much the observed distribution on the view sphere approaches or departs from random probability. The actual mean distances can be smaller (points are aggregated, i.e., $I \ll 1$), larger (points are regularly dispersed; $I \gg 1$), or not different from the expected distances (points are randomly dispersed; $I \approx 1$).

Figure 5 depicts sample spatial distributions associated with their 3D-NNI values generated by eye movements of the human worker on (a) the Yumi and (b, c) its gripper, as well as on (d, e) the 'laptop' object as source of information for a 'read aloud' task, during different working cycles.

Fixation distribution analysis based on 3D-NNI significantly contributes to measuring situation awareness (exploration behavior corresponds with multiple, distributed fixations and high I) and performance analysis (expert behavior corresponds with focused, clustered fixations and low I), as outlined in experimental results.

D. *Attentive Features to Discriminate Situation Awareness*

Situation awareness (SA) is a measure of an individual's knowledge and understanding of the current and expected future states of a situation. Eye tracking provides an unobtrusive measure to measure SA in environments where multiple tasks need to be controlled. In this context, [11] provided first evidence that fixation duration on relevant objects and balanced allocation of attention increases SA.



(a)

(b) (c)

(d) (e)

Figure 5. A novel measure (3D-NNI) for fixation distribution analysis (ref. [20]). (a) Gaze intersects with view sphere (probabilistic representation) of objects (robot: *purple*; gripper: *yellow*); intersection points are aggregated over time (≈20 sec.). (b) Point distribution on gripper for *dispersed* (*N=84 fixations, 3D-NNI 0.34*) and (c) *clustered* (*N=90; 3D-NNI 0.14*) intersections. Laptop sphere with (d) dispersed (*N=363, 3D-NNI 0.27*) and *focussed* (*N=269, 3D-NNI 0.12*) fixations.

In the frame of real-time tracking and analysis of eye movements, several features might be considered [30], however, in this work we focus on the following aspects:

- *look rate*: number of dwells ('gazes', 'glances', 'visits'), i.e., transitions in and out of an area of interest [30], mostly within a trial ('cycle' in this work). Increases in exploration stage and in complex situation, decreases with expertise.
- *average dwell time*: decreases with practice and expertise on a certain task.
- *turn rate*: number of changes of view between different objects. Decreases with control of the task.

IV. Predicting Situation Awareness & Performance

Research on eye tracking for situation awareness is focused on the analytical aspects, providing statistical results on eye movement features about the outcome of studies that lasted in the past. In this work we are interested in the exploitation of gaze informatics during operation, developing technical solutions for human-in-the-loop applications that enable assistive and timely response for the optimization of an integrated human-robot interaction system.

The framework implemented and presented in this work refers to statistical feature analytics on measurements from long-term monitoring of data that were typically captured after a duration of 2-3 minutes, and then to apply machine learning to determine estimators and classifiers that would be applied to short-term measurements, typically in the range of 20-24 seconds. The following classifier learning methodologies were applied:

- *Support Vector Machines* for regression [33] and classification [34].
- Binary regression & classification *decision trees* [35].

Further analysis and application to technical solution is dedicated to future work, here we focus on highlighting the potential of estimation and prediction of situation awareness measurements, especially in the industrial context.

V. Experimental Results

A. Rationale of Human-Robot Interaction Study

The performance evaluation of the presented methodology is based on a human factors study with a multiple task and handover scenario, performed within a robotics laboratory including several portable multisensory interfaces connected in a HRI system network.

The human worker is concentrated on a current primary task (read aloud of as many as possible text pages) and has to anticipate via gaze the handover procedure of an ABB Yumi robot with a randomized time period of fulfilling a sequence of actions and poses (Figure 6).

The fundamental rationale of this study is defined as follows: The characteristic value in human contribution to collaborative HRI systems – i.e., the contribution that will not be replaced by robots within the next decades – is understood as performing high level cognition tasks, represented by the (primary) reading task. Read aloud was chosen to enable emotion recognition from speaker's prosody in future evaluation stages of the study.

Figure 6. Sequence of postures of the robot. The worker had to visually anticipate the "handover" event (no. 1) from the actual appearance of any posture viewed during her task.

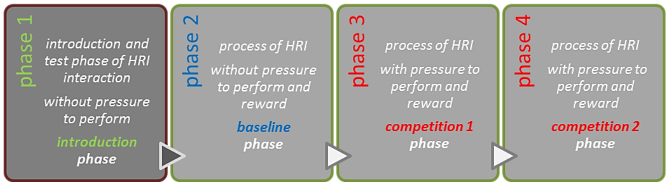


Figure 7. Phases and conditions of the human factors field study.

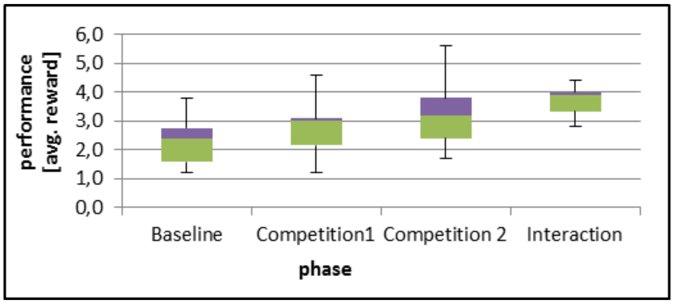


Figure 8. Increase of performance with experience of additional and increasingly competitive phases. In 'Interaction' phase, complementary information about SA was provided to the operators to boost performance.

The handover task – providing a domino just in time – represents a human specific value being transferred in a sensory-motoric manner to the physical aspect – the manipulator – of the robotic system. Important is the spatiotemporal aspect of human situation, i.e., getting to know, based on a human internal model of robot operation, about when to interact and where to, i.e., to the gripper. The completion of the overall task, after having received the handed over object, is left imaginary in this study.

The collaborative aspects, in particular, related to manufacturing and assembly team work, could be worked out in more detail in future studies, however, the key objective of this work is to focus on the aspect of the situation awareness, as well as on the distribution of attention to different processes, and how this can be measured.

The overall field of view I the environment was chosen to be sufficiently wide in order to urge human operators to turn the head and substantially change orientation of head gaze in order to enable to differentiate between contribution of head and eye gaze to overall situation awareness measurements.

The evaluation correlates – in a holistic manner – human situation awareness measured by standard SAGAT as well as SART questionnaires with measurements about probabilistic worker's gaze and human attention from eye movement features. The research question is whether questionnaires can be replaced by automated estimation and classification of measured human attention processes during task operation. SART questionnaire was applied after completion of the task as requested in [8]. For the application of the SAGAT questionnaire, the whole task has to be interrupted and task specific questions need to be posed to the human operator [7]. For this purpose, a human moderator sitting behind the participant of the study was able to stop the Yumi processing at any (randomly determined) time with a central control tablet and the participant was alerted and requested to immediately orientate to fill the SAGAT questionnaire. This procedure certainly interrupts the flow of task operation, however, it was necessary with the purpose to match standard SAGAT questionnaire output with human attention features.

B. *Descriptive Statistics*

Participants. The lab study involved 19 participants, 12 of them (5 female, 7 male, M 26.92, SD 4.96 of age) were used in the analysis of the presented work. Removal criteria involved casual malfunction of gripper and confused communication about the SAGAT stopping criterion.

Phases of study. Each participant of the study was involved in 4 phases within a complete session taking about 1.5 hours. First (phase 1, Figure 7), the concrete tasks of the worker were explained during operation by a moderator. Thereafter participants worked engaged (phase 2, '*baseline*') and, additionally, in two more phases, were told that they could earn money by the degree of success in the completion of the primary and secondary tasks (phases 3 & 4, '*competition1*' & '*competition2*'). Each phase consisted of 10 robot cycles with identical sequence of robot actions (postures). Cycles have randomly defined duration of the full action sequence with M 22, SD 2.2 seconds and uniform distribution upon the 18-26 seconds interval.

TABLE I. INFERENTIAL STATISTICS ON SITUATION AWARENESS FEATURES WITH SPEARMAN'S RANK (A) AND PEARSON (OTHER) CORRELATION MEASURE.

(A) SAGAT questionnaire	***SAL1: hit rate (gripper)***		
	phase 2	*phase 3*	*phase 4*
phase 4 (ρ)	.070	.196	.608*
(B) SART questionnaire	***SAL2: discriminative features***		
	look rate (laptop)	*turn rate (robot/laptop)*	*3D-NNI (gripper)*
all phases (r)	.553**	-.416	-.368
(C) SAGAT questionnaire	***SAL3: discriminative features***		
	3D-NNI (gripper)	*look rate (robot)*	*avg dwell (laptop)*
all phases (r)	.501**	.413	.375
(D) user performance	***discriminative features***		
	turn rate (robot/laptop)	*look rate (gripper)*	*look rate (robot)*
all phases (r)	-.617**	-.516**	-.457

**) p<0.05 (two-tailed) **) p<0.05.*

Payoff procedure and performance. Payoff for participation was a minimum 30 € with opportunity to extend by a maximum performance driven premium of 20 €. Participants were presented a maximum of 6 pages for read aloud with payoff for each read aloud page. Fundamental condition for any payoff was successful handover of a domino in time. Figure 8 provides an overview on performance increase with additional phases, depicting the learning effect. The boxplots demonstrate the median value separating the lower (green) and upper (violet) quartile of the distribution of the performance values. Figure 8 depicts particularly the increase in the median with increasing experience and competition. The 'interaction' phase will be evaluated in future studies, it refers to a phase where the user was informed about the next handover moment in time.

C. 3D Gaze Tracking Using Human Motion Capture

In order to achieve the highest level of gaze estimation accuracy in the study setup, the dynamic position of (i) the eye tracking glasses, (ii) the hands of the worker, and (iii) the gripper of the robot were tracked relative to the worker's environment (robot trunk, notebook with reading task). The choice of vision based motion capturing system was OptiTrack [25] using 9 cameras of type 'Prime 41' with 4.1 MP resolution, a 100' tracking range, 180 fps resolution of motion capture, and 51° field of view. The accuracy of marker positioning was reported to be 0.06 mm by the Motive Software from OptiTrack.

D. Inferential Statistics - Features for Situation Awareness

Inferential statistics was applied on all eye movement and attention features described in Sections III.C and III.D, and referring to SAL1, SAL2 and SAL3 processes. Situation awareness was investigated by (i) the SAGAT questionnaire measurements, by stopping the task, providing 196 SAGAT questionnaires with a score between 1 and 10 to rate the worker's prediction about the handover event, and by (ii) measuring SART in post-task questionnaires.

In the frame of (A) *SAGAT / SAL1*, the Spearman's rank correlation coefficient (ρ) between the SAGAT measure and the hit rate on the gripper of the robot was substantial and highest amongst all objects' hit rates (ρ=0.608; $p<0.05$, see TABLE I, (A)), being measured in final phase 4. This suggests that situation awareness in the frame of SAGAT analysis can be automatically estimated from gaze features that are set in context of HRI task relevant processes in the environment. The correlation is lower in phases 2-3 indicating that a substantial learning stage has taken place before phase 4 in the context of human situation awareness. In the context of (B) *SART / SAL2* (TABLE I, (B)), look rate, turn rate, but also 3D-NNI provide highly discriminative features. Note that these features were correlated between mean values obtained in each of phases 2-4 ('*all phases*'). In the context of (C) SAGAT / SAL3 (TABLE I, (C)), 3D-NNI, look rate, and average dwell time provide discriminative features.

Finally, the *user performance* of workers (Figure 8: overview on performance increase with additional phases, depicting the learning effect) is best estimated from correlated features about 3D-NNI, look rate and average dwell time.

SAGAT and SART correlated with r=-0.382; $p<0.05$, demonstrating some overlap, in contrast to [9]. Interestingly, *user performance* and *SART* values correlated with r=0.510, $p<0.05$, for low performance, and r=-0.207, $p<0.05$ for high performance values, respectively.

E. Prediction of Situation Awareness and Performance

Finally, discriminative features were selected to feed into neural network based machine learning approaches. The objective of this task is to evaluate the accuracy of predicting important human factors and performance measures and from this becoming capable to define assistive services.

Prediction of situation awareness (SAGAT). Values for SAGAT (SAL3) were found between 32.7% - 73.0%, with M 55.9%, SD 10.2%. *Two classes* were defined, SA_{SAGAT} > 51.5% for class 1 ('high values of SA, i.e., SAGAT), otherwise class 0 ('low values of SA'). A classification support vector machine network [34] was trained with features (*3D-NNI (gripper) and look rate (robot)*) and 4-fold cross-validation, it achieved a *classification accuracy of **83.3%** in the prediction*.

Prediction of situation awareness (SART). Values for SART (SAL2) were found between 14 - 42, with M 30.0, SD 7.6. *Two classes* were defined, SA_{SART} > 30 for class 1 ('high SA'), otherwise class 0 ('low SA'). A classification support vector machine network [30] was trained with features (*3D-NNI (gripper), look rate (laptop), turn rate (robot/laptop), and avg. dwell (gripper))* and 10-fold cross-validation, it achieved a *classification accuracy of **91.7%*** in the prediction.

Prediction of user performance. Values for user performance (Π) were found between 1.2 – 5.6, with M 2.92, SD 1.41. Figure 9 with predictions from SVM based regression fit [33] demonstrates very good accuracy. *Three classes* were defined, Π < 2.67 for class 1 ('low performance'), if Π < 4.13 class 2 ('medium performance'), otherwise 'good performance'. A classification support vector machine network [30] was trained with features (*3D-NNI (robot), look rate (gripper, robot), and turn rate (robot/laptop))* and 10-fold cross-validation, it achieved a *classification accuracy of **79.2%*** in the prediction.

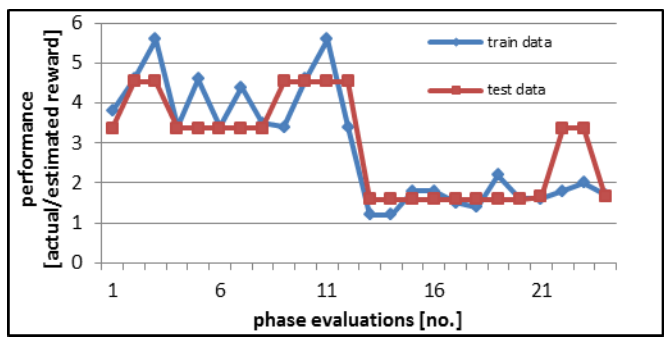


Figure 9. Prediction of train / test data, using a SVM regression network [33], demonstrates substantial accuracy for future assistance technologies.

Overall, classification performance – exclusively collected from real-time eye tracking data - is sufficiently accurate to serve as a solid basis for assistance technologies in the future.

VI. Conclusions and Future Work

The contribution of probabilistic situation awareness and vision based gaze localization in collaborative robotics results in successful estimation of gaze based awareness measures, to recover situation awareness in HRI during real-time tasks. The estimation of the specific status of situation awareness of the human worker can be crucial for the elaboration of performance, acceptance, executive function and interruption analysis. Future work will focus on developing improved parametrizations of the situation awareness model by gaze and performance measures, in particular, SAL3 modelling could be applied with anticipatory behavior modelling, such as, from neural network time series models. Furthermore, the impact of stress and emotion on attention and situation awareness will be modelled and evaluated.

REFERENCES

[1] L. Tiberio, A. Cesta, and M. O. Belardinelli, “Psychophysiological methods to evaluate user’s response in human robot interaction: a review and feasibility study”, *Robotics*, 2(2):92–121, 2013.

[2] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, „Common metrics for human-robot interaction”, *Proc. ACM SIGCHI/SIGART Human-robot interaction*, 2006.

[3] C.-M. Huang, and B. Mutlu. “Anticipatory robot control for efficient human-robot collaboration”, *Proc. ACM/IEEE HRI 2016,* 2016.

[4] M. R. Endsley, “Toward a theory of situation awareness in dynamic systems,” *Human Factors*, vol. 37, no. 1, pp. 32–64, 1995.

[5] C.-M. Huang, B. Mutlu, and M. Çakmak. Adaptive Coordination Strategies for Human-Robot Handovers, *Proc. Robotics: Science and Systems (RSS 2015)*, 2015.

[6] C. Heyer, “Human-robot interaction and future industrial robotics applications,” *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4749–4754, 2010.

[7] M. R. Endsley, “Direct measurement of situation awareness: Validity and use of SAGAT,” in *Situation Awareness Analysis and Measurement*, M. R. Endsley, and D. J. Garland, Eds., 2000.

[8] R. M. Taylor, “Situational awareness rating technique (SART): The development of a tool for aircrew systems design”, in *Situational Awareness in Aerospace Operations,* 1990, pp. 3/1–3/17.

[9] P. M. Salmon, N.A. Stanton, G.H. Walker, D. Jenkins, D., D. Ladva, L. Rafferty, & M. Young. Measuring Situation Awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics*, 39(3), 490-500, 2009.

[10] Beukel, A.P. van den and Voort, M.C. van der (2014) Driver's situation awareness during supervision of automated control - comparison between SART and SAGAT measurement techniques. In: *4th European conference on human centred design for intelligent transport systems*, Vienna, 5-6 June 2014, 05-06-2014 - 06-06-2014.

[11] K. Moore, and L. Gugerty, “Development of a novel measure of situation awareness: The case for eye movement analysis,” *Human Factors & Ergon. Soc. Ann. Meeting*, vol. 54:19, 2010.

[12] K.W. Kallus. Situationsbewusstsein und antizipative Prozesse. [Situation awareness and anticipatory processes] *Zeitschrift für Arbeitswissenschaft*, 1, 17.22, 2009.

[13] Stanton, N.A., P.M. Salmon, G.H. Walker, and D.P. Jenkins. Genotype and phenotype schemata and their role in distributed situation awareness in collaborative systems. *Theoretical Issues in Ergonomics Science* 10 (1):43---68, 2009.

[14] J. Lundberg, J. Situation Awareness Systems, States and Processes: A holistic framework. *Theoretical Issues in Ergonomics Science*, 2015.

[15] S. Mamessier, D. Dreyer, and M. Oberhauser, “*Calibration of Online Situation Awareness Assessment Systems Using Virtual Reality,”* in *Digital Human Modeling. Applications in Health, Safety, Ergonomics and Risk Management*, vol. 8529, V. G. Duffy, Ed. Switzerland: Springer International Publishing, 2014, pp. 124–135.

[16] W.J. Horrey, C.D. Wickens, and K.P. Consalus, “Modeling drivers’ visual attention allocation while interacting with in-vehicle technologies”, *Journal of Experimental Psychology: Applied*, 12(2), 67-78, 2006.

[17] G. Hauland, “Measuring individual and team situation awareness during planning tasks in training of en route air traffic control”, *The International Journal of Aviation Psychology*, 18(3), 290-304, 2008.

[18] S.A. Balk, K.S: Moore, J.E. Steele, and W Spearman, “Mobile phone use in a driving simulation task: Differences in eye movements”, *Poster at the annual meeting of the Vision Sciences Society*, 2006.

[19] R.J. Jacob, and S.K. Karn, “Eye tracking in human-computer interaction and usability research: Ready to deliver the promises”. In J. H. Radach & H. Deubel (Eds.), *In the mind's eye: Cognitive and applied aspects of eye movement research*, pp. 573–605, Amsterdam: Elsevier Science, 2003.

[20] M. Camilli, R. Nacchia, M. Terenzi, and F. Di Nocera, F. ASTEF: a simple tool for examining fixations. *Behavior Research Methods*. 2008 May, 40(2):373-82.

[21] Clark, P. J., & Evans, F. C. (1954). Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecolog*y, 35, 445-453.

[22] S. M. Munn, and J.B. Pelz, “3D point-of-regard, position and head orientation from a portable monocular video-based eye tracker”, *Proc. ETRA 2008*, pp. 181-188.

[23] F. Pirri, M. Pizzoli, and A. Rudi, “A general method for the point of regard estimation in 3D space”, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 921-928, 2011.

[24] K. Santner, G. Fritz, L. Paletta, and H. Mayer, “Visual recovery of saliency maps from human attention in 3D environments”, *Proc. ICRA 2013*, pp. 4297-4303.

[25] NaturalPoint, Inc. DBA OptiTrack. [Online] http://www.naturalpoint.com/optitrack/, 2016.

[26] T. Pfeiffer, and P. Renner, „EyeSee3D: a low-cost approach for analyzing mobile 3D eye tracking data using computer vision and augmented reality technology”, *Proc. Eye Tracking Research and Applications* (ETRA), pp. 367-374, 2014.

[27] A. Dini, G. Lodron, and L. Paletta. Random Dot Marker based Positioning Accuracy for Continuous Attention Monitoring in Work Cells, *Proc. OAGM & ARW Joint Workshop on “Vision, Automation & Robotics*”, submitted for review, 2017.

[28] R.D. Wright, and L.M. Ward, *Orienting of Attention*, Oxford University Press, 2008.

[29] M.I. Posner, Orienting of attention, *Quarterly Journal of Experimental Psychology,* Vol. 32 (1): 3–25, 1980.

[30] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhusrt, H. Jarodzka, and J. van de Weijler. *Eye Tracking – A Comprehensive Guide to Methods and Measures*, Oxford University Press, 2011, pp. 187.

[31] Belardinelli, A., Herbort, O., and Butz, M.V. (2015). Goal-oriented gaze strategies afforded by object interaction. *Vision Research*, 106.

[32] L. Paletta, A. Dini, C. Murko, S. Yahyanejad, M. Schwarz, G. Lodron, G., S. Ladstätter, G. Paar, & R. Velik. Towards Real-time Probabilistic Evaluation of Situation Awareness from Human Gaze in Human-Robot Interaction, *Proc. ACM/IEEE HRI 2017*, Vienna.

[33] Fan, R.-E., P.-H. Chen, and C.-J. Lin. Working set selection using second order information for training support vector machines. *Journal of Machine Learning Research*, Vol 6, 2005, pp. 1889–1918.

[34] Scholkopf, B., and A. Smola. Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond, *Adaptive Computation and Machine Learning*. The MIT Press, 2002.

[35] Breiman, L., J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees.* Boca Raton, FL: CRC Press, 1984.