ORIGINAL ARTICLE

Using visual feedback to improve hand movement accuracy in confined-occluded spaces in virtual reality

Yu Wang[1] · Ziran Hu[1] · Shouwen Yao[1] · Hui Liu[1]

Abstract

Accurate and informative hand-object collision feedback is of vital importance for hand manipulation in virtual reality (VR). However, to our best knowledge, the hand movement performance in fully-occluded and confined VR spaces under visual collision feedback is still under investigation. In this paper, we firstly studied the effects of several popular visual feedback of hand-object collision on hand movement performance. To test the effects, we conducted a within-subject user study (n=18) using a target-reaching task in a confined box. Results indicated that users had the best task performance with see-through visualization, and the most accurate movement with the hybrid of proximity-based gradation and deformation. By further analysis, we concluded that the integration of see-through visualization and proximity-based visual cue could be the best compromise between the speed and accuracy for hand movement in the enclosed VR space. On the basis, we designed a visual collision feedback based on projector decal,which incorporates the advantages of see-through and color gradation. In the end, we present demos of potential usage of the proposed visual cue.

Keywords Visual collision feedback · Occluded interaction · 3D occlusion management · Hand movement performance · See-through visualization

1 Introduction

Virtual reality (VR) provided users with a vivid virtual environment where users can observe the virtual world in stereo vision and interact with virtual objects in natural behavior. However, due to the absence of haptic feedback, users perform hand manipulation tasks in VR mainly based on proprioception and visuo-motor synchronization [1–3], which seriously limits the hand manipulation performance in VR [4,5]. In the confined-occluded VR spaces, however, the hand manipulation tasks become more challenging. On one hand, users cannot see their hands movement further degrading the manipulation accuracy [6,7]. On the other hand, the user needs to carefully control the hand movement to avoid the collisions with spatial boundaries to ensure correctness, which in turn requires high precision of the movement of the hand. This issue limits the applications of VR in simulating fine hand manipulation tasks under occlusion conditions, such as VR-aided assessment of assembly task in complex gearbox, or lottery game of finding and catching blocked objects from an unknown treasure box.

Methods of providing collision feedback in virtual environments have been extensively studied. The physical feedback method [8–12] uses mechanical arm or vibro-tactile suit to provide collision feedback actively, which has good accuracy and intuitiveness [9,13]. However, the methods mostly rely on expensive additional devices which could be cumbersome to wear in practice, and lacks mobility. In addition, the use of haptic feedback devices conflicts with the ongoing trend towards natural gesture interaction based on marker-less full-body motion capture [14,15]. The visually enhanced methods, which provides collision feedback via various visualization technologies and visual cues, are more flexible than the physical feedback solutions in applications. Previous works [16–19] have illustrated that the visual cues can efficiently improve task performance and sense of reality in virtual environments, in some cases, visual feedback are less ambiguous than vibrotactile feedback [13]. However, most relevant studies focus on improving manipulation performance in unobstructed and unrestricted view, and few have conducted comprehensive studies on the effects of visual feedback in confined-occluded VR space. This has formed

✉ Shouwen Yao
  armysw@bit.edu.cn

[1] Beijing Institute of Technology, Beijing, China

gaps in current research, for example, how does the spatial information of collision encoded in visual feedback affect the hand movement performance in confined and occluded VR spaces?

To explore this gap, in this paper, we first reviewed related works to identify the informativeness levels of visual feedback used in hand-object interaction in VR. Then, we conducted a within-subject user study to compare the effects of the visual feedback on hand movement performance using a target-reaching task in a confined box, followed by the analysis of the experiment results. Finally, we discuss the findings and proposed a visual collision feedback based on projector decal, which reaps the advantages of see-through visualization and proximity-based color gradation. In the end, we presented demos of the application of the proposed visual cue in confined-occluded VR interaction cases. Our results provide reference for understanding hand movement behavior in confined-occluded spaces and designing effective visual collision feedback.

2 Related work

2.1 Sensory feedback in aiming movement

The aiming movement can be defined as a two-phase model which contains a free movement phase and a fine positioning phase [4,20]. Human complete the aiming movement using multiple modalities of sensory feedback, including visual feedback, proprioception and somatosensory (haptic) feedback [3]. Visual feedback is believed to play the major role of building the spatial plan of movements, while the proprioception provides sensory input for the translation of the kinematic plan into variables corresponding to forces and torques for the execution of movement [21]. Somatosensory feedback plays a role of controlling the grip force [22], it also works as a confirmation signal of whether successfully reach the target [7], and a receptor of the location information of the touched objects [23]. The somatosensory feedback as well as other modalities of feedback, such as auditory feedback, provide less spatial information of target compared to the visual feedback [13,23]. They usually work as supplementary sensory sources when the visual feedback is unavailable.

Different forms of sensory feedback integration influence the performance of the aiming movement. Visuo-motor colocation is the basic sensory integration for aiming movement, although it has less influence on the movement performance in virtual environment [2,21]. Integration of haptic feedback usually increases the movement speed, this might be attributed to the fast processing speed of human brain on haptic feedback [7,24]. Whereas, lack of visual feedback often leads to a lower movement accuracy [1,6,7,13]. Thus, integrating haptic and visual feedback could be the best trade-off between the speed and accuracy in aiming movement [7].

2.2 Spatial information in hand-object interaction

To understand environments and achieve the interaction goal properly, one should access sufficient spatial knowledge from the environments. Therefore, it is necessary to specify in advance what type of spatial knowledge human need to acquire to understand collision events or to perform fine aiming movements in enclosed spaces. Waller et al. [23] presented that human understand the spatial relations with the surrounding environments based on two types of sensory information: (a) External information about the nature of one's environment such as the environment layout including relative directions, distance as well as scale; and (b) Internal information about the status of one's body or effectors, e.g., position, orientation, and movement of musculature. Fan et al. [25] proposed an HOR pattern of *hand-object, object-reference, hand, object, reference* to interpret human's intention in atomic hand-object interaction. The pattern includes not only the state descriptors of the pose and appearance of hands, objects, and references, but also position-related interactions between hands and objects as well as objects and references. Researches about aiming movement have shown that the visibility of the target and the movement of the effector affect movement efficiency [1,6,7]. Niklas et al. [26] suggested that the visual tasks are influenced by three geometrical properties of the environment which refers to the spatial interaction between objects, the object density and the detail level of individual objects. They also indicated that the 3D occlusion management technology should consider four properties of the targets which are appearance, depth, geometry, and location. For the spatial knowledge needed to understand collisions, in addition to the movement state of the objects about to collide [27], the location of the impending collision can also help predict and avoid collisions [28]. The direction, depth, force and effort of the collisions contribute to understand the status of the ongoing collisions [16].

2.3 Visual feedback for collision

Contact-based coloring cue changes the global color of the affected object, and is widely used to indicate the events of the contact or grasp [17]. Proximity-based coloring cue changes the color of object (local or global) with a color gradation according to the target-effector proximity [16,18]. The local highlight simply adjust the lighting intensity at the collision point according to the collision status [16]. These two cues are thought to have better performance to indicate contact region and the distance [16]. Deformable glyphs were proved to be an efficient cue to hint the contact depth and

force by attaching a deformable indicator at the collision point [16,29]. Virtual shadow seems to be the most natural way to implicitly inform the collision, but it is not practical in the enclosed environment [30]. See-through (translucent) visualization is also thought to be an intuitive solution for displaying information in the occluded spaces [31]. Ariza et al. [32] suggested that the binary proximity-based feedback outperform the continuous feedback in terms of faster movement speed and overall higher throughput in 3D selection task.

The previous researches have provided sufficient evidences of the effects of visual feedback on improving hand manipulation performance in visible VR spaces. However, their works did not reveal the effects of visual feedback in the confined-occluded spaces. Sreng's work [16] is the most relevant, however, they only studied the subjective preference, did not reveal the hand movement behavior in the confined-occluded spaces with the given visual feedback. Bloomfield et al [8] designed an experiment to examine the effects of visual cues and vibrotactile cues on manipulation accuracy in confined space, however, their experiment only involved the local highlight, which cannot represent all forms of visual cues. Thus, it is necessary to study the effects of the visual collision feedback on hand movement performance in confined-occluded VR spaces if we would like to explore more challenging hand interaction tasks in VR.

3 Informativeness of visual feedback

In order to better understand the influence of spatial information of collision encoded in different visual feedback on hand motor performance, it is necessary to know in advance how informative the feedback is and what information can be conveyed by the feedback. We referred to previous work [16,23,25] and designed a empirical metric containing two dimensions of detail and understanding to describe the informativeness of visual feedback. We excluded the velocity and position of hand and object from detail dimension because the motion information of hand is involved in the proprioception of hand movement, while the motion information of object highly overlaps with the given elements of detail dimension. Moreover, we added an element of *direction* in depth dimension based on Sreng's scale [16] since we consider this will help understand the collision situation in 3D environment. The definitions of the metric dimensions are presented below.

- **Detail**: Describes the granularity of the spatial information encoded in visual feedback, which consists of the following elements.
  - *Event*: Whether a hand collides with a object.
  - *Location*: The location of an ongoing collision or an imminent collision.
  - *Distance*: The distance between hand and object in an imminent collision or the penetration depth of an ongoing collision.
  - *Direction*: The direction of an ongoing collision or an imminent collision.
  - *Physics*: The force and effort between hand and object in collision.
- **Understanding**: Describes how easily the spatial information can be understood by users, which consists of the following levels:
  - *Hybrid*: There are multiple symbols or metaphors representing the same detail element.
  - *Easy*: The information is displayed by a consistent symbol, e.g., using a line to express distance or an arrow to express direction.
  - *Medium*: The information is displayed by a relevant metaphor, e.g., using number or color gradation to represent distance, which users can understand through very little reasoning.
  - *Hard*: There is no enhanced visual feedback, but users can identify the information through instinct or lots of reasoning, e.g., binocular vision or translational parallax.
  - *None*: Users cannot acquire any information from the visual feedback.

The detail dimension refers to the number of spatial information details retained in visual feedback. The understanding dimension is an ordinal dimension which describes the intuition of a visual feedback in every detail dimension. At last, the informativeness level of a visual feedback is the sum of all details multiplied by their corresponding understanding level. In general, a visual feedback with more detail elements and higher understanding level is considered as richer in informativeness, and vice versa. We used the proposed metric to describe the informativeness of a visual collision feedback in the following experiment.

4 Experiment

Since the aiming movement of hand under different occluded and feedback conditions have been comprehensively studied [6,7,32], in this work, we want to understand the relations between hand movement behavior in confined-occluded space and the visual feedback with different level of informativeness. We cared about the movement efficiency, i.e. speed and accuracy, since they are important to the manipulation task performance. In a straightforward thought, a rich informativeness visual feedback should be superior in movement performance. Thus, we put forward the following hypotheses:

– H1: The visual collision feedback with rich informativeness can make the hand movement more quick in the confined-occluded space.
– H2: The visual collision feedback with rich informativeness can make the hand movement more accurate in the confined-occluded space.

Moreover, we are also interested in some exploratory questions without prior hypotheses to gain inspirations of design improvements and future research. The questions are listed as follows:

In confined-occluded VR spaces:
- RQ3: What are the effects of spatial information on hand movement performance in the detail dimension?
- RQ4: What are the effects of spatial information on hand movement performance in the understanding dimension?
- RQ5: Whether high informativeness visual feedback can raise a feeling of visuo-tactile synesthesia?

To verify the hypotheses and exploratory questions, we conducted a within-user study which tested five visual feedback that are commonly-used in VR. The included visual feedback were carefully selected according to their informativeness. The details of the experiment are presented below.

4.1 Experiment scene and task

Our experiment task is to reach a ball inside a confined box. The experiment scene is shown in Fig. 1. The length of the box is 400mm to ensure that the subject's arm must be fully inserted into the box to reach the ball. The size of the box's opening is 100 x 100mm, which is slightly larger than the average width of a male adult's palm [33]. The box hovers in front of subject with the opening facing to the right side of the subject. A ball with a diameter of 80mm is placed at the bottom of the box, which is the target of the reaching task. A capsule-style virtual hand is used to represent subjects' hand and lower-arm in the virtual scene. The box is opaque, so the virtual hand and the ball are completely occluded by the box except under certain experimental condition. A start button, which the subject interacts with to start an experiment trial, hovers under the box.

The task design of our experiment refers to Bloomfield's work [8]. Subjects start the task by pressing the start button. Then, the subjects insert their hands through the box opening and try to touch the ball without colliding with the box, while the subjects are not allowed to move their head to peek into the box. When the subjects touch the ball, a bell sound is given as the feedback of the valid touching. Finally, the subjects withdraw their hands and wait for the next trial.
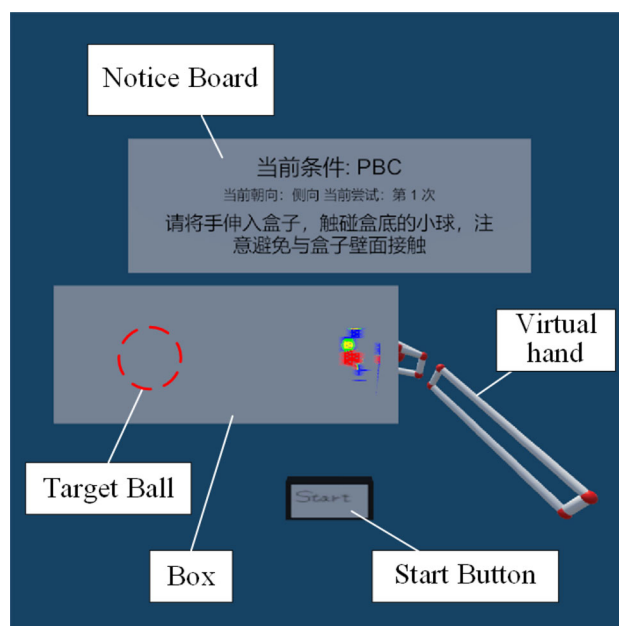
Fig. 1 Experiment scene of the reaching ball task in a box. It should be noticed that we mark the position of the ball using the red circle for readers convenience. The ball was actually invisible for subjects during the experiment except in See-through condition

4.2 Visual collision feedback

According to the literature review in Sect. 2.3, we selected five popular forms of visual collision feedback which represents different levels of informativeness. The selected feedback includes three visual collision cues mentioned in Sreng's work [16], namely the proximity-based coloring gradation (Gradation), the proximity-based deformation (Deformation), and the hybrid of both (Hybrid). The three proximity-based cue conditions represents the situation where users perform the movement with proprioception and different collision information. The contact-based coloring (Coloring) is included since it is the most commonly used collision feedback, indicating a situation where users are informed of the *event* detail of collisions only. Finally, the see-through visualization (See-through) is involved because it is a common method to display the occluded environments and represents a condition where users perform the movement based on the proprioception and basic visuo-motor synchronization. We rated the informativeness level of the selected visual feedback according to the proposed metric, and the results are presented in Fig. 2. We set the understanding ratings of all detail elements of See-through as *hard* based on the assumption that the complicated structure in an enclosed space increases the difficulty of the understanding the spatial information, leading to more reasoning efforts to judge the collision.
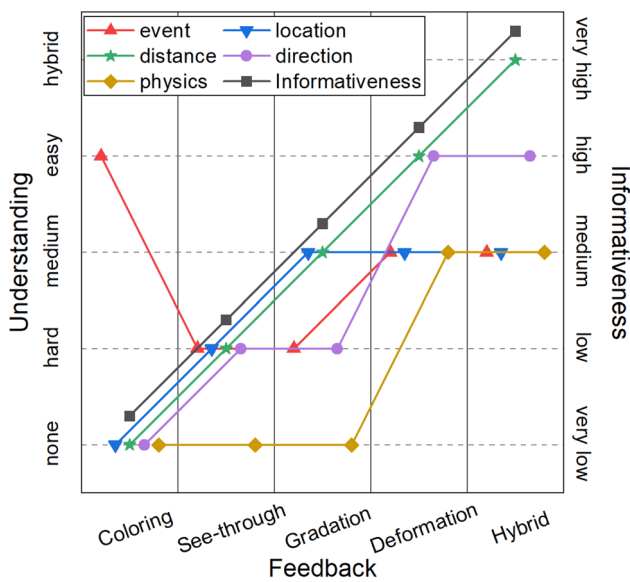
Fig. 2 Informativeness level of the visual feedback. The order of the informativeness level from low to high is Coloring, See-through, Gradation, Deformation, and Hybrid

To compare the effects of the selected visual feedback, a baseline condition without any visually-enhanced feedback was set as a reference, representing a situation where users perform the movement only based on proprioception. Considering that it is difficult for users to perform the task in the baseline condition, we added an auxiliary auditory cue (bird sound) to notice the collisions in all experimental conditions. This may slightly affect the experiment results regarding the detail element of *event*. Correspondingly, we added an amendment question in the questionnaire to indicate the bias. In summary, there are six conditions included in the experiment.

We used the Unity engine to implement the visual collision feedback. The visualization of the three proximity-based visual cues (Gradation, Deformation, Hybrid) were implemented based on a wireframe shader [34], where the visibility, color, and offset of the object vertex are updated

according to the local minimum distance between the vertex and the hand joints. The gradation setting refers to [16]. For deformation, we implemented a bump effect, where the object surface is sunken as the hand approaches the point of imminent collision, and after contact, the surface bulges with the depth of penetration. The sunken depth and the bulges height are linearly correlated to the proximity and the penetration depth respectively. The hybrid cue is a simply mixture of the effects of coloring gradation and deformation. The see-through effect was achieved by adjusting the alpha channel of the object texture to 0.5. The contact-based coloring effect was realized based on the event callback of the Unity's physics engine, where the color of the entire object is changed to red. Fig. 3 shows the visual effects of the six visual collision feedback in our experiment.

4.3 Apparatus

We used HTC Vive (resolution 1080 x 1200 pixels per eye, refresh rate 90 Hz, field of view 110) as the head-mounted display (HMD) device. To ensure the hand-tracking stability, we used two Leap Motion Controllers (LMC) to capture subjects' hands movements (see Fig. 4). One of the LMC was placed at a in front of the subject, the other one was mounted on the front surface of HMD helmet. The hand-tracking data from the two LMCs were fused by taking the weighed-averages according to the tracking confidence from the LMC SDK [35]. The auditory feedback was provided through a voice box.

The HTC Vive were connected to a graphics workstation (Intel Xeon Gold6128 CPU, NVIDIA Quadro RTX 6000, 128GB RAM, Microsoft Windows 10) via the DP port and the USB 3.0 port. Due to the bandwidth limitation, the two LMCs cannot connect to one computer simultaneously. Thus, the LMC mounted on the HMD was connected to the graphics workstation through the USB cable, the desktop LMC was connected to an Intel NUC computer (Intel Core i5-8259U at 2.3 GHz, 8GB RAM, and Iris PlusGraphics 655). Then, the hand tracking data captured by the desktop LMC was
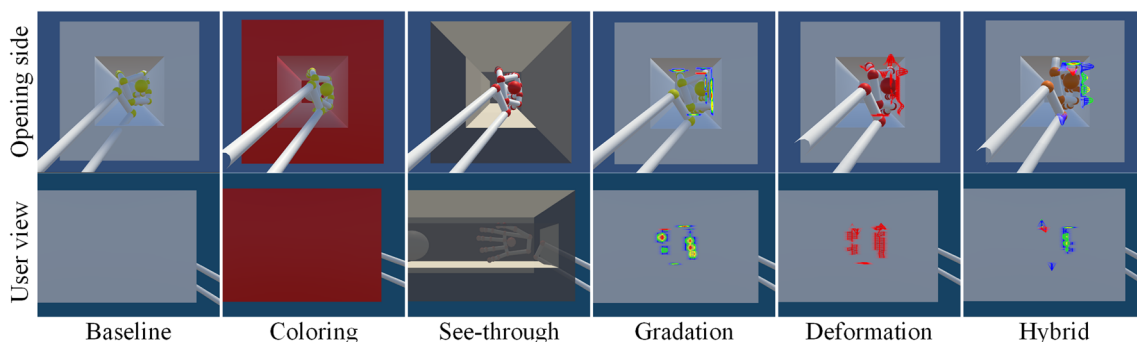

Fig. 3 Visual effects of the feedback in each experimental condition. The first row shows the situation in the box, the second row shows user's view
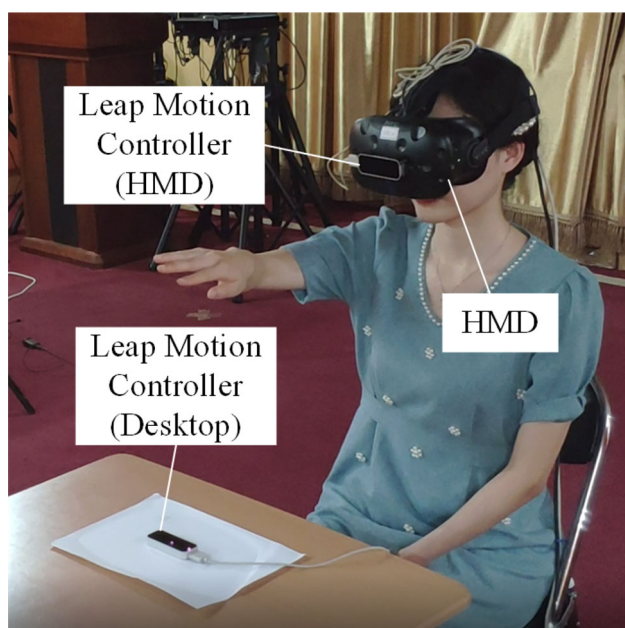
Fig. 4 Experiment devices

transmitted to the graphics workstation through the local area network. The latency of the data transmission is less than 5ms. The experiment program was developed in Unity game engine (2019.4.7f1) with the Leap Motion plugin (Core 4.4.0) [36]. The rendering rate of the experiment program was about 60 Hz with our setup.

4.4 Subjects

We invited 18 college students (8 females, aged from 22 to 30, M=25.78, SD=6.89) via posters and social media. All subjects had normal vision or corrected vision, and are right-handed. The palm widths of subjects' right hands were measured using a caliper, and the results are M=77.44mm, SD=6.74mm. The palm widths of subjects' virtual hand were automatically scaled by Leap Motion, and the statistical results are M=81.66mm, SD=2.48mm. When asked to rate their experience of using VR using a 5-point Likert scale (1:very unfamiliar, 5:very familiar), 14 subjects rated their experience beneath 2 (unfamiliar), and four subjects rated as 3 (neutral). All subjects volunteered to join the experiment with no payment.

4.5 Measurement

We used the task time and the collision time ratio (CTR) as the objective measures to evaluate the hand movement performance. The task time is defined as the period from the time when the subjects' hand firstly enter the box to the time when the hand firstly touches the ball. The CTR is defined as the ratio of accumulated time of the hand colliding with the

box's walls to the task time. We used the Interaction Engine [37] provided by Leap Motion SDK to detect the collision between the virtual hand and the box's walls. In addition, the trajectory data of palm joint were also recorded for analyzing the hand movement status in the box.

To collect subjective opinions, We designed a post-test questionnaire which contains four statements, as shown in Table 1. The first two questions subjectively assess subjects' collision awareness and spatial awareness during the task. The third question collects subjects' opinions on the helpfulness of each visual feedback compared to the auditory cue to reflect the bias in the *event* detail. We also designed a question to collect subject's feeling of visuo-tactile synesthesia shown as Q4 referring to Biocca's questionnaire [29,38]. Subjects provided their answers for Q1, Q2 and Q4 using a 7-point Likert scale from 1 to 7, where 1 represents strongly disagree and 7 represents strongly agree. For Q3, we used a 11-point Likert scale from 0 to 10 to collect subject's opinions on the contribution proportion (0: the audio cue made all the contributions; 5: two cues contributed similarly; 10: the visual cue made all the contributions). To collect user's preference and open-ended comments, a post-experiment questionnaire was conducted.

4.6 Experiment procedure

When subjects arrived, we briefly explained the experiment to them. Then, the subjects familiarized themselves with the virtual environments and the visual feedback in a demo scene. After the subjects felt ready, the formal experiment began.

We firstly collected the subject's profile using a pre-test questionnaire before the experiment. At the beginning of each condition, we adjusted the position of the box to ensure the best comfort for each subject, and the size of the box opening was scaled according to the subject's virtual palm width to maintain the task difficulty. Then, the subject performed the task successively under all conditions. The order of the experimental conditions was counter-balanced using the Latin-square. In each condition, the subjects performed the task repeatedly for 15 trials, during which the subjects were told to complete the task as accurately and quickly as possible, meanwhile, the task time, CTR, and the trajectory data of palm were recorded. There was a ten-seconds break between each trial and a one-minute rest between every five trials. After the subjects finished all trials of a condition, they were asked to evaluate the current condition using the post-test questionnaire. After all conditions were completed, a post-experiment questionnaire was given to the subjects to collect their overall preference and open-ended comments. In summary, each subject needed to complete the reaching task for 15 x 6=90 trials, taking about 90 minutes to complete the experiment.

Table 1 Post-test questionnaire

Measures	Question
Collision awareness	Q1. I can easily avoid the collision between my hand and the box
Spatial awareness	Q2. I can easily understand the spatial relations between my hand and the surrounding objects
Helpfulness	Q3. Regarding the visual cue and the audio cue, which do you think is more helpful?
Synesthesia	Q4. I feel like I was touching a real object during the task

(a) Task time (b) Collision time ratio
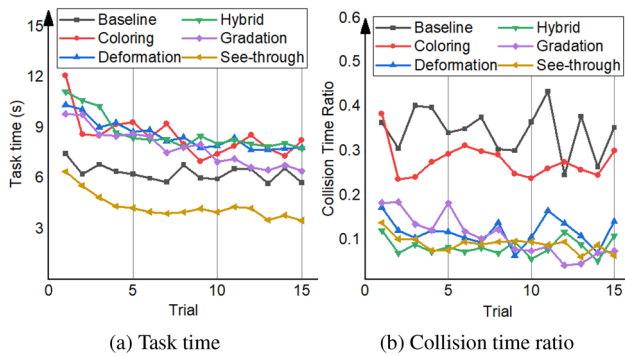
Fig. 5 Learning curve of task time and collision time ratio under all experimental conditions

5 Results

We reports on the experimental results by three parts: (1) Objective measurements, where the measurements of the task time, CTR, and the distribution of palm position inside the box are reported; and (2) Subjective feedback, where the scores of the questionnaire is analyzed; and (3) The open-ended comments. The data analysis was performed in OriginPro 2021. The normality was tested using the Shapiro-Wilks test. Results show that the distribution of all measurements reject normality. Thus, the Friedman ANOVA was used to test the difference, and the Wilcoxon signed rank tests were conducted for the post-hoc analysis. The mean difference was significant at .05 level, and the Bonferroni correction was automatically applied for multiple comparisons unless noticed otherwise.

5.1 Task performance

5.1.1 Learning effects

We first checked the learning effects on the measurements. Fig. 5 shows the learning curve of task time and CTR against 15 trials of the task. To test the learning effects, we split the measurements of the 15 trials into three blocks and conducted an ANOVA over the three blocks. Results show that in the measurements of task time under all conditions, significant

differences were found in the pair of block one and block two ($p < 0.01$), as well as the pair of block one and block three ($p < 0.001$), but no significant difference was found in block two and block three ($p > 0.264$), indicating a significant learning effect exists in task time. However, there was no significant difference found in the three blocks of CTR, indicating no significant learning effect exists in CTR ($p > 0.058$). Based on the findings, we used the data of latter two blocks in the following analysis to eliminate the bias caused by the learning effects.

5.1.2 Task time

Statistical results of task time under the six conditions are shown in Fig. 6a. The averaged task time is the longest under Hybrid condition ($m = 8.11$, $SD = 3.95$), while is the shortest under See-through condition ($m = 3.95$, $SD = 2.07$). Result of the Friedman ANOVA shows that there are significant differences found in the task time of all conditions ($\chi^2(5) = 38.6$, $p <= 0.001$). The post-hoc analysis shows that there is no significant difference found among the three proximity-based cue conditions ($p > 0.37$), whereas, the task time is significantly longer task time in Hybrid condition than in Baseline condition ($Z = 4.28$, $p = 0.030$). The task time in See-through condition is significantly shorter than in all other conditions ($p < 0.007$) except Baseline condition ($Z = 3.4$, $p = 0.156$). There is no significant difference found between the task time in Coloring condition and the proximity-based cues condition ($p > 0.81$).

5.1.3 Collision time ratio

Statistical results of the CTR under the six conditions are shown in Fig. 6b. The averaged CTR is highest under Baseline condition ($m = 0.34$, $SD = 0.17$), while is lowest under Hybrid condition ($m = 0.082$, $SD = 0.068$). Result of the Friedman ANOVA shows that there are significant differences found in the CTR of all conditions ($\chi^2(5) = 51.40$, $p < 0.001$). The post-hoc analysis shows that the CTR is significantly lower in the proximity based cue conditions and See-through condition than in Baseline and Coloring conditions ($p < 0.03$). But no significant difference is found among the CTR in See-through, Gradation,
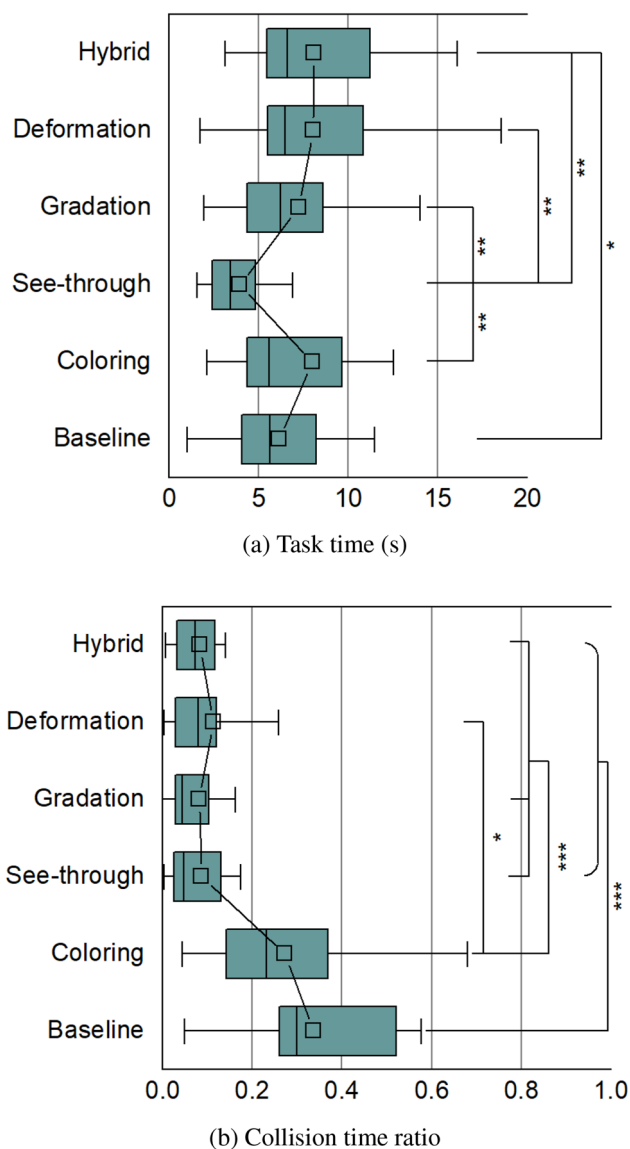
(a) Task time (s)

(b) Collision time ratio

Fig. 6 Results of the task performance under different condition. Boxes represent the inter-quartile intervals, cube dots represent the means, whiskers represent the 95% confidence intervals, $*p < 0.05, **p < 0.01, ***p < 0.001$

5.1.4 Hand trajectory distribution

We used 2D Gaussian kernel density to calculate the distribution density of the palm movement trajectory on the cross-sectional plane (Vertical-Depth) across all subjects against the six experimental conditions, as Fig. 7 shows. In general, the SD of distributions are smaller in conditions with visual collision feedback than without. Among the visual feedback conditions, the SD of the distributions are smaller

under the three proximity-based cue conditions. Separately, Hybrid condition ($SD = 10.45$) has the smallest SD in depth direction, while the largest SD appears in Deformation condition ($SD = 19.76$). In vertical, the Deformation condition has the smallest SD ($SD = 6.47$), while Coloring ($SD = 9.34$) and See-through (10.28) conditions has the largest SD among the visual feedback conditions. In addition, Coloring condition has the largest SD in vertical direction among all proximity-based conditions.

In terms of deviation, the mean of the distributions are generally larger in the three proximity-based conditions among all visual feedback conditions. To be specific, See-through condition shows the nearly zero deviation in vertical, while Coloring condition shows the largest bias ($m = 3.85$). In depth, the smallest deviation appears in Coloring condition ($m = 5.44$) among all visual feedback conditions, while Deformation condition shows the largest bias ($m = 11.02$).

By qualitatively evaluating the shape of the distributions, we found that the palm of subjects moved more balanced in box under Baseline, Coloring, and See-through conditions, as illustrated by the relatively balanced contacts with the four walls of the box. In contrast, obvious unbalanced movements were found in the three proximity-based conditions. In Gradation and Hybrid conditions, the hand tended to approach the corner lying between the front (close to subjects) wall and the bottom wall of the box. In Deformation condition, however, the hand were more likely to swing back and forth along depth direction. We also found abnormal movement behavior under Deformation conditions showing as the high distribution density in the area of the back wall marked in Fig. 7e.

5.2 Subjective feedback

5.2.1 Collision awareness

Statistical results of the rating scores of Q1 under the six conditions are shown in Fig. 8a. As shown in the figure, the averaged rating score is the highest under Hybrid condition ($m = 5.94, SD = 0.87$), while is the lowest under Baseline ($m = 2.33, SD = 1.81$) and Coloring conditions ($m = 2.33, SD = 1.46$). Result of the Friedman ANOVA reveals that significant differences exist in the rating scores ($\chi^2(5) = 46.91, p < 0.001$). The post-hoc analysis indicates that the rating scores are significantly higher in Gradation, Hybrid, and See-through conditions than in Coloring and Baseline conditions ($p < 0.001$). There is no significant difference found among the rating scores in the proximity-based cue conditions as well as the See-through condition ($p > 0.156$).
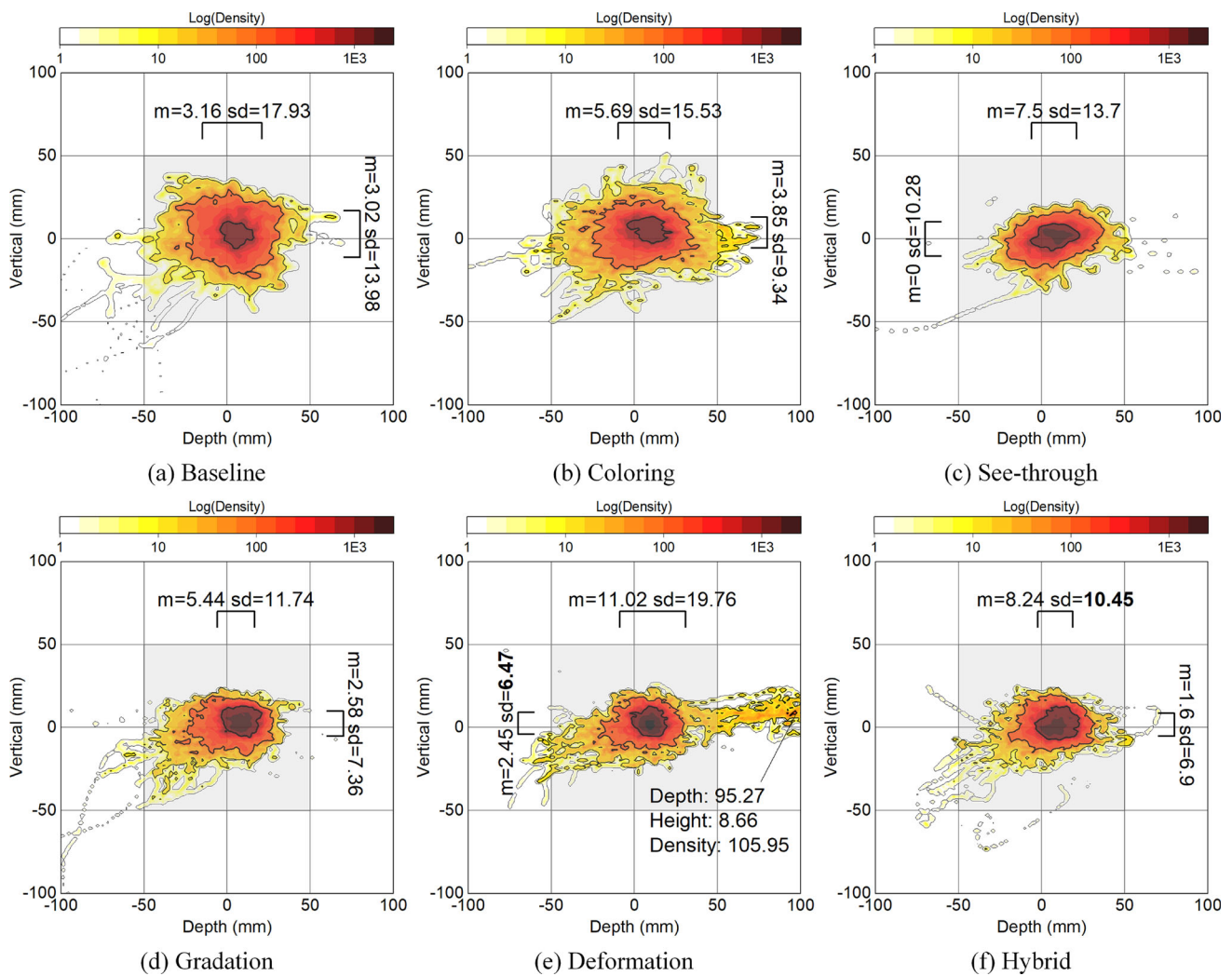
Deformation and Hybrid conditions ($p > 0.82$), in addition, the standard deviation of CTR significantly increased in Deformation condition ($SD = 0.135$).

Fig. 7 Hand trajectory distribution on the cross-section plane of the box. The gray area represents the enclosed space of the box

5.2.2 Spatial awareness

Statistical results of the rating scores of Q2 under the six conditions are shown in Fig. 8b. As shown in the figure, the averaged score is the highest under Hybrid condition ($m = 5.72$, $SD = 1.01$), and is the lowest under Coloring condition ($m = 1.72$, $SD = 0.96$). Result of the Friedman ANOVA shows that there are significant differences found in the rating scores ($\chi^2(5) = 54.91$, $p < 0.001$). The post-hoc analysis shows that the rating scores are significantly higher in See-through condition as well as all of the proximity-based cue conditions than in Coloring condition ($p < 0.022$), but no significant difference is found among the rating scores in the proximity-based cue conditions and the See-through condition ($p > 0.227$).

5.2.3 Helpfulness

Statistical results of the rating scores of Q3 under the six conditions are shown in Fig. 8c. It can be seen that the averaged score is the highest under Hybrid condition ($m = 8.67$, $SD = 1.75$), and is the lowest under Baseline condition ($m = 1.78$, $SD = 2.75$). The Friedman ANOVA shows that there are significant differences found in the rating scores ($\chi^2(5) = 50.7$, $p < 0.001$). The post-hoc analysis show that the rating scores are significantly higher in Gradation, Hybrid, and See-through conditions than in Coloring condition ($p > 0.006$). There is no significant difference found among the rating scores in the proximity-based cue conditions as well as See-through condition ($P > 0.126$). No significant difference is found between the rating scores in Coloring condition and Baseline condition ($p = 0.83$).
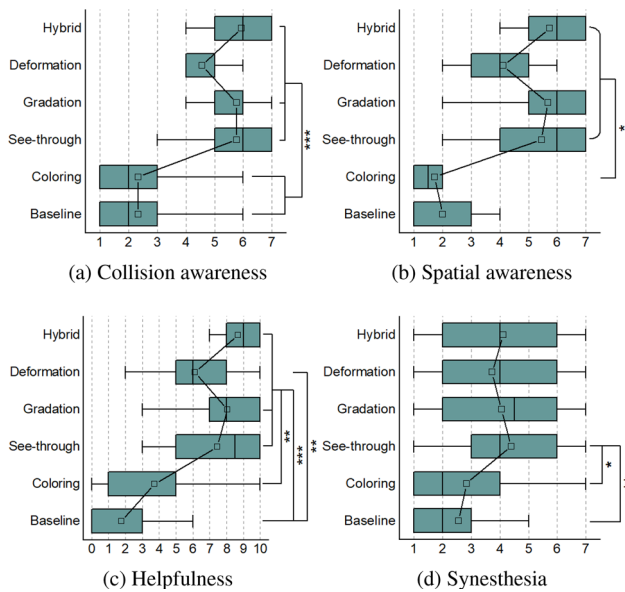
Fig. 8 Statistical results of the subjective ratings under different conditions. Boxes represent the inter-quartile intervals, cube dots represent the averaged rating scores, whiskers represent the 95% confidence intervals. $*p < 0.05, **p < 0.01, ***p < 0.001$

Fig. 9 Subjects preference on the six conditions (the upper the preferred)

5.2.4 Synesthesia

Statistical results of the rating scores of Q4 in the six conditions are shown in Fig. 8d. The figure shows that the averaged rating score is the highest under See-through condition ($m = 4.39$, $SD = 1.79$, while is the lowest under Baseline condition ($m = 2.55$, $SD = 1.81$). Result of the Friedman ANOVA reveals significant differences exist in the rating scores ($\chi^{(5)} = 17.65$, $p = 0.003$). The post-hoc analysis indicates that the rating scores are significantly higher in See-through condition than in Baseline and Coloring conditions ($p < 0.019$). No significant difference is found among the rating scores in other conditions ($p > 0.14$).

5.2.5 Preference

We summarised subjects' preference in Fig. 9. It can be seen that Hybrid ($M = 2.06$, $SD = 0.94$) and Gradation ($M = 2$, $SD = 0.91$) are the most two preferred conditions, and no significant difference was found between them ($Z = -0.068$, $p = 0.946$). See-through ($M = 2.72$, $SD = 1.64$) is the third preferred condition of which the rank is slightly lower than Hybrid and Gradation (p>0.146), and slightly higher than Deformation ($M = 3.56$, $SD = 0.78$, $Z = -1.611$, $p = 0.107$). Baseline ($M = 5.83$, $SD = 0.38$) and Coloring (4.94, $SD = 0.73$) are the least two preferred groups, but the rank of Coloring is not significantly different from the rank of Baseline ($p = 0.11$).
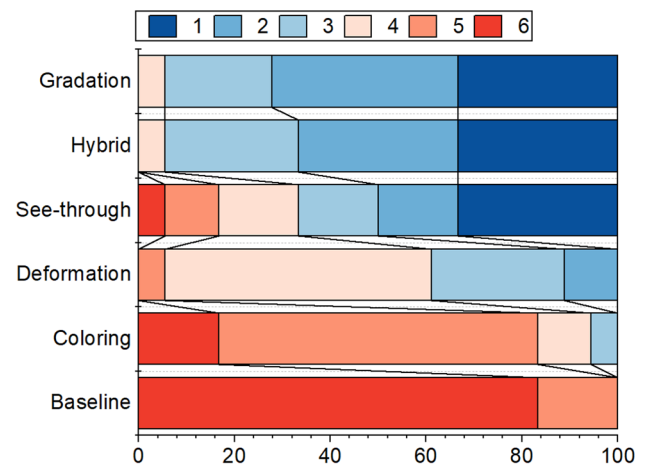
5.3 Open-ended comments

At the end of experiment, subjects were asked to describe the pros and cons of each conditions in a post-experiment questionnaire. We collected the feedback of all subjects and listed the reoccurring themes with number of mentions in Table 2.

Coloring. Subjects disliked Coloring condition for many reasons. Most subjects complained that the Coloring feedback cannot provide enough collision information. Four subjects mentioned that the Coloring feedback "annoyed me because I did not know how to get rid of the collision", "The stimuli was too strong and made me stressed" (S9). Four subjects mentioned that "the Coloring cue had no difference with the baseline condition" (S7). But four subjects liked the Coloring feedback because they felt the effect of the Coloring cue was "obvious and made the reaction quicker" (S11).

Gradation vs. Deformation. When asked to compare the Gradation and Deformation cue, 11 subjects clearly expressed that they preferred the Gradation cue rather than the Deformation cue. However, subjects were split when asked whether the cue is intuitive to present proximity and collision. Seven subjects mentioned that the Gradation cue was more intuitive for displaying proximity and collision than the Deformation cue, while four subjects presented the inverse opinions. The subjects who preferred Gradation thought that the cue "is easy to understand" (S13), "has a strong visual stimulus" (S16, S17, S18) and "is neat and clear" (S10). While, the Deformation cue was considered to be "confusing when the hand simultaneously approaching two adjacent walls" (S4, S10), "difficult to identify the distance by the deformation" (S4, S11, S15, S18), and "the bulge affected my judgement" (S10, S17, S18). On the other hand, subjects who preferred Deformation thought the cue is "more intuitive to understand the proximity and penetrating depth"

Table 2 Open-ended comments of all visual feedback conditions, the numbers in brackets are the occurrence times of the themes

Visual cue	Pros	Cons
Coloring	Quick reaction, obvious(4)	No difference with no-cue (4), Annoying, Stressed(4), Cannot tell where the collision occurred (7), No help to adjust hand movement (5)
See-through	A clear view of the interior (12), Intuitive (7), Good in non-depth condition (4)	Awkward in depth condition (4), Boundary is unclear (4), Inconspicuous collision feedback (10)
Gradation	Intuitive (7), Good in the depth direction (3), Strong visual stimuli (3)	Lots of cognitive workload (2), bad in the non-depth direction (3)
Deformation	Intuitive (4), Good in the non-depth direction (4)	Bad in the depth direction (7), Confusing (2), Unclear (4), Affect judgement (3)
Hybrid	Abundant information (2), Pre-warn, pre-adjust (2)	Too much information, heavy cognitive workload (3), Confusing (3)

(S2, S3, S4, S14). S14 mentioned that the deformation effects were more similar with the real touch feeling, especially after penetration. They disagreed with the Coloring cue because "The change of color brings a lot of cognitive load, I need to keep processing which color corresponds to which level of proximity", "Color-changing needs more practice" (S2, S4).

When asked to talk about the helpfulness, three subjects mentioned that the Gradation cue "worked better for presenting the information in depth direction", but worse in non-depth direction because "the 2D coloring cues become invisible on the surfaces parallel to user's sight line" (S7). Four subjects indicated that the Deformation cue performed better in the non-depth direction because "the bulge is obvious". But seven subjects mentioned that the Deformation cue worked badly in the depth direction because "the bulge is parallel to the sight line and become invisible in the depth direction" (S12). Regarding this phenomenon, S3 recommended to use Gradation cue in non-depth direction and use Deformation cue in depth direction.

Hybrid. Subjects' attitudes to the Hybrid cue were neutral. Five subjects liked the Hybrid cue because it provided "rich information" (S4, S16) and "integrate the advantages of gradation and deformation" (S9). But five participants disliked it because "it displayed too much information and brought heavy cognitive workload" (S3, S6, S17), and also "made me confused when the hand approaching two adjacent walls" (S3, S9, S11) which was mainly caused by the deformation effect.

See-through. See-through condition was remarked as the most natural and intuitive condition (S11: I can clearly see the movement inside the box, intuitively understand the hand position). But several explicit drawbacks were also mentioned. Eight subjects mentioned the difficulty of perceiving collision (S5: The cue of collision was inconspicuous, I could only use audio cue to judge the collision). Two subjects thought that in See-through condition, "the collision in depth direction cannot be warned" because "the collision with the rear wall was occluded by the hand" (S18). S14 mentioned an interesting opinion that the See-through condition goes against the reality, "In the real case, people cannot see through the box walls, but the See-through effect gives people a kind of 'super power', making the experience unreal."

Suggestions. In addition to S3's recommendation, several subjects gave suggestions to improve the visual cues. S4 and S12 mentioned that "The color used in Gradation can be simplified by just using two colors to respectively represent approach and collision". This accords with the suggestions of Ariza [32]. S4, S7 and S16 suggested that "the See-through condition can be improved by merging other visual effect in it", which is similar to the cue effects proposed by Sreng [16].

6 Discussion

In this section, we first discuss the effects of the informativeness level of the visual collision feedback on the hand movement performance in confined-occluded spaces. Next, we give a deep insight on the effects of the spatial information encoded in the visual collision feedback regarding the detail elements and understanding levels. Finally, we provide suggestions for visual collision feedback in enclosed spaces, and propose a promising technology and its implications.

6.1 Effects on hand movement performance

6.1.1 Movement speed

Experiment results cannot support the hypothesize H1, showing as the non-linear relations between the informativeness level and the task time. To be more specific, the task time was the shortest in See-through condition which was rated as a median-lower level of informativeness, while the time was longest in Hybrid condition which was rated as the highest level of informativeness, even longer than in Baseline condition. This accords with Zhang's work [6], where the pointing movement time becomes longer when the visual feedback of the hand movement is restricted. We assume the possible reason may be that limited by the visual feedback of hand movement, users may need longer time to correct the movement based on proprioception or the provided collision metaphor (supported by the open-ended comments of S2, S4), in contrast, users could easily confirm the corrections in See-through condition, making users more confident to move quickly. Another possible reason could be that due to the collision cue is inconspicuous in See-through condition, users may focus more on reaching the target instead of avoiding the collisions, leading to the hand movement may change from a steering movement [39] to a standard Fitts movement [40], and the task time was shortened consequently [9]. In the proximity-based cue conditions, subjects may be more likely to use the visual cue as a guide for hand movement (showing as the hand movement trend toward the front-bottom corner of the box in Fig. 7d, e and f instead of concentrating on trajectory planning, which leads to longer task time [13]. In Baseline condition, however, subjects probably took the "quick and dirty" approach to quickly complete the task with the cost of accuracy due to the task difficulty [13]. Our finding also extends the previous work on improving hand movement speed in the enclosed space, where the effects of visual feedback on hand-object interaction is limited compared to clear visualized visuo-motor synchronization.

6.1.2 Movement accuracy

The experiment results support the hypothesize H2 as lower CTR and more concentrated hand trajectory distribution were found in the visual feedback condition with higher informativeness level.

The contact-based coloring feedback has very limited effect on improving user's collision awareness and spatial awareness, manifested as poor objective and subjective performance in the experiment. Possible reason could be that the collision information provided by the Coloring feedback may not be efficient enough to help subjects to leave the collisions. Our results are in line with some of works investigating other modalities of event feedback, such as audio, vibrotactile, and haptic, in which users under conditions of single modality of feedback with obscure information are more likely to be confused and to make more mistakes in tasks [7,13].

The see-through visualization is believed to be the most natural feedback of collision according to subjects' feedback. However, although it has the same level of CTR compared to the three proximity-based feedback, the hand trajectory distribution under See-through condition was relatively decentralized, and subjective feedback reported that additional information was needed to judge collisions under See-through condition (supported by a relatively lower rating of Helpfulness compared to Gradation and Hybrid condition).

The improvements on the hand movement accuracy made by the three proximity-based feedback are significant, manifested by the superior performance in both CTR and hand trajectory distribution. Subjective feedback and the ranking results also shows user's preference on the Gradation and Hybrid feedback to inform collisions. However, we found obviously different performance in the depth and vertical direction. Specifically, the proximity-based gradation effect is more advantageous in depth direction, while the deformation effects is more conspicuous in vertical direction. Whereas, the deformation effect appears to be detrimental to depth perception, showing as the large amount of trajectory outliers marked in Fig. 7e. The possible reason, as mentioned in the open-ended comments, could be the interference between the effective direction of the visual cues and the user's view direction. Another reason may be that due to the visual effects are triggered according to the local minimum distance, it is possible to trigger the visual effects of two near parallel surfaces perpendicular to depth direction at the same time. The gradation effects are workable in this situation because users can judge the distance according to the color difference between the two surfaces. However, it is impossible for deformation because the distance cues are invisible in depth direction, and judging the depth relations between the two surfaces could be difficult since the two sur-

faces are the same color and overlap with each other in user's view. The hybrid effect seems to integrate the advantages of the gradation and deformation, however, rich information may lead to a higher cognitive load according to subjects' feedback.

6.2 Effects of the spatial information

In this part, we further discuss some post-hocs findings on the effects of the visual collision feedback in view of the information details and the understanding level, which is to answer the exploratory questions proposed in Sect. 4.

6.2.1 Information detail

There are five detail elements included in the detail dimension of our metric. According to the experiment results, we believe that the *location*, *distance*, and *direction* are the three most effective information details to judge the collisions, as they were included in the high performance conditions (See-through, Gradation, Deformation, and Hybrid) and excluded in the low performance conditions (Baseline, Coloring). However, we cannot distinguish their effectiveness. The *event* element appears to be less effective in our context, but related researches suggest that the *event* element works better in simple tasks, such as pointing movement [32] or grab task [17]. The *physics* element may not help to understand collisions either, as the experimental conditions with the *physics* element did not show significant difference in task performance compared to the conditions without. The *physics* element is proved to be useful to simulate haptic effects [41] and support more complicate interaction task, such as perceive the softness or texture [42]. However, our results contradict with previous works [38], showing as the visuo-tactile synesthesia illusion was reported in See-through condition, not in the conditions with metaphors of force (Deformation and Hybrid). We assumed the possible reason could be that the absence of hand representation in Deformation and Hybrid conditions reduced user's sense of embodiment, breaking links between the visual feedback and the haptic illusion [43,44].

6.2.2 Information understanding

Based on the definition, the three understanding level *hard, medium, and easy* can be mapped to natural representation, implicit artificial metaphor, and explicit artificial symbol [45]. By comparing the experimental results under the conditions with different understanding level, we found that, in general, there is a trend where the movement speed is quicker in the condition with more natural and implicit representations of spatial information (See-through) than in the conditions with more artificial and explicit representations of the spatial information (Gradation, Deformation, and

Hybrid). The trend goes inversely in view of the movement accuracy. However, because the experimental conditions were not orthogonal grouping, this trend needs to be further proved.

In summary, location, distance and direction are key detail elements for understanding the spatial information about collisions. In addition, the integration of See-through visualization and proximity-based visual cues might be the best compromise between the speed and accuracy of the hand movement in confined and occluded spaces in VR.

6.3 Suggestions and implications

Based on our findings, we put forward suggestions about the visual collision feedback for hand interaction in confined-occluded VR spaces. We redesigned the feedback with the following considerations in mind:

> *Compromise between accuracy and speed.* In order to balance the accuracy and speed, local See-through visualization and proximity-based coloring gradation cues was integrated into the new feedback technology. The local see-through effect provides a natural view of the hand surroundings to ensure the movement speed, while the proximity-based coloring gradation cues provides accurate collision information that works well both in depth and vertical.
> *Feedback filter.* As suggested by Sreng [16] and subjects, appropriate filters that can hide the feedback effects of trivial collisions are favorable to reduce cognitive load and increase focus on the key collisions. Despite the filter proposed by Sreng, the multi-trigger phenomenon on the adjacent parallel surfaces should also be concerned, this is of vital importance to the feedback display in depth direction.
> *Modeling.* A concern arose during the implementation phase in our study, that is, the visual effects based on wireframe shader extremely count on the quality of object's mesh model. The feedback works well if the meshes are carefully modeled and evenly distributed on the model surface, otherwise, the display effects of feedback could be weird. However, most of the virtual assembly simulation models are exported from computer-aided design software, where the mesh models will be greatly simplified to non-uniform triangle mesh. This brings tremendous work to model preprocessing.

Based on the considerations, we proposed a new visual collision feedback based on projector decal, which looks like a handprint projected on the affected object, with different shading effects to display the collision information. The idea came from our daily act of putting hands on glass, and was improved based on the works of Greene [46] and Reddy [47],
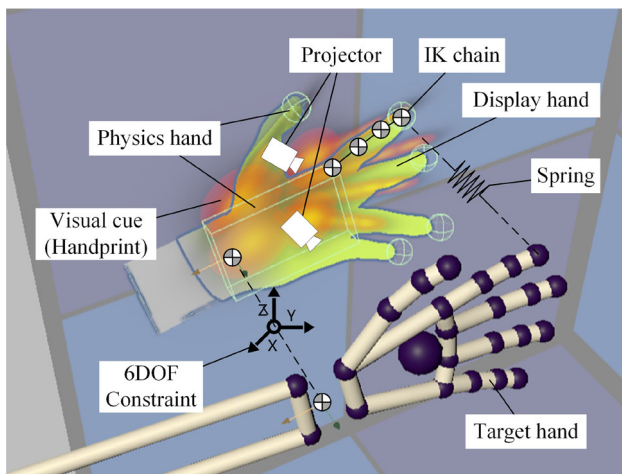
Fig. 10 Technology principle

Fig. 11 Proximity texture

where the depth map textures of five fingers and palm are calculated individually to achieve collisions feedback for more complex hand-object interaction, and a adaptive projection algorithm is developed. The implementation details are introduced below.

Figure 10 shows the principle of the feedback technology. We applied six projectors to project the collision information of different parts of hand on objects, that is, five projectors to inform the collision states of the five fingers, and one projector to display the collisions of palm. The clipping volume of the projector is an orthogonal spaces. The width and height of the volume are determined by the maximum dimensions of the hand parts in the X-Y plane, the far and near planes are symmetrical along the X-Y plane in a distance of an adjustable parameter *Range*. Similar to Reddy's work [47], we realized a physical interaction hand to update the pose of projectors adaptively. As shown in Fig. 10, the physical interaction hand includes:(a) The target hand refers to the tracking position of the user's hand; (b) Physics hand, i.e. a group of colliders bound to the target hand, and interact physically with the objects in the virtual scene; and (c) Display hand, showing the final interaction effect, whose movement is controlled by the physics hand through inverse kinematics (IK) algorithm. The projectors move with the display hand. In each update frame, the finger projector is positioned at the midpoint of the line connecting the proximal finger joint to the fingertip, the palm projector is positioned at the palm center, and the z-axis of all projectors are always facing to the corresponding target joints.

Different from Greene's work [46], the depth map texture is calculated only based on the hand joint positions. In the clipping space of the projector, the texture sampling coordinates of each pixel in the affected fragment are calculated
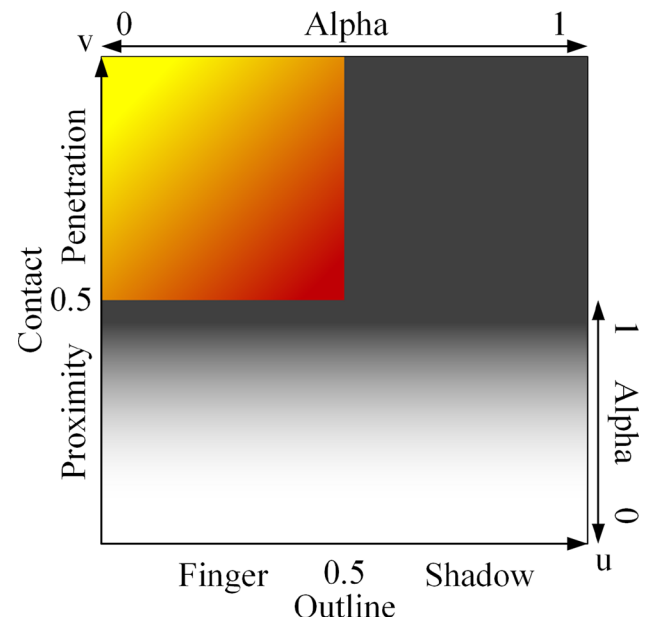
using a 2D Gaussian method given as:

$$u = \max \exp \left(- \left(\frac{(x - p_x)^2}{2\sigma_x^2} + \frac{(y - p_y)^2}{2\sigma_y^2} \right) \right) \quad (1)$$

$$v = \max p_z u \quad (2)$$

Where, p_x, p_y, p_z are the coordinates of hand joints transformed to the projector's clipping volume. For finger projector, the hand joints refer to the proximal, intermediate, distal, and fingertip joints; for palm projector, they refer to the metacarpal and proximal joints of the index and pinky fingers. σ_x, σ_y control the size of the cues of the hand joints. After calculating the sampling coordinates, the RGBA of the pixel are sampled from a 2D proximity texture as Fig. 11 shows. It can be seen that, u dimension is used to outline the hand part shape, while the v dimension accounts for displaying the distance information. To clearly identify the collision event, the color gradations for penetration and proximity are set as two obviously different schemes (gray and red). Using the gray scheme to indicate proximity also could help to reduce the distraction by simulating a shadow effect [46]. The alpha channel is adjusted to fade the edges.

Finally, in order to visualize the collision cue in occluded spaces, the ZTest is turned off and the cue is rendered after all objects in the scene completes the rendering. To filter the unwanted feedback on the backward surfaces, the surfaces are clipped if the inner product between the normal of the surface and the projection direction is greater than zero. Furthermore, according to our experiment findings, we set the display hand to be visible in the occluded spaces, providing visuo-motor synchronization to improve the movement
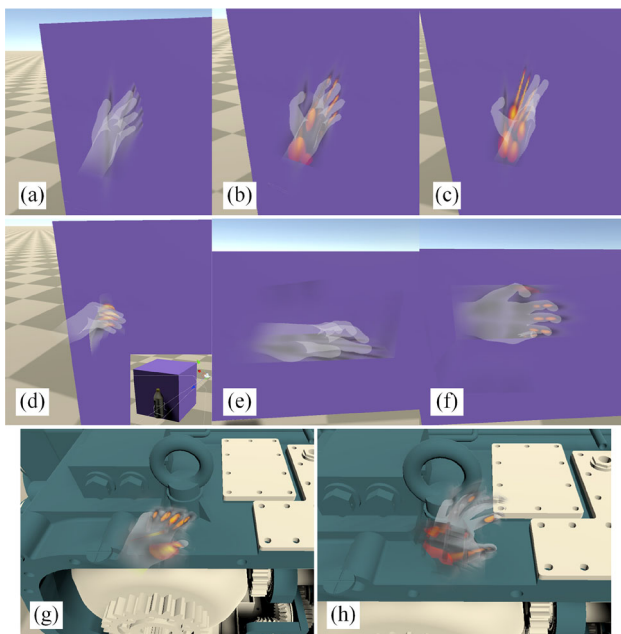
Fig. 12 Application demos of the proposed visual collision feedback, namely: Hand **a** approaches, **b** contacts and **c** collides with an occluded plane, **d** grabs a bottle behind a wall, touches the **e** bottom and **f** back walls in the confined box, and **gh** searches in a complex gear box

speed. Fig. 12 shows the application demos of our proposed collision feedback. The demos were implemented in Unity3D engine 2019.4.7f1.

Since most of the computation is done by GPU, our technology achieves excellent computer efficiency, with a refreshing rate of about 250Hz in a scene with 35.6k triangles, and 100Hz in an extreme situation with 2 million triangles, on a household gaming laptop (Intel core i7, 16GB RAM, Nividia RTX 3060 8GB).

7 Conclusion

In this paper, we investigated the use of visual collision feedback to improve hand movement performance in confined-occluded VR spaces. Firstly, the relations between hand movement performance and visual collision feedback of different informativeness levels were studied. Based on our results, we proposed a promising visual collision feedback based on projector decals, which combines the advantages of see-through visualization and proximity-based coloring gradation to achieve an optimal compromise between hand movement speed and accuracy in a confined-occluded VR space.

In the experiment, we tested the hand movement performance under five commonly-used visual collision feedback conditions. An empirical informativeness metric was proposed to measure the spatial information content encoded in the feedback. Experimental results show that the informativeness level has a significant effect on the hand movement speed, however, the relation between them appeared to be non-linear. Specifically, the see-through visualization is advantageous in improving hand movement speed, while the hybrid of the proximity-based coloring gradation and deformation significantly slowed down the hand movement. In terms of accuracy, there was a clear trend that the more informative the feedback, the more accurate the hand movement. We further analyzed the impacts of collision information from the aspects of information details and understanding level. We found that location, distance, and direction information may be the key details to understand a collision, and the movement speed appears to be quicker in natural and implicit feedback environment, while slower in artificial and explicit feedback environment. Finally, we concluded that the integration of see-through visualization and the proximity-based visual cues might be the best compromise between the speed and accuracy of the hand movement in confined-occluded VR spaces.

We redesigned the visual collision feedback according to the findings, and proposed a promising feedback technology based on projector decal. The new technology is achieved by six projectors which cast the information cues about the collision states of five fingers and palm individually. The projector pose is updated by an adaptive algorithm, and collision information is displayed as a color gradation, which is sampled from a proximity texture based on the sampling coordinates calculated by a 2D Gaussian method. Demos of the proposed technology were presented, and the technology shows remarkable computing efficiency on a household gaming laptop.

Our results provide reference for understanding hand movement behavior in confined-occluded spaces and designing effective visual collision feedback. In future work, we will investigate the validity of the proposed informativeness metric, and further explore the use of visual feedback in more complex tasks, such as assembly simulation of complex gearbox.

Declaration

References

1. González-Alvarez, C., Subramanian, A., Pardhan, S.: Reaching and grasping with restricted peripheral vision. Ophthalmic Physiol. Opt. **27**(3), 265–274 (2007)

2. Fu, M.J., Hershberger, A.D., Sano, K., Çavuşoğlu, M.C.: Effect of visuomotor colocation on 3d fitts' task performance in physical and virtual environments. Presence **21**(3), 305–320 (2012)

3. Valentina, G.: The role of peripheral visual cues in planning and controlling movement:| ban investigation of which cues provided by different parts of the visual field influence the execution of movement and how they work to control upper and lower limb motion. PhD Thesis, University of Bradford. See also http://hdl.handle.net/10454/5715 (2013)

4. Liu, L., van Liere, R., Nieuwenhuizen, C., Martens, J.-B.: Comparing aimed movements in the real world and in virtual reality. In 2009 IEEE Virtual Reality Conference, IEEE, pp. 219–222 (2009)

5. Chapoulie, E., Tsandilas, T., Oehlberg, L., Mackay, W., Drettakis, G.: Finger-based manipulation in immersive spaces and the real world. In 2015 IEEE Symposium on 3D User Interfaces (3DUI), IEEE, pp. 109–116 (2015)

6. Zhang, L., Yang, J., Inai, Y., Huang, Q., Wu, J.: Effects of aging on pointing movements under restricted visual feedback conditions. Hum. Mov. Sci. **40**, 1–13 (2015)

7. Bell, J.D., Macuga, K.L.: Goal-directed aiming under restricted viewing conditions with confirmatory sensory feedback. Hum. Mov. Sci. **67**, 102515 (2019)

8. Bloomfield, A., Badler, N. I.: Collision awareness using vibrotactile arrays. In 2007 IEEE Virtual Reality Conference, IEEE, pp. 163–170 (2007)

9. Louison, C., Fabien, F., Daniel, R.M.: Spatialized vibrotactile feedback improves goal-directed movements in cluttered virtual environments. Int. J. Human-Comput. Int. **34**(11), 1015–1031 (2018)

10. Sagardia, M., Hulin, T., Hertkorn, K., Kremer, P., Schätzle, S.: A platform for bimanual virtual assembly training with haptic feedback in large multi-object environments. In Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology, Association for Computing Machinery, pp. 153–162 (2016)

11. Xia, P., Lopes, A.M., Restivo, M.T., Yao, Y.: A new type haptics-based virtual environment system for assembly training of complex products. Int. J. Adv. Manuf. Technol **58**(1), 379–396 (2012)

12. Lindeman, R. W., Page, R., Yanagida, Y., Sibert, J. L.: Towards full-body haptic feedback: the design and deployment of a spatialized vibrotactile feedback system. In Proceedings of the ACM symposium on Virtual reality software and technology, pp. 146–149 (2004)

13. Weber, B., Sagardia, M., Hulin, T., Preusche, C.: Visual, vibrotactile, and force feedback of collisions in virtual environments: effects on performance, mental workload and spatial orientation. In: Int. Conf. Virtual, pp. 241–250. Springer, Augmented and Mixed Reality (2013)

14. Lawson, G., Salanitri, D., Waterfield, B.: Future directions for the development of virtual reality within an automotive manufacturer. Appl. Ergon. **53**, 323–330 (2016)

15. Zimmermann, P.: Virtual reality aided design. a survey of the use of vr in automotive industry. *Product Engineering*, pp. 277–296 (2008)

16. Sreng, J., Lécuyer, A., Mégard, C., Andriot, C.: Using visual cues of contact to improve interactive manipulation of virtual objects in industrial assembly/maintenance simulations. IEEE Trans. Visual Comput. Graphics **12**(5), 1013–1020 (2006)

17. Prachyabrued, M., Borst, C. W.: Visual feedback for virtual grasping. In 2014 IEEE Symposium on 3D User Interfaces (3DUI), IEEE, pp. 19–26 (2014)

18. Vosinakis, S., Koutsabasis, P.: Evaluation of visual feedback techniques for virtual grasping with bare hands using leap motion and oculus rift. Virtual Reality **22**(1), 47–62 (2018)

19. Samad, M., Gatti, E., Hermes, A., Benko, H., Parise, C.:Pseudo-haptic weight: changing the perceived weight of virtual objects by manipulating control-display ratio. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2019)

20. Vance, J. M., Dumont, G.: A conceptual framework to support natural interaction for virtual assembly tasks. In World Conference on Innovative Virtual Reality, Vol. 44328, American Society of Mechanical Engineers, pp. 273–278 (2011)

21. Sarlegna, F. R., Sainburg, R. L.: The roles of vision and proprioception in the planning of reaching movements. *Progress in motor control*, pp. 317–335 (2009)

22. Shumway-Cook, A., Woollacott, M.H.: Motor control: translating research into clinical practice. Lippincott Williams & Wilkins, US (2007)

23. Waller, D., Hodgson, E.: Sensory contributions to spatial knowledge of real and virtual environments. In *Human walking in virtual environments*. Springer, pp. 3–26 (2013)

24. Ng, A.W., Chan, A.H.: Finger response times to visual, auditory and tactile modality stimuli. In Proceedings of the International Multiconference of Engineers and Computer Scientists **2**, 1449–1454 (2012)

25. Fan, H., Zhuo, T., Yu, X., Yang, Y., Kankanhalli, M.: Understanding atomic hand-object interaction with human intention. *IEEE Trans. Circuits. Syst. Video Technol.* (2021)

26. Elmqvist, N., Tsigas, P.: A taxonomy of 3d occlusion management for visualization. IEEE Trans. Visual Comput. Graphics **14**(5), 1095–1109 (2008)

27. Iachini, T., Ruotolo, F., Vinciguerra, M., Ruggiero, G.: Manipulating time and space: collision prediction in peripersonal and extrapersonal space. Cognition **166**, 107–117 (2017)

28. Fricke, N., Thüring, M.: Complementary audio-visual collision warnings. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 53, SAGE Publications Sage CA: Los Angeles, CA, pp. 1815–1819 (2009)

29. Biocca, F., Kim, J., Choi, Y.: Visual touch in virtual environments: an exploratory study of presence, multimodal interfaces, and cross-modal sensory illusions. *Presence: Teleoperators & Virtual Environments,* **10**(3), pp. 247–265 (2001)

30. Cortes, G., Argelaguet, F., Marchand, E., Lécuyer, A.: Virtual shadows for real humans in a cave: Influence on virtual embodiment and 3d interaction. In Proceedings of the 15th ACM Symposium on Applied Perception, pp. 1–8 (2018)

31. Eren, M.T., Balcisoy, S.: Evaluation of x-ray visualization techniques for vertical depth judgments in underground exploration. Vis. Comput. **34**(3), 405–416 (2018)

32. Ariza, O., Bruder, G., Katzakis, N., Steinicke, F.: Analysis of proximity-based multimodal feedback for 3d selection in immer-

sive virtual environments. In 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), IEEE, pp. 327–334 (2018)

33. Dey, S., Kapoor, A.: Hand length and hand breadth: a study of correlation statistics among human population. Int. J. Sci. Res. **4**(4), 148–150 (2015)

34. Andreas, B., L, N. S., Mikkel, G., D, L. B., Jørgen, C. N.: Single-pass wireframe rendering. In *ACM SIGGRAPH 2006 Sketches*. pp. 149–es (2006)

35. UltraLeap, 2017. Hand — leap motion c# sdk v2.3 documentation. On the WWW, December. https://developer-archive.leapmotion.com/documentation/v2/csharp/api/Leap.Hand.html?proglang=csharp

36. UltraLeap, 2010. Unity modules of leap motion. On the WWW. https://leapmotion.github.io/UnityModules/

37. UltraLeap, 2020. Unity modules leap motion's unity sdk 4.5.0: Interaction engine, leap motion. On the WWW, May. https://leapmotion.github.io/UnityModules/interaction-engine.html

38. F, B., Yasuhiro, I., Andy, L., Heather, P., Arthur, T.: Visual cues and virtual touch: Role of visual stimuli and intersensory integration in cross-modal haptic illusions and the sense of presence. *Proceedings of presence*, pp. 410–428 (2002)

39. Zhai, S., Accot, J., Woltjer, R.: Human action laws in electronic virtual worlds: an empirical study of path steering performance in vr. Presence **13**(2), 113–127 (2004)

40. MacKenzie, I.S.: Fitts' law as a research and design tool in human-computer interaction. Human-Comput. Interact. **7**(1), 91–139 (1992)

41. Kawabe, T.: Mid-air action contributes to pseudo-haptic stiffness effects. IEEE Trans. Haptics **13**(1), 18–24 (2019)

42. Li, M., Konstantinova, J., Secco, E.L., Jiang, A., Liu, H., Nanayakkara, T., Seneviratne, L.D., Dasgupta, P., Althoefer, K., Wurdemann, H.A.: Using visual cues to enhance haptic feedback for palpation on virtual model of soft tissue. Med. Biol. Eng. Comput. **53**(11), 1177–1186 (2015)

43. Argelaguet, F., Hoyet, L., Trico, M., Lécuyer, A.: The role of interaction in virtual embodiment: Effects of the virtual hand representation. In 2016 IEEE Virtual Reality (VR), IEEE, pp. 3–10 (2016)

44. Schwind, V., Lin, L., Di Luca, M., Jörg, S., Hillis, J.: Touch with foreign hands: The effect of virtual hand appearance on visual-haptic integration. In Proceedings of the 15th ACM Symposium on Applied Perception, pp. 1–8 (2018)

45. Dienes, Z., Perner, J.: A theory of implicit and explicit knowledge. Behav. Brain Sci. **22**(5), 735–808 (1999)

46. Greene, E. D.: Augmenting visual feedback using sensory substitution. Master's thesis, University of Waterloo (2011)

47. Reddy, G. R., Rompapas, D. C.: Visuotouch: Enabling haptic feedback in augmented reality through visual cues. In IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (2020)

Shouwen Yao received his Ph.D degree from China Academy of Railway Sciences, Beijing, China, in 2000. He has been working as a professor in Beijing Institute of Technology (BIT) since 2002. He is also Professor and Director of Scene Simulation Lab at BIT. His current research interests include virtual reality aided assemblability evaluation, 3D point-cloud reconstruction with RGB-D sensors and virtual reality aided simulation of UGV, etc

Yu Wang is currently pursuing his PhD degree in mechanical engineering at School of Mechanical Engineering, Beijing Institute of Technology, Beijing, China, where he received his MS degree in vehicle engineering in 2016. He joined HIT Lab NZ at UC as a research assistant from 2018-2019. His current research interests include theory, method and application of VR-aided industrial design, natural interaction technology, marker-less full-body tracking and avatar control in VR.

Ziran Hu received her Bachelor in 2019 and M.S. degree in 2021 from Beijing Institute of Technology, Beijing, China. Her research interests include novel user interface for VR-aided ergonomics evaluation and human-computer interaction.

Hui Liu received her Ph.D degree of Mechanical Engineering from Beijing Institute of Technology (BIT), Beijing, China, in 2003. She has been working as a professor in BIT since 2003. She is also the director of Vehicle Engineering Department of the School of Mechanical Engineering of BIT. Her current research interests include vehicle dynamics and electromechanical transmission for hybrid electric vehicles.