Intelligent Service Robotics (2018) 11:279–288
https://doi.org/10.1007/s11370-018-0253-1

ORIGINAL RESEARCH PAPER

SoTCM: a scene-oriented task complexity metric for gaze-supported teleoperation tasks

Haitham El-Hussieny[1,2] · Samy F. M. Assal[3,4] · Jee-Hwan Ryu[1]

Received: 28 May 2017 / Accepted: 3 May 2018 / Published online: 16 May 2018

Abstract
Recent developments in human–robot interaction (HRI) research have heightened the need to incorporate indirect human signals to implicitly facilitate intuitive human–guided interactions. Eye-gaze has been widely used nowadays as an input interface in multi-modal teleoperation scenarios due to their advantage in revealing human intentions and forthcoming actions. However, to date, there has been no discussion about how the structure of the environment, that the human is interacting with, could affect the complexity of the teleoperation task. In this paper, a new metric named "Scene-oriented Task Complexity Metric" (SoTCM) is proposed to estimate the complexity of a certain scene that is involved in eye-gaze-supported teleoperation tasks. The proposed SoTCM objectively estimates the effort that could be exerted by the human operator in terms of the expected time required to point at all the informative locations retrieved from the scene under discussion. The developed SoTCM depends on both the density and distribution of the informative locations in the scene, while incorporates the eye movement behavior found in the psychology literature. The proposed SoTCM is subjectively validated by using the time-to-complete index in addition to the standard (NASA-TLX) workload measure in eight varying structure scenes. Results confirmed a significant relation between SoTCM and the measured task workload which endorses the applicability of using SoTCM in predicting scene complexities and subsequently the task workload in advance.

Keywords Human–robot interaction · Teleoperation · Eye-gaze tracking · Task workload · Complexity metrics

1 Introduction

Despite the significant advance in involving robots in many real-life applications, their potential has not been fully realized. Once it is programmed, the robot is expected to work in a relatively structured environment. Nowadays, research directions have focused on taking advantage of applying human–robot interaction (HRI) schemes to get benefit from the synergy between the human intelligence and the robot accuracy [8,21]. Robotic teleoperation is identified as one of the examples that could benefit from this kind of synergy between human and robots. It is defined as controlling a robotic manipulator (slave) at a distance by a human operator through a master device to achieve high-level, planning, or cognitive decisions in unstructured or unknown environments [31].

Recent developments in teleoperation have heightened the need for finding new interfacing methods that grant efficient, intuitive, and easier interaction [12]. Recently, a considerable literature has grown up around the use of eye-gaze as an indirect input modality to support classical input interface in multi-modal teleoperation scenarios [2,17,18,27]. In this regard, eye trackers are placed in front of a human to measure the raw gaze movements to predict his/her region of interest (ROI) on a front screen as depicted in Fig. 1. Since the human gaze during teleoperation is mainly directed within the objects required for the task [20], gaze information could be beneficially used by the remote robot to implicitly anticipate the forthcoming task goals. Not only the eye-gaze is considered one of the fastest ways to interact in the daily

✉ Haitham El-Hussieny
haitham.elhussieny@feng.bu.edu.eg

1 School of Mechanical Engineering, Korea University of Technology and Education, Cheonan-City, South Korea

2 Electrical Engineering Department, Faculty of Engineering (at Shoubra), Benha University, Benha, Egypt

3 Mechatronics and Robotics Engineering Department, School of Innovative Design Engineering, Egypt-Japan University of Science and Technology (E-JUST), New Borg El Arab, Egypt

4 Faculty of Engineering, Department of Production Engineering and Mechanical Design, Tanta University, Tanta, Egypt

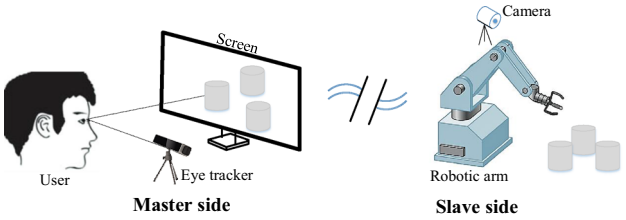


Fig. 1 Block diagram of using eye-gaze modality in robotic teleoperation

active perception tasks, but it also affords valuable information about the user's intent in real time. Furthermore, with little conscious effort, the human can select the ROI faster than using one of the traditional input devices, such as mouse, touch-pads, or haptic devices [37].

Despite the advances in incorporating the eye-gaze modality in teleoperation scenarios, there has been no discussion about how the structure of a certain remote environment could affect the level of task complexity in a gaze-supported teleoperation tasks. Particularly, how both the density and distribution of the probable ROIs in a certain scene could be relate to the human performance in that task. To answer this question, a scene understanding technique is badly needed to grade the given scene according to the level of complexity from the eye-gaze perspective. Subsequently, anticipating the task complexity in advance could be beneficial in more than one scenario. For instance, in shared autonomy teleoperation [1], a task-aware assistance could be offered to the human operator, where the level of assistance is directly proportional to the complexity of the task at hand [11]. In other words, aggressive assistance is preferable on difficult tasks [41], while small assistance is preferable on easy tasks [25]. Moreover, in predicting the human intention from motion, the degree of confidence regarding this prediction is mainly related to the task difficulty [13]. This could be easily anticipated beforehand for the given scenario if we have such a task complexity metric. Finally, in HRI applications, the performance of different interaction algorithms could be fairly compared to each other if we have a fixed task complexity level even if the structure of the environment is not remaining the same.

In this paper, a Scene-oriented Task Complexity Metric (SoTCM) is proposed to objectively assess the complexity of a certain environment from eye-gaze pointing perspective. The motivation behind this is to predict in advance how much complex a certain environment could be if it has been involved in a gaze-supported teleoperation scenario. Subsequently, with the help of SoTCM, the human performance could also be roughly estimated since it is tightly related to the task complexity. The SoTCM takes into account both the density and the distribution of the informative goals retrieved from a certain scene to assess its complexity in terms of the time required to scan all these informative goals. This time is estimated in an off-line manner from the scene configurations with no need to conduct a subjective evaluation to assess the task difficulty.

The remainder of this paper is organized as follow, in Sect. 2, the work related to the task complexity metrics is reviewed, while in Sect. 3 the mathematical background is briefly reviewed. In Sect. 4, the proposed SoTCM is explained and discussed in details. Results of validating the proposed SoTCM are detailed and discussed in Sect. 5. Finally, conclusion is given in Sect. 6.

2 Related work

The concept of "Task Complexity" has previously proposed to define criteria for grading a particular task from easy/simple to difficult/complex in a reasoned way that reflects the amount of workload has paid on fulfilling that task [28]. The term workload is mainly divided into two types: either a physical workload that could be measured quantitatively by one of the physiological metrics or a mental workload that reflects the difference between the task demand and the human capabilities [33]. In general, task complexity has studied in both objective and subjective measures. The former considers the attributes of the task itself [32], while on contrary, the subjective perspective takes both the task attributes and the human capabilities into consideration [36].

Subjective measures of task complexity depend on participants' perceptions regarding the amount of workload imposed on them. NASA Task Load Index (NASA-TLX) is considered one of the standard subjective metrics used to assess the mental workload and perhaps the task complexity [19]. It is based on asking participants a questionnaire that compromises six factors including mental demand, temporal demand, physical demand, performance, effort, and frustration level. A scale from 0 to 100 (simple to difficult) is chosen by the human participant for each metric after completing the specified task to quantify its difficulty. NASA-TLX and other types of evaluation questionnaires, such as Rating Scale Mental Effort (RSME, [44]) are commonly used after the experiment to capture the workload experience and the task complexity from the participant perspective. Despite their wide use in workload assessment, subjective measures suffer from being subjective, where the computed task complexity is mainly influenced by the participant performance that may differ from one to other.

On the other hand, objective evaluations of task complexity are conducted with the help of one physiological metrics such as heart rate variability [6,24] or eye activity measures including blinking, fixations, or saccades rate [30]. Objective metrics can detect changes in mental workload in an online fashion without being influenced by the subjectivity factor as

in subjective metrics. However, they require special types of equipment that are required to be in contact with participants during the time of the experiment. In addition, they could be affected by other environmental factors and conditions preceding the workload measurement [9].

Due to the mentioned limitations of both subjective and objective complexity measures, research has been conducted to find a way to anticipate the task complexity in advance from the attributes of the task itself. For instance, in [10] the complexity of scanning an advertisement is measured based on some of its visual features such as color, luminance, and edges. Scenes with higher variations in their features are rated as complex tasks compared to low variation scenes. In [7], they investigated how the structure of a scene could affect the attention of human eyes. Thus, a saliency map is built to predict the most probable fixations of the eye in a certain scene before starting the experiment and subsequently, the task complexity could be obtained by conducting one of the objective eye-related workload measures. In [3], the regularity of the scene image over fixed regions is proposed as a measure of image complexity. The more the irregularity of a scene, the more random walks the human eyes will make and hence a more complex scene exists. The term regularity is defined as the difference in certain features between the adjacent regions in the scene. Despite their applicability, the mentioned approaches don't take the gaze-based teleoperation task into account where the eye movement behavior should be considered. In this research, a SoTCM is proposed as a task complexity metric that is computed in advance taking into account both the eye movement behavior and the required task to fulfill.

3 Mathematical background

In literature, a well-known predictive model called Fitts' law has been widely used to assess the complexity of pointing and selection tasks [14,29]. It states that, if a pointer is required to move from a certain point, s, to an intended target t, the task index of difficulty ID is formulated as follows

$$\mathrm{ID} = \log_2\left(1 + \frac{D}{W}\right) \quad \text{(bits)} \tag{1}$$

where the target t has a width W and is located at distance D from s. Consequently, the task Movement Time MT is linearly related with ID as follows

$$\mathrm{MT} = a + b \cdot \mathrm{ID} \quad \text{(sec.)} \tag{2}$$

where coefficients a and b depend on the task starting time and the used pointing device.

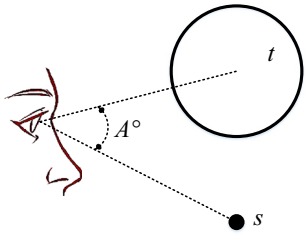


Fig. 2 Illustration of the eye-gaze-based pointing task to reach a target t from point s

However, recent studies have failed to incorporate the Fitts' law in modeling the eye-gaze pointing tasks [38,43]. Psychology literature states that the eye generally moves in a ballistic, sudden, and rapid movement which opposes Fitts' law [4,23]. In this regard, the Carpenter model [5] has been used to predict the time T required for the eye to reach on a certain target t from a point s as follows

$$T = 21\,\mathrm{ms} + 2.2\,\mathrm{ms}/^\circ \cdot \mathrm{A} \quad \text{(ms)} \tag{3}$$

where A is the angular distance between s and t in degrees since the eye moves in a rotational way as illustrated in Fig. 2. This early eye movement to reach a certain object is called the saccadic eye movement, where the goal is to foveate the object of interest rapidly without taking into account the target dimension [35]. In the proposed SoTCM, we mainly rely on this model, with some adaptations, to anticipate the effort that could be exerted on human subjects in terms of the required time to scan all the objects in a certain scene as will be discussed.

4 The proposed scene-oriented task complexity metric (SoTCM)

In this paper, the SoTCM of a certain scene, that is involved in a gaze-supported task, is assumed to be consistent with the effort that human subjects could put in terms of time to point to all of the targets found in that scene. In other words, a scene with a low SoTCM value is identified as low complex scene that all of its objects could be scanned by the human eyes faster than other more complex scenes with high SoTCM values. In this regard, the expected scanning time, with some elaborations to Eq. (3), will be used to estimate the task difficulty in advance for a certain scene based on its objects distribution and density.

4.1 Extraction of the scene probable objects

The probable objects of interest within the scene that the human will interact with should be recognized at first to identify their features that will be used in the calculation of

SoTCM. Extraction of these objects could be done by using one of the *Objects Segmentation* techniques [15,16] to partition a scene into separate clusters based either on depth, color or the structure of objects. Consequently, the geometrical characteristics of each object i could easily be obtained, such as the centroid, width, depth, and height of each object. Object extraction could also be done based on the Saliency Detection [22]. It is responsible for identifying the most interesting regions that draw the human attention in a given scene. These regions are ranked according to size, orientation, color, and illuminations.

In this research, SoTCM depends on the first mentioned extraction approach. The segmentation is done using the tabletop depth segmentation package [26] which is developed under Robotic Operating System (ROS) platform.[1] This package segments a certain scene s that is given in 3D point-cloud (PCL) data [34] into N objects based on depth information. Subsequently, the extracted objects' features are subsequently projected into 2D image space since the human eyes are always pointing to a projected 2D scene. Projection is an essential step if the original scene is captured by a 3D imaging device such as a stereo camera. Otherwise, if the scene is originally given into a projected 2D image, only the segmentation step is applied with no required projection into 2D. At the end, the bounding boxes for the segmented objects are computed based on the point-cloud density of each object, where each of the probable object O_i is defined by a 3-tuple $\{c_i, h_i, w_i\}$ where c_i denotes the center while h_i and w_i represent its associated bounding box height and width in the image 2D space as depicted in Fig. 3. The key assumption behind the proposed SoTCM is that all the extracted objects are static in the scene. Although such an assumption could perhaps limit the applicability of the proposed SoTCM, it could be beneficial for a wide range of applications that the teleoperation task is achieved in a static environment.

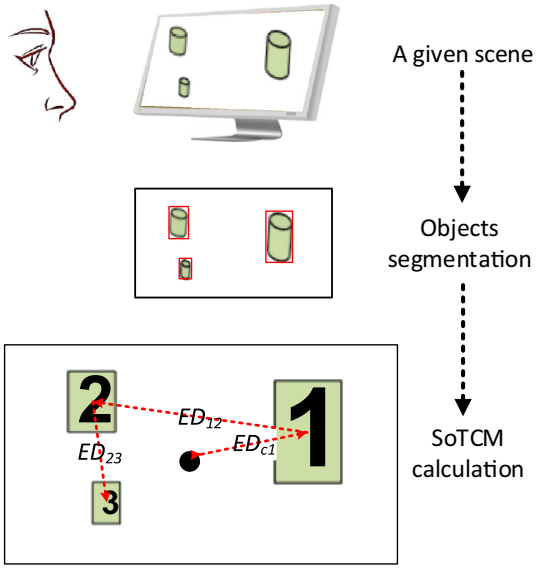


Fig. 3 Work flow for measuring the SoTCM for a given scene

4.2 Calculation of SoTCM

After extracting the probable objects of interest, the SoTCM complexity value, $\eta(s)$, of a certain scene s is measured based on the expected effort that will be exerted on the human subject when interacting with that scene. This effort is anticipated by predicting, in advance, the total pointing time required to accomplish the gaze-driven task based on Carpenter model in Eq. (3). Particularly, the complexity $\eta(s)$ is computed by summing up all the individual pointing times T_i, where $1 \leq i \leq N$, starting from a selected point and passing through all the retrieved informative objects with a specific order as follow

$$\eta(s) = \sum_{i=1}^{N} T_i = 21\,\mathrm{N} + 2.2\,\beta\,E_\mathrm{d} \quad (\mathrm{ms}) \tag{4}$$

where E_d represents the total expected distance in pixels between the N informative objects in the scene, while the constant β is used to convert the angular distance from degrees into pixels in image space. It is defined as the angular resolution in degrees per unit pixel. As depicted in Fig. 4a, the value of β depends on the human distance d from the screen, the viewing angle γ, screen dimensions and screen resolution as follow

$$\beta = 2\,\mathrm{atan2}(s\,\cos(\gamma), 2\,d\, - s\,\sin(\gamma)) \qquad (^\circ/\mathrm{pixel}) \tag{5}$$

where s is the dimension of one pixel on the screen in cm that is basically equals the screen horizontal (or vertical) dimension divided by the horizontal (or vertical) screen resolution. As noted, the value of β in Eq. (5) depends on the human distance and the viewing angle from the screen. However, at human distances ranging from 40 to 80 cm and viewing angles from $-45°$ to $+45°$, the value β has small variations as illustrated by the green area in Fig. 4b with 32-in., 1920×1080 pixels front screen.

4.3 Calculation of the total expected distance, E_d

Indeed, two main concerns should be taken into account while calculating the total expected distance E_d between the retrieved objects in Eq. (5). First, from which target we should start computing the total expected distance, and second, what will be the order we have to follow in summing up

[1] www.ros.org.

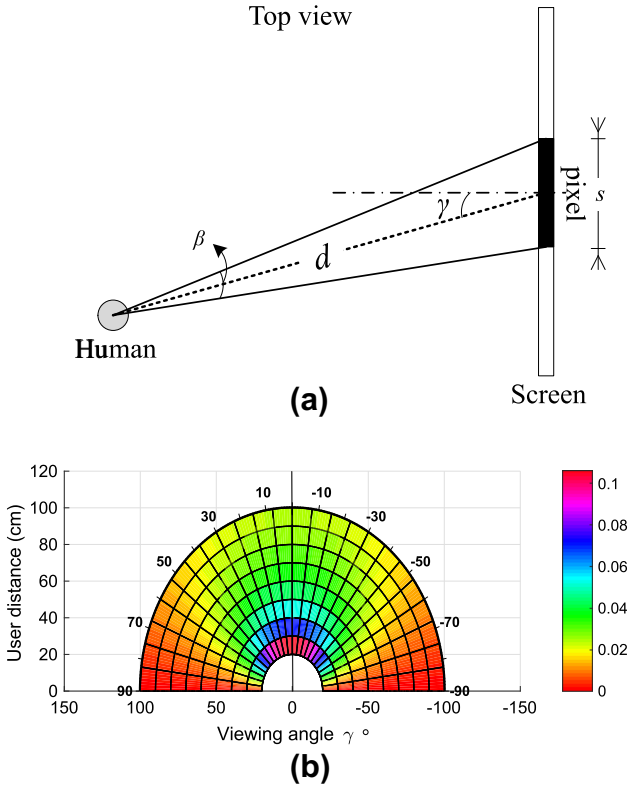


Fig. 4 **a** Calculation of angular resolution β with respect to the distance d and the viewing angle γ. **b** The value of constant β at different d and γ with a 32-in., 1920 × 1080 pixels, front screen

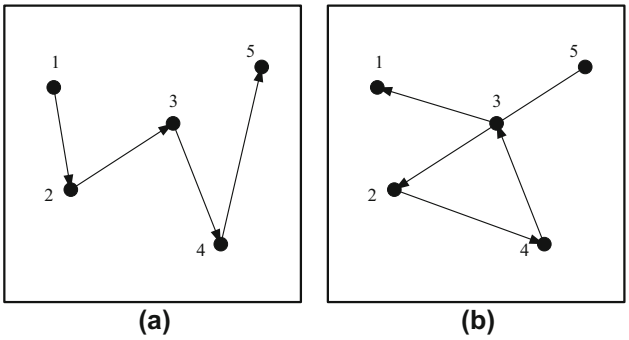


Fig. 5 Effect of computational order on the total expected distance between objects

these distances between objects. For example, if we have in a certain scene five retrieved objects $O_1, O_2, \ldots O_5$ that are distributed in the projected 2D image. Then, summing up the distances starting from O_1 then O_2 and ending up with O_5 as depicted in Fig. 5a will be different if we start computing from O_5, O_2, O_4, O_3 and finally O_1 as shown in Fig. 5b.

The psychological literature of the eye-gaze movement has been followed in this research to address the two mentioned concerns. First, it is reported that the middle of a certain scene is considered as the first point the human is generally looking at during his/her daily active perception tasks [42]. Thus, the center of the scene under discussion is selected as the first starting point to compute the total pointing time. To cope with the ordering ambiguity, the proposed SoTCM follows the order that the human gaze generally follows when looking at a certain scene. The literature states that humans generally shift their gaze toward the more salient regions at first then the less and less. This saliency is mainly depending on the object's features such as size, color, illumination and orientation [39]. In this research, the size feature of the retrieved object in terms of the bounding box dimensions is only considered with the potential to incorporate more saliency features in the future work. Thus, summing up the expected distances E_d for a certain scene starts from the center of that scene then the biggest retrieved object then smaller and smallest as illustrated in Fig. 3. Fortunately, the used tabletop segmentation package segments the given scene in the same order, based on the dimensions of objects in the scene. Hence, the retrieved objects are automatically labeled according to their size and no further sorting step is required to calculate the total expected distance. Thus, to conclude, in Eq. (4), the total expected distance E_d in pixels could be calculated as follows:

$$E_\mathrm{d} = \mathrm{ED}_{c1} + \sum_{i=1, j=i+1}^{N-1} \mathrm{ED}_{ij}. \tag{6}$$

where ED_{ij} is the expected distance between centers of the objects i and j following the mentioned computational order, while ED_{c1} is the expected distance from the middle point c of the scene toward the first object ($i = 1$). The calculation of these distances between objects does not take into account the dimension of objects themselves as mention by the Carpenter model (3). However, it should be noted that many uncertainty sources could exist in retrieving the goals in a given scene. These sources might be due to the camera calibration error, projection model approximation, or the segmentation process itself. Thus, the total expected distance, and subsequently the task complexity, are measured taking into account this kind of uncertainty in the objects locations c_i. Thus, the expected distance ED_{ij} in Eq. (6) is measured in pixels with an additive Gaussian distribution representing the uncertainty in objects location as follow [40],

$$\mathrm{ED}_{ij} = \sqrt{d_{ij}^2 + 0.25 \cdot \Sigma_{ij}} \tag{7}$$

$$d_{ij}^2 = (c_{x_i} - c_{x_j})^2 + (c_{y_i} - c_{y_j})^2 \tag{8}$$

$$\Sigma_{ij} = \sigma_{x_i}^2 + \sigma_{y_i}^2 + \sigma_{x_j}^2 + \sigma_{y_j}^2 \tag{9}$$

where object i is characterized by the mean $\mu_i = [c_{x_i}, c_{x_i}] \in \mathbb{R}^2$ and both the horizontal and vertical variances, σ_{x_i} and σ_y in the modeled uncertainty. With reference to Fig. 6, the

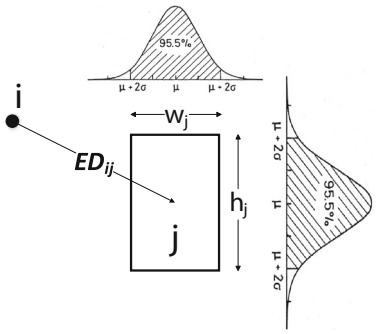


Fig. 6 Measurement of the expected distance ED_{ij} from a point i to uncertain object j with a Gaussian distribution

bounding box of each object is assumed to be covered by around 95.5% of the Gaussian distribution. Thus, the variances could be calculated as follows:

$$\sigma_{x_i} = 2w_i/4, \qquad \sigma_{y_i} = 2h_i/4. \tag{10}$$

5 Results and discussion

In the following experimental evaluations, we demonstrate the validity and sensitivity of our proposed SoTCM with different scenarios varying in their structures.

5.1 Effect of scene structure on SoTCM

In the first set of experiments, a Kinect sensor is placed in front of a table with some standing objects to capture RGBD images for the scenes involved in experiments. To vary the scene structure with ease, different paper-cups are put on the table and used as probable objects of interest. At first, the effect of the object size on the developed SoTCM is studied by stacking cups over each other as depicted in Table 1 (first row). For each scene, the computational order is highlighted in blue arrow lines while both the image center and objects centers are highlighted in green and red colors respectively. As highlighted, decreasing the object size (left to right) quietly affects SoTCM values labeled under each scene image. In addition, increasing the number of objects in the scene, Table 1 (second row), increases the SoTCM value rapidly as shown. The SoTCM values are shown under each scene used as illustrated in Table 1. Additionally, Fig. 7 shows the measured SoTCM for some real-world scenario selected from the Willow Garage object recognition dataset, where some of the daily-life objects are incorporated.

To discuss the obtained results, the measured scene complexity from the gaze pointing perspective depends on both the density and distribution of the informative objects in a certain scene. As obtained in the first experiment, the SoTCM is quietly affected by the size of objects in the scene which could be interpreted due to two different sources. First, in Eq. (10), increasing the size of an object i will increase the variances, σ_{x_i} and σ_{y_i}, around the object bounding box and subsequently the complexity will increase. However, increasing the object size will also change its center position and therefore its distance d_{ij} from either the image center or the subsequent object j and hence the complexity will be affected. Therefore, increasing the object size has different effects on the total expected distance and hence the scene complexity as shown in Table 1. It is also shown in the second row of Table 1 that the scene complexity is directly affected by the number of objects in the scene. This is reasonable as the SoTCM is directly proportional to N where increasing the number of objects will require more effort to point to all of them and hence the task complexity will increase.

It is worth to mention that although the proposed SoTCM is an absolute measure in milliseconds, different scenes in a certain teleoperation scenario could be evaluated in their complexities relative to each other by use of max–min normalization. This will help to sort the involved scenes according to their complexity values ranging from zero to one or from easiest to complex.

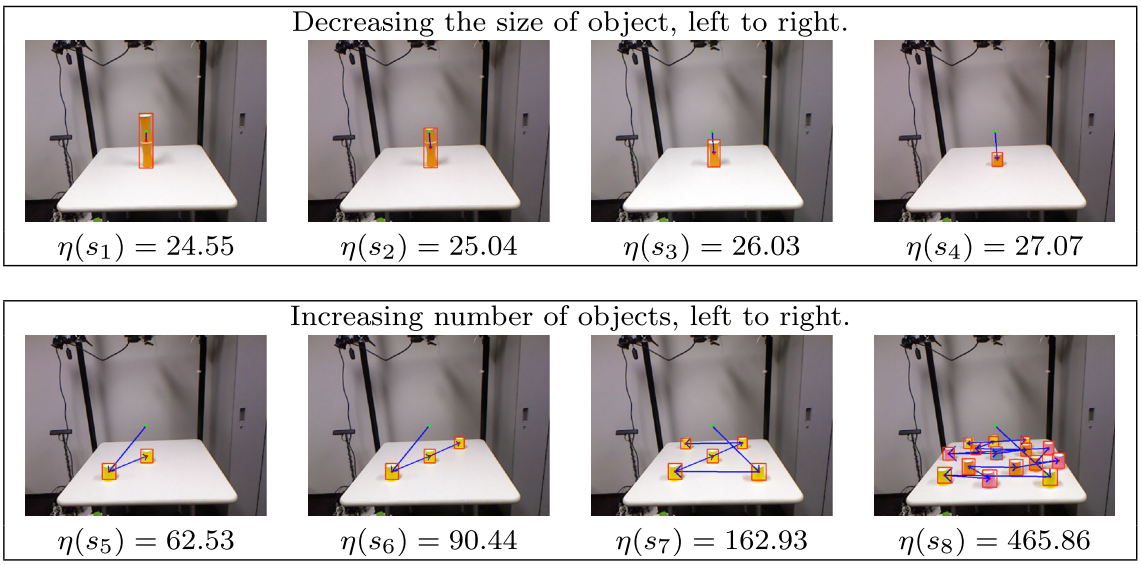
5.2 Validation of SoTCM

To check the validity, another set of experiments have been conducted to show the relation between the proposed SoTCM and the task demand in terms in both objective and subjective perspectives.

In the objective perspective, the actual scene scanning times that the human subject needs to point to all objects in each of the eight available scenes (Table 1) is measured and plotted versus the in advance computed SoTCM values, Fig. 8. For each scene, one of eight subjects were asked to point to all the objects in the given scene in any order he/she would prefer. The experiment is repeated twice for each scene per each subject with a total of 16 trials per scene. An Eyetribe[2] eye tracker is placed in front of human subjects to retrieve the gaze movement at 60 Hz rate. Subjects were looking at a 32-in., 1920 × 1080 pixels, screen while pointing to the intended goals at distance $d = 60$ cm and viewing angle $\gamma \approx 0$ from the from the screen. The object that is already reached by the human eye is highlighted with a bounding circle as a kind of visual feedback for the subject. The object is considered as reached if the subject fixed his/her eye on the intended object for at least the decision time (100 ms) within its bounding box. Due to the difference in SoTCM values between the first and the last four scenarios, the relation between the actual scene scanning time and the SoTCM is divided into two sets. In Fig. 8a, the relation is plotted for the first four scenes with complexities $\eta(s)$ ranging from

[2] www.theeyetribe.com.

Table 1 SoTCM values for different scenes



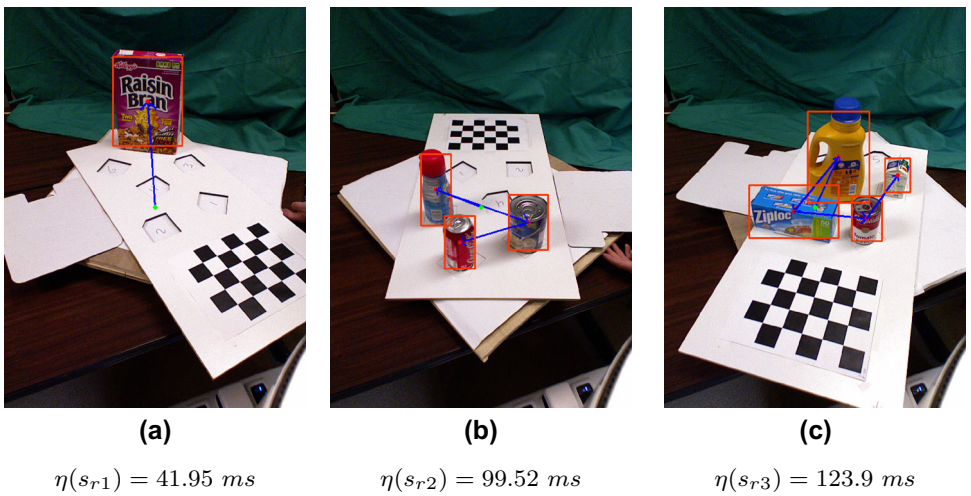


Fig. 7 SoTCM of scenes with different real-world objects

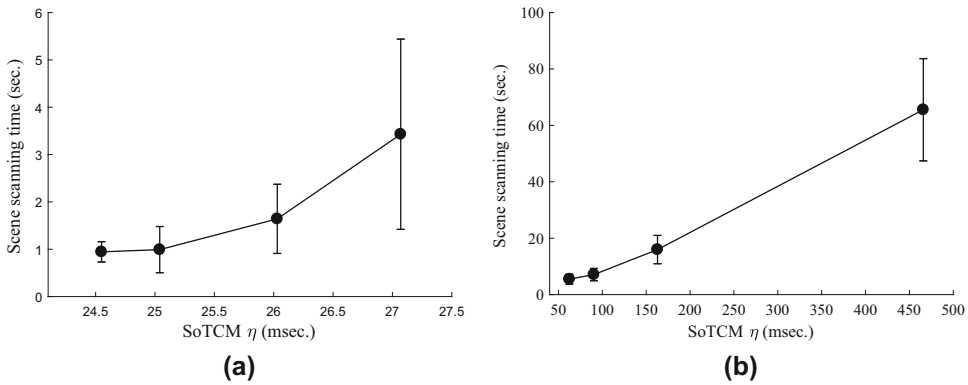


Fig. 8 Relation between the developed SoTCM and the measured scanning time for the first four scenarios in **a** and for the second four scenarios in **b**

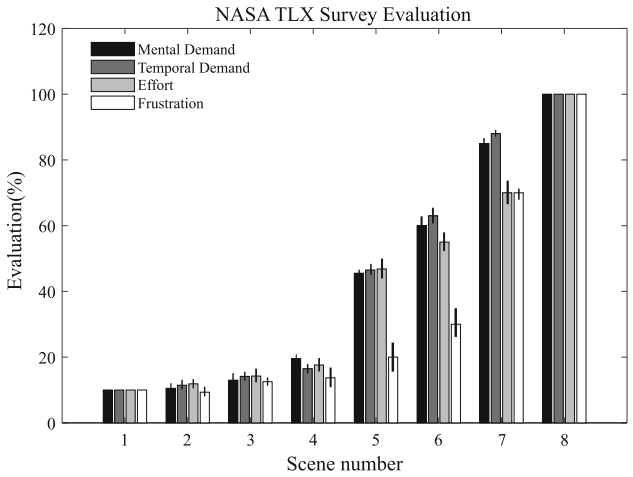


Fig. 9 Computation of NASA Task Load Index (TLX) for each of the available eight scenarios in terms of: mental demand, temporal demand, effort and frustration

24.55 to 2.07, while in Fig. 8b, the relation is plotted for other four scenes with complexities $\eta(s)$ ranging from 62.53 to 465.86. As observed, the proposed SoTCM shows a significant relationship with the actual scene scanning time required by human subjects to point at all objects in that scene. The obtained results support our working hypothesis that the proposed SoTCM is a representation for the required effort to be exerted on a certain scene with the advantage that it is calculated in advance from the scene characteristics. However, the difference between the measured scene scanning time (in seconds) and the computed SoTCM (in milliseconds) is due to the fact that the decision time and the delay time in the gaze filtration process are included in the measure scanning time. On the other hand, the SoTCM in terms of time represents the abstract time with no further processing time.

Furthermore, the relation between the proposed SoTCM and the perceived cognitive load is subjectively evaluated by incorporating the standard NASA task load index (NASA-TLX) [19] for the eight mentioned scenes. Once a subject has completed the two trials previously mentioned for each scene, he/she was asked to rate how much workload they experience (from 0 for 'low' to 100 for 'high'). Each participant is asked four questions to rate the: mental load, temporal load, effort, and frustration scale of the eight pointing scenes. To make the human subject compare relatively between the eight scenes, the four NASA-TLX scales are explicitly set to 10 for the first scenario s_1 in Table 1, while they are set to 100 for the last scenario s_8. Responses to the NASA-TLX questionnaire are found to be consistent with the proposed SoTCM as highlighted in Fig. 9 where the horizontal axis represents scenes from 1 to 8 arranged in ascendant according to their SoTCM values. As noted, no significant difference between the four NASA-TLX rates of the first four scenes since as mentioned their structure are relatively comparable to each other. However, significant differences are found in those rates of the other four scenes due to theirs significant structure difference.

5.3 SoTCM sensitivity

To figure out the effect of the user distance and the viewing angle on the proposed SoTCM, the complexity value η is obtained for the scenario shown in Fig. 10b for different d and γ. The user distances are chosen as $d = [30, 60, 90]$ cm while the viewing angle is varied from $-50°$ to $+50°$ and for each pair the complexity $\eta(d, \gamma)$ is obtained as shown in Fig. 10a. For a small distance from the screen, η is greatly affected by the value of the viewing angle. In contrary, in large distances from the screen, the viewing angle has a small effect on η especially in angles ranging from $-45°$ to $+45°$. Since the objective of the proposed SoTCM is to compare among the different available scenarios in terms of task complexity, the values of η could be obtained for the scenarios under investigation while explicitly mentioning the chosen d and γ.

In another set of experiment, the effect of changing the computational order over the proposed SoTCM is investigated. Two different scenes are involved with three kinds of computational orders are chosen as shown in Table 2. The first calculation order is the default order found by the segmentation package as mentioned earlier according to the object size while the other two orders are chosen arbitrarily. As depicted, the order of SoTCM calculation slightly affects the complexity values by order of around 15 ms for both scenarios. As mentioned, the calculation path is managed by the segmentation package where the calculation order depends on the size of the retrieved informative objects. However, more accurate results could be obtained in future if the saliency map is incorporated to predict the eye-gaze path according to not only the objects size but also colors, orientation, and the context.

6 Conclusion

In this paper, we examine the effect of the structure of a certain remote environment on the task complexity of gaze-supported multi-modal teleoperation tasks. In this regard, a new task complexity metric called "Scene-oriented Task Complexity Metric" (SoTCM) is proposed to assess in advance how the scene structure could affect the complexity of the pointing task and hence the human performance. The proposed metric is taking into account the characteristics of the eye-gaze movement behavior found in the physiological literature. The SoTCM values for eight different environment scenes differing in their structure are obtained to show how both density and distribution of scene objects could contribute to the total expected task complexity. Moreover,

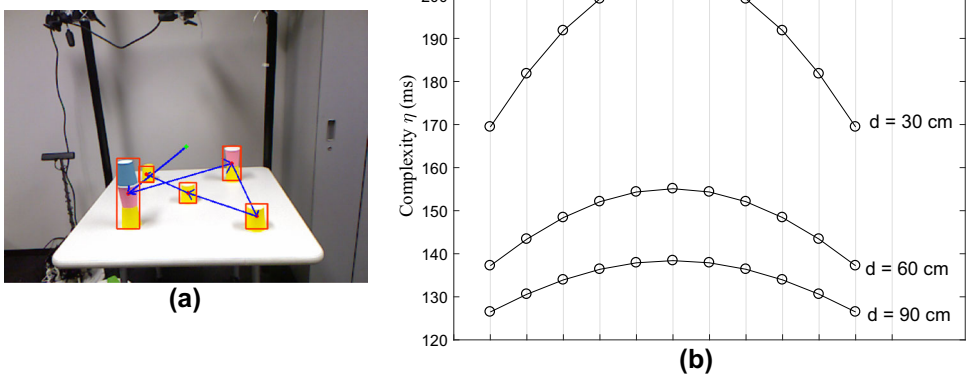
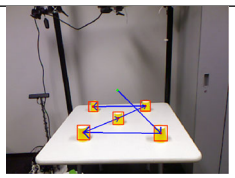
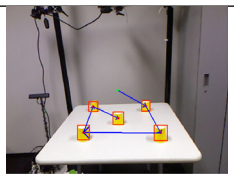
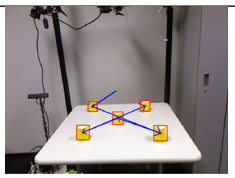
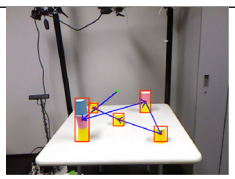
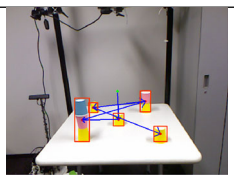
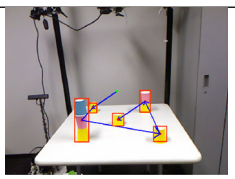


Fig. 10 **a** Calculation of SoTCM η variation with respect to the change in distance d and the viewing angle γ from the screen

Table 2 SoTCM values for different path order

Path order	$1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5$	$4 \rightarrow 1 \rightarrow 2 \rightarrow 5 \rightarrow 3$	$5 \rightarrow 1 \rightarrow 3 \rightarrow 2 \rightarrow 4$
Scene #1			
SoTCM	$\eta(s_{11}) = 163.07$	$\eta(s_{12}) = 149.04$	$\eta(s_{13}) = 161.48$
Scene #2			
SoTCM	$\eta(s_{21}) = 155.01$	$\eta(s_{22}) = 163.02$	$\eta(s_{23}) = 149.70$

results show evidence of the connection between the proposed SoTCM and the task difficulty has been shown in both objective and subjective perspectives when compared with the actual pointing time and the standard NASA-TLX, respectively, in the given tasks. The proposed complexity metric should prove usefulness in grading gaze-supported teleoperation tasks in both shared autonomy and intention prediction contexts. In our future research, we plan to incorporate more saliency feature such as color, illumination, and orientation in the detection of the computational order not only the size feature. Another goal is to elaborate our proposed method to estimate the complexity of dynamic tasks where the scene objects are not stationary as assumed.

Acknowledgements This research was partially supported by the Civil-Military Technology Cooperation Program (15-CM-RB-09) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF- 2016R1E1A1A02921594). The first author is also supported by a Post-doctoral fellowship from Korea University of Education and Technology (KOREATECH) which is gratefully acknowledged.

References

1. Anderson RJ (1996) Autonomous, teleoperated, and shared control of robot systems. In: Proceedings of the 1996 IEEE international conference on robotics and automation, 1996, IEEE, vol 3, pp 2025–2032
2. Aronson RM, Santini T, Kübler TC, Kasneci E, Srinivasa S, Admoni H (2018) Eye-hand behavior in human-robot shared manipulation. In: Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction, ACM, pp 4–13
3. Bonev B, Chuang L, Escolano F (2013) How do image complexity, task demands and looking biases influence human gaze behavior? Pattern Recogn Lett 34(7):723–730
4. Bruce V, Green PR, Georgeson MA (2003) Visual perception: physiology, psychology, and ecology. Psychology Press, London
5. Carpenter RH (1988) Movements of the eyes, 2nd edn. Pion Limited, London

6. Castaldo R, Montesinos L, Wan TS, Serban A, Massaro S, Pecchia L (2017) Heart rate variability analysis and performance during a repeated mental workload task. In: EMBEC & NBC 2017, Springer, Berlin, pp 69–72
7. Cheng M, Mitra NJ, Huang X, Torr PH, Hu S (2015) Global contrast based salient region detection. IEEE Trans Pattern Anal Mach Intell 37(3):569–582
8. Dautenhahn K (2007) Methodology & themes of human-robot interaction: a growing research field. Int J Adv Rob Syst 4(1):15
9. De Waard D (1996) The measurement of drivers' mental workload. Groningen University, Traffic Research Center, Groningen
10. Donderi DC (2006) Visual complexity: a review. Psychol Bull 132(1):73
11. Dragan AD, Srinivasa SS (2013) A policy-blending formalism for shared control. Int J Robot Res 32(7):790–805
12. Drewes H (2010) Eye gaze tracking for human computer interaction. Ph.D. thesis, lmu
13. El-Hussieny H, Assal SF, Abouelsoud A, Megahed SM (2015) A novel intention prediction strategy for a shared control telemanipulation system in unknown environments. In: 2015 IEEE international conference on mechatronics (ICM), IEEE, pp 204–209
14. Fitts PM (1954) The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol 47(6):381
15. Freixenet J, Muñoz X, Raba D, Martí J, Cufí X (2002) Yet another survey on image segmentation: region and boundary information integration. In: Computer VisionECCV 2002, Springer, Berlin, pp 408–422
16. Fu KS, Mui J (1981) A survey on image segmentation. Pattern Recogn 13(1):3–16
17. Gomes J, Marques F, Lourenço A, Mendonça R, Santana P, Barata J (2016) Gaze-directed telemetry in high latency wireless communications: the case of robot teleoperation. In: IECON 2016-42nd annual conference of the IEEE industrial electronics society, IEEE, pp 704–709
18. Hansen JP, Alapetite A, MacKenzie IS, Møllenbach E (2014) The use of gaze to control drones. In: Proceedings of the symposium on eye tracking research and applications, ACM, pp 27–34
19. Hart SG, Staveland LE (1988) Development of nasa-tlx (task load index): results of empirical and theoretical research. Adv Psychol 52:139–183
20. Hayhoe M, Ballard D (2005) Eye movements in natural behavior. Trends Cognit Sci 9(4):188–194
21. Heyer C (2010) Human-robot interaction and future industrial robotics applications. In: 2010 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 4749–4754. IEEE
22. Hou X, Zhang L (2007) Saliency detection: a spectral residual approach. In: IEEE conference on computer vision and pattern recognition, 2007. CVPR'07, IEEE, pp 1–8
23. Jacob RJ (1995) Eye tracking in advanced interface design. Virtual Environ Adv Interface Des pp 258–288
24. Jorna PG (1992) Spectral analysis of heart rate and psychological state: a review of its validity as a workload index. Biol Psychol 34(2):237–257
25. Kim DJ, Hazlett-Knudsen R, Culver-Godfrey H, Rucks G, Cunningham T, Portee D, Bricout J, Wang Z, Behal A (2012) How autonomy impacts performance and satisfaction: results from a study with spinal cord injured subjects using an assistive robot. IEEE Trans Syst, Man Cybernet, Part A: Syst Hum 42(1):2–14
26. Kramer J, Burrus N, Echtler F, Daniel H.C, Parker M (2012) Object modeling and detection. In: Hacking the Kinect, Springer, Berlin, pp 173–206
27. Latif HO, Sherkat N, Lotfi A (2009) Teleoperation through eye gaze (telegaze): a multimodal approach. In: 2009 IEEE international conference on robotics and biomimetics (ROBIO), IEEE, pp 711–716
28. Liu P, Li Z (2011) Toward understanding the relationship between task complexity and task performance. In: Internationalization, design and global development, pp 192–200
29. MacKenzie IS (1992) Fitts' law as a research and design tool in human–computer interaction. Hum–Comput Interact 7(1):91–139
30. Marquart G, Cabrall C, de Winter J (2015) Review of eye-related measures of drivers mental workload. Proced Manuf 3:2854–2861
31. Niemeyer G, Preusche C, Hirzinger G (2008) Telerobotics. In: Springer handbook of robotics, Springer, Berlin, pp 741–757
32. Rouse WB, Rouse SH (1979) Measures of complexity of fault diagnosis tasks. IEEE Trans Syst Man Cybernet 9(11):720–727
33. Rubio S, Díaz E, Martín J, Puente JM (2004) Evaluation of subjective mental workload: a comparison of swat, nasa-tlx, and workload profile methods. Appl Psychol 53(1):61–86
34. Rusu RB, Cousins S (2011) 3d is here: point cloud library (pcl). In: 2011 IEEE international conference on robotics and automation (ICRA), IEEE, pp 1–4
35. Saeb S, Weber C, Triesch J (2011) Learning the optimal control of coordinated eye and head movements. PLoS Comput Biol 7(11):e1002253
36. Schwab DP, Cummings L (1976) A theoretical analysis of the impact of task scope on employee performance. Acad Manag Rev 1(2):23–35
37. Sibert LE, Jacob RJ (2000) Evaluation of eye gaze interaction. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, New York, pp 281–288
38. Vertegaal R (2008) A fitts law comparison of eye tracking and manual input in the selection of visual targets. In: Proceedings of the 10th international conference on multimodal interfaces, ACM, New York, pp 241–248
39. Wolfe JM, Horowitz TS (2004) What attributes guide the deployment of visual attention and how do they do it? Nat Rev Neurosci 5(6):495–501
40. Xiao L, Hung E (2007) An efficient distance calculation method for uncertain objects. In: CIDM 2007. IEEE symposium on computational intelligence and data mining, 2007, IEEE, pp 10–17
41. You E, Hauser K (2012) Assisted teleoperation strategies for aggressively controlling a robot arm with 2d input. In: Robotics: science and systems, vol 7, p 354
42. Zhang L, Tong MH, Marks TK, Shan H, Cottrell WG (2008) Sun: a bayesian framework for saliency using natural statistics. J Vis 8(7):32–32
43. Zhang X, MacKenzie IS (2007) Evaluating eye tracking with iso 9241-part 9. In: HCI intelligent multimodal interaction environments human–computer interaction, Springer, Berlin, pp 779–788
44. Zijlstra FRH (1993) Efficiency in work behaviour: a design approach for modern tools. Delft University of Technology, Delft