Free hands interaction in augmented reality

Conference Paper · July 2013

DOI: 10.1145/2491367.2491370

2 authors:

Dragos Datcu
Delft University of Technology

Stephan Lukosch
Delft University of Technology

International Journal of Human-Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: http://www.tandfonline.com/loi/hihc20

On the Usability and Effectiveness of Different Interaction Types in Augmented Reality

Dragoş Datcu, Stephan Lukosch & Frances Brazier

To cite this article: Dragoş Datcu, Stephan Lukosch & Frances Brazier (2015) On the Usability and Effectiveness of Different Interaction Types in Augmented Reality, International Journal of Human-Computer Interaction, 31:3, 193-209, DOI: 10.1080/10447318.2014.994193

To link to this article: http://dx.doi.org/10.1080/10447318.2014.994193

Intl. Journal of Human–Computer Interaction, 31: 193–209, 2015

ISSN: 1044-7318 print / 1532-7590 online
DOI: 10.1080/10447318.2014.994193

Taylor & Francis
Taylor & Francis Group

On the Usability and Effectiveness of Different Interaction Types in Augmented Reality

Dragoş Datcu, Stephan Lukosch, and Frances Brazier
Technology, Policy and Management, Delft University of Technology, Delft, the Netherlands

One of the key challenges of augmented reality (AR) interfaces is to design effective hand-based interaction supported by computer vision. Hand-based interaction requires free-hands tracking to support user interaction in AR for which this article presents a novel approach. This approach makes it possible to compare different types of hand-based interaction in AR for navigating using a spatial user interface. Quantitative and qualitative analyses of a study with 25 subjects indicate that tangible interaction is the preferred type of interaction with which to determine the position of the user interface in AR and to physically point to a preferred option for navigation in augmented reality.

1. INTRODUCTION

Over recent years, augmented reality (AR) has been extensively researched with remarkable results for its application in various fields such as entertainment, maintenance, design, medical domain, crime scene investigation (CSI), for learning, real-life testing, and simulation. Viable solutions for hand motion and gesture interaction are provided in both low-level (Marco, Cerezo, & Baldassarri, 2012; Mateu, Lasala, & Alamán 2014; Xu, Read, Mazzone, & Brown, 2007) and high-level education purposes (Chen, Chi, Hung, & Kang, 2011), biology (Ness et al., 2010), assembly (Radkowski & Stritzke, 2012; Wiedenmaier, Oehme, Schmidt, & Luczak, 2003) and chemical engineering, architectural projects (Broll et al., 2004; Gu, Kim, & Maher, 2011; Nagel & Heidmann, 2011), wastewater treatment plant control (Bertelsen & Nielsen, 2000), entertainment (Newton-Dunn, Nakano, & Gibson, 2003; Piumsomboon, Clark, & Billinghurst, 2011; Schiettecatte & Vanderdonckt, 2008), home (Smith, 1995), office (Ehnes, 2009), shopping (Merrill & Maes, 2007), and military (Livingston, Ai, Karsch, & Gibson, 2011) environments.

This article addresses one of the major challenges in AR, namely, to provide a generally reliable solution for hand-based interaction within AR supported by computer vision. The use of computer vision does not necessarily restrict the applicability of AR interaction by hand gestures. On the contrary, computer vision is often embedded in devices for AR interaction such as, for example, the visible spectrum camera found in mobile devices, including Head Mounted Devices (HMDs).

In the broad context of AR interaction, it is often essential that AR users can easily point to or select small details in the augmented view (Billinghurst & Thomas, 2011). These details may refer to menu items and other real or augmented objects that are part of the AR interface. Recent research (Datcu, Swart, Lukosch, & Rusak, 2012) reveals the limitations of automatic hand registration using computer vision methods and indicates the need for further improvement on menu-based interaction types for AR navigation.

In previous research on CSI with the Netherlands Forensic Institute (Datcu et al., 2012; Poelman, Akman, Lukosch, & Jonker, 2012; CSI The Hague, 2012; http://www.forensicinstitute.nl/) an interface with *free-hand gestures for user interface operation for mobile users with HMDs*, supported by computer vision technology,[1] similar limitations have been recognized, specifically

1. The system's hand gesture recognition is not precise enough, impeding accurate selection of the icons in the menu interface (Poelman et al., 2012).
2. Users explicitly stated interest in the use of everyday objects for interaction within the AR system (Swart, 2012).
3. The positioning of the user interface in the user's view is considered disturbing, especially when a larger set of menu icons is displayed at once.

This article addresses the aforementioned aspects. It introduces a novel approach for free-hands tracking to support user interaction in AR that (a) provides robust and precise hand detection and tracking in conditions of varying illumination and limited occlusions, (b) supports the selection of the menu items using free-hand gestures, and (c) makes it possible to use a hand as a 3D pointing device. This approach enables the comparison

Address correspondence to Dragoş Datcu, Multi-Actor Systems/Systems Engineering, Delft University of Technology, Jaffalaan 5, Delft, 2628BX, the Netherlands. E-mail: D.Datcu@tudelft.nl

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hihc.

[1]No additional hardware equipment such as data gloves were used. Voice-based interfaces were also out of consideration due to their limited performance in noisy environments.

of free-hand gesture interaction with other types of interaction. This article describes and evaluates the design of five different interaction types for navigating within a spatial user interface together with an evaluation of their usability and effectiveness for interacting with in AR.

This article is structured as follows: Section 2 discusses background literature and the motivation for this article. Section 3 presents the hand-tracking approach for free-hand interaction. Section 4 presents the five types of interaction and provides details on their implementation emphasizing concepts of interaction and system architecture to support interaction with physical objects. Section 5 describes the experimental setup for the evaluation of the usability of the five interfaces. Section 6 presents the quantitative and qualitative results of the study. Section 7 concludes with a discussion on future research.

2. BACKGROUND

Accurate hand-tracking capabilities in AR are encountered in designs of local systems, that is, systems that do not require the user to move around, often use sensors mounted as part of the environment, have a relatively small set of events to handle, and benefit from controlled view and lighting conditions when supporting hand-gesture-based interaction (Chen et al., 2011; Cheng, Liang, Chen, Liang, & Kuo, 2010; Ehnes, 2009; J. Y. Lee, Rhee, & Seo, 2010; Maier et al., 2010; Marco et al., 2012; Nagel & Heidmann, 2011; Ness et al., 2010; Radkowski & Stritzke, 2012; White, Feng, & Feiner, 2009; Xu et al., 2007). First-Person Vision leveraged by mobile video cameras is currently considered the most optimal way to sense an environment and a subject's activities from a wearable sensor, as indicated by Kanade and Hebert (2012).

First-Person Vision–based pointing that uses computer vision technologies still requires attention from researchers working on hand tracking and hand gesture recognition. Designing hand tracking in AR with the same level of expressiveness and naturalness as in real life human communication is a research endeavor involving both technical and usability challenges (Kölsch, Turk, Hollerer, & Chainey, 2004).

Hand-based interaction in AR, for example, often requires users to perform specific hand and finger gestures like tapping a marker (Cheng et al., 2010; Nagel & Heidmann, 2011), clicking on a physical object (Newton-Dunn et al., 2003), showing gestures of fist (Radkowski & Stritzke, 2012) or palm (Poelman et al., 2012) postures, pushing with the fingertip and pinching (Nagel & Heidmann, 2011; Piumsomboon et al., 2011), rotating and dragging the physical object (Cheng et al., 2010; Ehnes, 2009; Ness et al., 2010), positioning physical cubes relative to each other and moving them around (Schiettecatte & Vanderdonckt, 2008). Several types of gestures have been studied but not in relation to others. Some AR systems limit the interaction for menu navigation to using one hand only (Merrill & Maes, 2007; Radkowski & Stritzke, 2012). Other AR systems are designed to support two hands interaction. Even though such systems can handle more complex operations and have shown to be 3 times faster than standard interaction with a computer and mouse (Reifinger, Wallhoff, Ablassmeier, Poitschke, & Rigoll, 2007), the question of whether one hand interaction is more effective for AR menu navigation still remains.

One characteristic of our previous AR system (Poelman et al., 2012) is the interaction based on bare hands (López, López, Guerrero, & Bravo, 2014; Mateu et al., 2014; Radkowski & Stritzke, 2012). Other AR systems use everyday physical objects for hand-based interaction to provide tangible feedback and simplify gesture input to the user (Henderson & Feiner, 2008). Actions applied on objects or blocks reported in the literature include selection, gripping, cutting, copying and pasting (J. Y. Lee et al., 2010), pushing and pulling (Piumsomboon et al., 2011), grasping (Radkowski & Stritzke, 2012), zooming and panning a map (Nagel & Heidmann, 2011), typing in open documents (Ehnes, 2009), and generating audio content (Ness et al., 2010; Newton-Dunn et al., 2003; Schiettecatte & Vanderdonckt, 2008) or vibro-tactile feedback (J. Y. Lee et al., 2010). Such tangible interaction with physical objects provides higher engagement (Chen et al., 2011; Mateu et al., 2014) and a more natural feeling when manipulating physical objects. Interaction with two hands and a physical object also minimizes the interaction errors and provides more immersive and natural feeling in AR (J. Y. Lee et al., 2010).

Often not only one object but two objects are used in hand-based interaction with physical devices. A polyhedral object enhances interaction on a table surface by selecting and filtering various metadata facets (Nagel & Heidmann, 2011). A piece of paper becomes an active device to help select and navigate the AR interface (White et al., 2009). Users can interact with everyday objects and stylus on a tabletop surface to build their own physical world (Jones, Sodhi, Campbell, Garnett, & Bailey, 2010). Tangible interaction on table surfaces is used for document handling, sharing, and keyboard forwarding (Datcu et al., 2012). Tangible objects on a flat surface can be used to combine operational properties such as location, movement, arrangement, and layout (Ness et al., 2010; Newton-Dunn et al., 2003; Schiettecatte & Vanderdonckt, 2008) or to reinforce physical manipulation and colocated interaction design in tabletop-oriented AR games (Marco et al., 2012; Xu et al., 2007). Summarizing, tangible, graspable, and movable everyday objects can be easily transformed into controllers for AR leading to seamless interaction in AR environments (Cheng et al., 2010; M. Lee, Green, & Billinghurst, 2008).

The display data in AR menus generally include text entries, images, icons, or 3D objects (Dachselt & Hübner, 2007), which follow different types of alignment, relative to the world, object, head, body, or the device (Bowman, Kruijff, LaViola, & Poupyrev, 2004). The world-fixed alignment traces its roots in classical desktop systems and is still commonly used in current AR/VR systems.

The review just presented shows that there is a huge variety of interaction types but that there is no evaluation on the usability and effectiveness of different interaction types in a uniform environment. Without losing generality and validity of the findings, a CSI scenario from our previous research is used to illustrate the hand-tracking approach and as the domain of application for the comparative study.

3. HAND-TRACKING APPROACH

From the functional point of view, hand segmentation and tracking can be done by skin color segmentation (Piumsomboon et al., 2011; Shen, Ong, & Nee, 2011), depth information (Mateu et al., 2014; Newton-Dunn et al., 2003; Poelman et al., 2012), markers (Jones et al., 2010) or colored markers attached to user fingertips (Newton-Dunn et al., 2003).

A large number of systems use sensors others than visible spectrum cameras, due to their improved performance in tracking the user and the physical objects of interaction. Sensors tracking body, hand, and finger movements such as Microsoft Kinect ("Kinect," n.d.), Asus WAVI Xtion, or Leap Motion (http://en.wikipedia.org/wiki/Leap_Motion) have opened and continue to expand the horizon of possibilities for interaction (Merrill & Maes, 2007). For such systems, the natural interaction in AR is improved by the use of accurate location information generated by the depth sensor.

This article focuses on systems in which users are mobile and free to move around. The approaches just discussed are not applicable to this situation, due to lack of extra sensors and changing lighting conditions. The hand-processing methods proposed in the article are based on robust video analysis only. The 2D hand detection is handled by a Viola&Jones object detector (Viola & Jones, 2002) and the hand contour detector is based on Active Appearance Models (AAM; Cootes, Edwards, & Taylor, 2001). The tracking of hands is done using Lukas-Kanade optical flow (Lucas & Kanade, 1981) running on AGAST visual features (Mair, Hager, Burschka, Suppa, & Hirzinger, 2010). This novel approach that uses AAM on hand objects and Lukas-Kanade optical flow for tracking shows robustness and accuracy (Datcu & Lukosch, 2013).

For the purpose of hand tracking, this article distinguishes three different hand postures for both the dominant hand and the nondominant hand as depicted in Figure 1. A detection model for palm hand posture is assigned to the nondominant hand to control the menu positioning. A detection model for fist hand posture is assigned to the dominant hand to control the cursor (Figure 1c).[2]

Touchless triggers are linked to the event of holding the hand or the physical object over sensitive areas for a prespecified amount of time. For free hands interaction, the hand postures

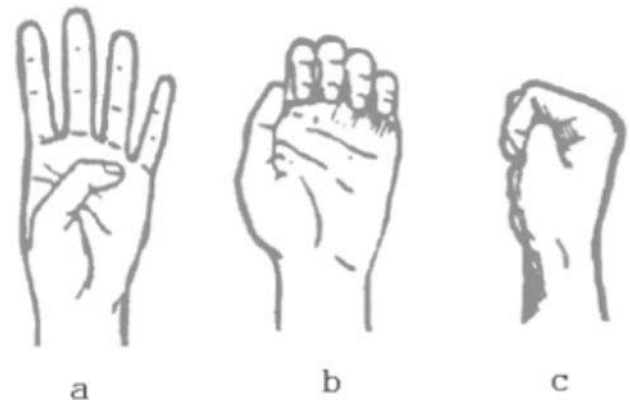


FIG. 1. Hand postures considered for the interaction in augmented reality.

mapped to the nondominant hand trigger repositioning of the menu. The user can display the hand postures of the dominant hand to trigger repositioning of the cursor and selection of the menu icons. Navigation in the menu require a sequence of actions that involve gestures of one hand or two hands with or without physical objects.

To test the hand detection for pointing, a single user performed menu navigation and menu item selection in AR. The user wore a colored marker on his fingertip. The data generated by the video cameras of the HMD are stored for later offline processing. To test the accuracy of pointing, Viola&Jones and AAM algorithms are applied on 200 frames randomly selected from the original video sequence.

As a first step, an automatic color detector determines and tracks the location of the color marker in each frame. This information serves as the ground truth for the actual test. All results are manually checked to ensure that the color marker is detected correctly in all test frames.

Second, Viola&Jones hand detection algorithm is run automatically for all frames. The result consists of rectangular areas identifying hand regions (red rectangles in Figure 2). Given a rectangular area, the location of the fingertip is approximated at a relative location within the hand rectangle (yellow circles in Figure 2).

Third, the location of the tip of the finger is computed as the average location of three AAM key points at the tip of the pointer finger of the dominant hand (blue circles in Figure 2). The same figure depicts the AAM shapes with green line segments and key points along each hand contour.

Figure 2 shows errors for estimating the location of the fingertip (a) by Viola&Jones approximation (blue line segment) and (b) by AMM (orange line segment). Figure 2a shows an example for which both Viola&Jones and AAM provide results close to the ground truth of the pointer finger tip of the dominant hand. Viola&Jones pointing errors are small (up to 20 pixels) when the hand is not angled and the detected rectangular area matches the real dimensions of the hand. Figure 2b, c, and d

[2]Initially, fist hand posture detector was mapped to the nondominant hand to control the location of the menu. The palm hand posture detector was mapped to the dominant hand to control the position of the on-screen cursor. Feedback from testers in the preexperiment session showed that this setup is highly uncomfortable and tiresome. This was also caused by small field of view limiting the work volume for the hands and imposing for hands rather fully stretched to interact properly with the AR system.

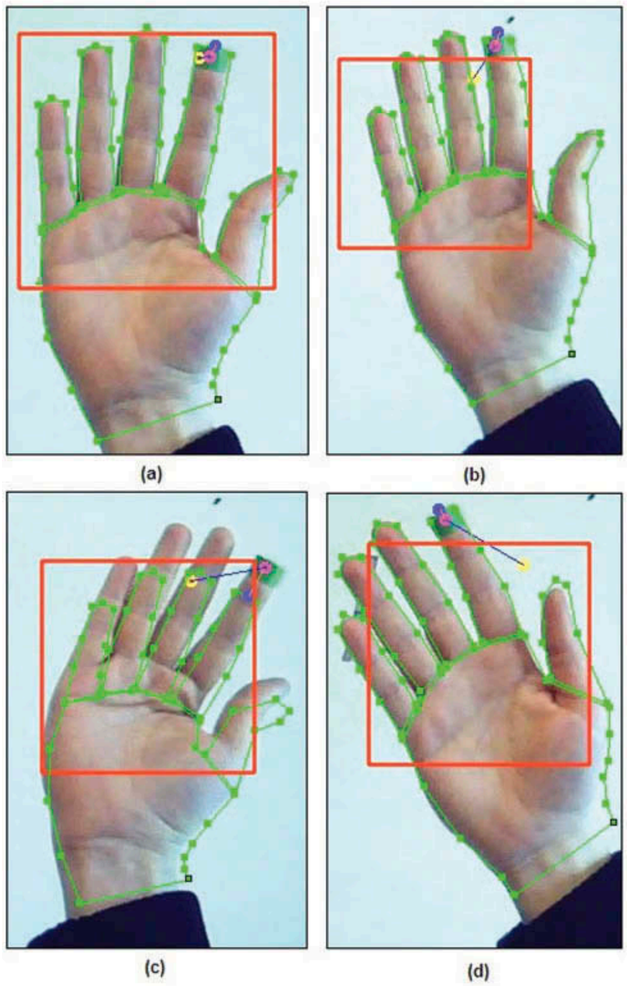


FIG. 2. Examples of palm detection and shape extraction.

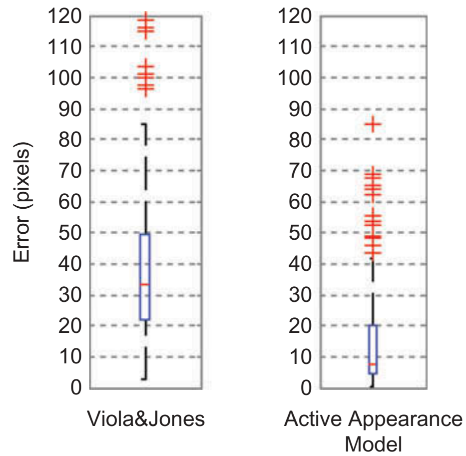


FIG. 3. Box plots showing finger pointing errors using Viola&Jones (left) and Active Appearance model (right).

show that rotated hand poses generally lead to high approximation errors and wrong pointing results with Viola&Jones only. In turn, deformable models like AAM are able to extract the contour of the palm, even in conditions of limited rotation, scaling, and translation, relative to the initialization step provided by the Viola&Jones model. Depending on the initialization, generalization in modeling palm appearance and strategy of updating model parameters, some cases are still difficult to handle with both Viola&Jones and AAM. An example of wrong hand shape extraction by AAM is showed in Figure 2c.

As shown in the box plots from Figure 3, the median error by AAM is 7.76 pixels. This level of error is acceptable for handling free-hands pointing in AR by only computer vision methods. The median error of the Viola&Jones model is 33.02 pixels, more than 4 times bigger than in case of the AAM model. This result clearly indicates the superiority of AAM over Viola&Jones palm detector for pointing estimation. The results indicate that computer vision-driven hand models are robust for free-hand pointing in AR.

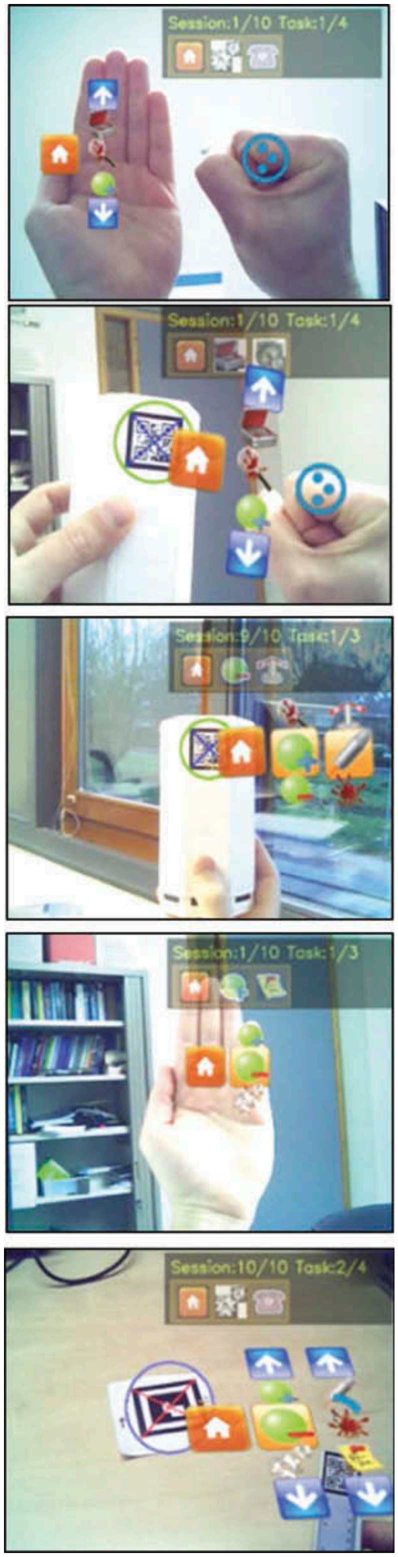
4. INTERACTION DESIGN

In the physical world, switches and buttons are commonly integrated in an environment and used to trigger specific events. They involve a mechanical process that involves natural tactile feedback. Triggering an event in AR can follow a similar course, with touchless interaction.

The five types depicted in Figure 4 represent a broad range of free-hand and tangible interactions that are representative also for other AR systems found in the literature. These types of interaction are explored and compared. For all types of interaction, a menu interface is activated and deactivated by specific triggers coupled to the presence and absence of a hand or of a physical object. The set of interaction cases analyzed is a representative selection of interaction capabilities in AR for a large range of application contexts in general and for the CSI scenario in particular.

The five interaction types are depicted in Figure 4. Interaction types (a) and (d) require free-hand interaction, whereas interaction types (b), (c), and (e) require interaction by hand and physical objects. The dominant hand is used to control the location of the augmented cursor when using two hands (Figure 4a), when interacting with a physical object and two hands (Figure 4b), and when interacting with one hand and two objects, by using a so-called magic wand (Figure 4e). For all five interaction types, the dominant hand navigates the menu and to select menu items.

The first free-hand interaction type assumes the use of one hand (Figure 4d), and the second assumes the use of two hands (Figure 4a). Both interaction types use the palm of a user's hand as a projection surface for the menu interface. The other three interaction types (Figure 4b, c, and e) use the surface of a physical object to project the menu. The interaction type in Figure 4e uses a flat surface as projection surface and is the only type that makes use of the local physical environment as means of



(a) Two hands interaction — Nondominant hand controls the position of a menu and dominant hand controls the position of the cursor.

(b) Interaction with two hands and object — Nondominant hand holds a physical object and dominant hand controls the position of the cursor.

(c) Interaction with one hand and object — A hand (either dominant or nondominant) holding a physical object controls menu navigation and selection.

(d) One hand interaction — Nondominant hand gesture controls menu navigation and selection.

(e) Interaction with one hand and two objects — One hand (either dominant or nondominant) using a physical object (ruler Figure 6, right side) controls the location of the cursor; the menu is displayed on the surface of a second object (e.g., a table or wall).

FIG. 4. Interaction types.

interaction. The interaction types in Figure 4b and Figure 4c use a physical object resembling an everyday object as projection surface. The augmented menu is either projected on a fixed area in the view of the user or follows the position of the hand or of the physical object.

The five types depicted in Figure 4 cover a broad range of free-hand and tangible interactions, which are representative also for other AR systems found in the literature. In our study without losing generality, we opted for a menu as visual projection and part of the user interface. The same navigation and selection triggers can be easily mapped to control also other interfaces and representations for handling educational, entertainment, or complex industrial systems, similar to the interfaces discussed in Section 2.

For the interaction type displayed in Figure 4a, the nondominant hand controls the position of a menu and the dominant hand controls the position of the cursor. By moving and keeping the cursor on a menu item, the menu item will be selected and activated. Selecting Up/Down arrow icons allows a user to scroll through the icons on the same menu level. Selecting a nonterminal icon allows for scrolling through icons on different menu levels. The first selection of a nonterminal icon expands the menu with the icons at the next depth level in the menu. Similarly, a second selection of a nonterminal icon when its icons associated to the next depth level are already displayed triggers the collapse of the menu by hiding the icons associated to the selection icon. The home icon allows a user to return to the main menu. The menu disappears automatically after a few seconds from the moment the nondominant hand was detected last time or after some period without interaction involving the dominant hand.

For the type of interaction depicted in Figure 4b, the nondominant hand holds a physical object and dominant hand controls the position of the virtual cursor. The nondominant hand indirectly controls the position of the menu, given the system marker-based detection of the object. The scroll through the menu follows a similar course as for the interaction presented in Figure 4a by selecting Up/Down arrow icons and the nonterminal icons. The menu is automatically hidden after a few seconds after the object was last detected or after a period within which there has been no interaction by the dominant hand.

The interaction type illustrated in Figure 4c requires one hand (either dominant or nondominant) holding a physical object to control menu navigation and selection. The navigation through different menu levels is done by a motion pattern of rotating the object toward left or right on its vertical axis and quickly returning to the original position. The rotation toward right on a nonterminal icon expands the menu and the rotation toward the left on a nonterminal icon returns to the previous depth level by hiding the icons associated to the previously selected nonterminal icon. Navigation through icons on the same menu level is done by a motion pattern of shifting the object upward or downward and quickly returning to the original position. The selection event is triggered when a user applies the motion pattern to the right while positioned on an icon on the terminal depth level. The menu is automatically hidden after a few seconds after the object was last detected or after a period within which there has been no interaction by the dominant hand.

For the interaction type shown in Figure 4d, the nondominant hand gesture controls menu navigation and selection. Navigation and selection in the menu is done by motion patterns of shifting the hand in one of the four directions and quickly returning to the original position. Navigation through icons on the same depth level is done by motion patterns upward and downward. To move to icons on the next depth level, a user applies a motion pattern to the right and a quick return to the original position, the action being succeeded by displaying the icons associated to the previously selected nonterminal icon. To move back to icons on the previous depth level, a user applies a motion pattern to the left and a quick return to the original position, the action being succeeded by hiding the set of icons associated to the current nonterminal icon. The menu is automatically hidden after a few seconds from the moment the hand was last detected or after a period without interaction.

The interaction types in Figure 4c and Figure 4d do not require handling a virtual cursor for navigation and selection, these two events being triggered by the execution of motion patterns and direct control of object and hand.

Figure 4e illustrates an interaction type that requires one hand (either the dominant or nondominant) holding a physical object to control the location of the cursor. One object is represented by a flat surface in the physical space of the user (i.e., a table) and a second object is used as a *magic wand*. The user uses the *magic wand* to interact with a menu that is associated to the table object. The navigation and selection follow similar actions as for the interaction types illustrated in Figure 4a and Figure 4b, with the difference that the virtual cursor is controlled through an object ("magic wand") and not by the free hand. The menu is automatically hidden after a few seconds from the moment after the object was last detected or after a period within which there has been no interaction by the dominant hand.

To instruct future research and development on interaction types in AR, the comparison of the interaction types in this article investigates the following hypotheses:

H1: Free-hand interaction is more effective than tangible interaction.

H2: One-hand interaction is more effective than two-hand interaction.

Although navigation tasks in world-fixed menu systems indicate the shortest average time and the smallest error rates, they lack intuitiveness and preference among users (White et al., 2009). The rather high discrepancy between pro and contra arguments further points to the following hypothesis:

H3: Fixed menu user interface (UI) layouts allow for better interaction than non-fixed UI layouts.

Hypothesis 1 is tested by comparing the interaction types in Figure 4a and d with the interaction types depicted in Figure 4b, c, and e. More, Hypothesis 2 is tested by comparing the interaction types in Figure 4c, d, and e with the interaction types depicted in the Figure 4a and b.

4.1. Tangible Interaction

Users of AR systems have expressed interest in using everyday objects to support interaction (Swart, 2012). Capabilities for tangible interaction in AR can be assigned based on the affordances of these physical objects. A flashlight or ruler could, for example, be used as a pointing device, whereas a container could be used as a dial. The interaction types in Figure 4b, c, and e are designed with regard to Hypothesis 1 for this purpose.

A prerequisite for interacting with physical objects in AR is a system's capability to accurately detect and track user's hand and objects. If a user is holding a container object (Figure 4b and c) or a ruler object (Figure 4e) in his or her hand, the interaction is determined based on spatially tracking the objects and recognizing the patterns of motion. The table object in the situation of interacting with one hand and two objects is stationary (Figure 4e), therefore only detection and tracking of the object are considered.

To facilitate the detection and tracking for the types illustrated in Figure 4b, c, and e, markers are placed directly on the surface of the physical objects. The interaction with a flat surface (Figure 4e) is based on the *magic wand* concept (Figure 5, left) that enables the user to control the cursor to select menu items. The magic wand emulates regular physical objects like a ruler or a pointer. In the case of the magic wand, the selection of menu items is done by holding the augmented cursor over the surface of the item for a specific amount of time.

Two interaction types (Figure 4b and c), are based on the container dial concept (Figure 5, right) for which a menu is displayed onto and around the object using an AR overlay. The navigation and selection is done by moving or rotating the physical object that scrolls up or down one menu item in the menu hierarchy.

FIG. 6. The design of the physical object (Swart, 2012): cylinder-shape container (left) and ruler (right).

The current 3D object analysis techniques are unable to do real-time reconstruction, detection, and tracking of physical objects. As a consequence, tangible interfaces most often detect objects of interaction using markers (Billinghurst & Thomas, 2011; Chen et al., 2011; Datcu & Lukosch, 2013; Jones et al., 2010; Lucas & Kanade, 1981; bin Mohd Sidik, bin Sunar, bin Ismail, bin Mokhtar, & Jusoh, 2011; Nagel & Heidmann, 2011; Xu et al., 2007). To reduce the complexity of the problem, this study uses a generalized concept of physical objects in AR applications. Figure 6 shows a generic physical object that can take different physical shapes and can be easily detected by computer vision methods by simple markers attached to its different surfaces.

4.2. User Interface Alignment/Positioning

A graphical user interface enhances the view of the physical world with an augmentation using two different visual elements, namely, a cursor and a menu. The augmented elements are rendered in stereo AR view via an HMD. The stereo view for visualization enables the user to spatially perceive the augmented content.

In the presented study, the menu contains icons for the CSI domain, which relate to typical actions of the field investigator.

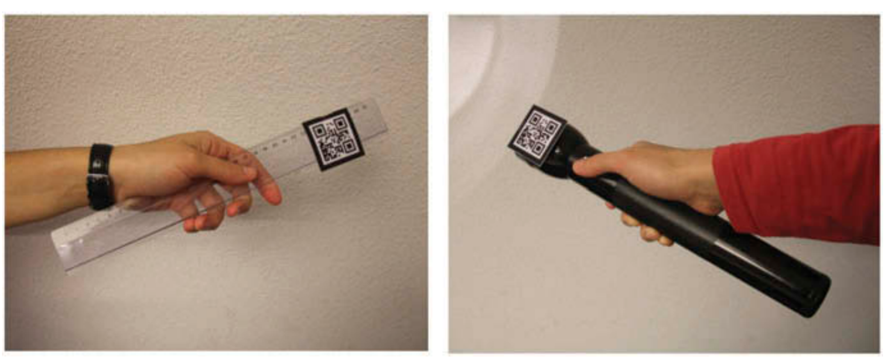


FIG. 5. Magic wand concept (left). Container dial concept (right).

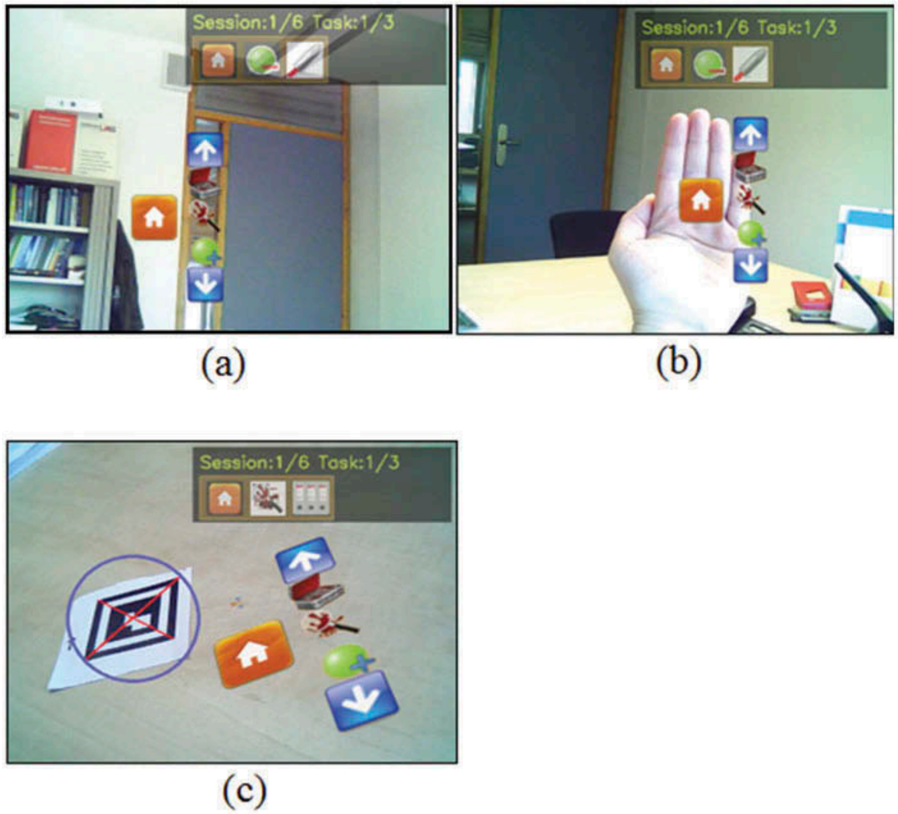


FIG. 7. Types of menu alignment: View—menu aligned with screen only, does not change position or orientation when object or user's hand moves (a); 2D Hand/object—menu changes 2D position when object or user's hand moves (b); 3D world—changes 3D position and orientation when object or user moves (c).

Such menu icons trigger functions such as taking photos and recording videos, running bloodstain pattern analysis, bullet trajectory ballistics, for manual annotation, for tagging physical evidence at the crime scene and for 3D scanning of the physical environment. The icons are positioned in the menu in a structured way, on several levels in depth so as to fit groups of similar actions and to facilitate their localization by the user. In addition to the domain specific icons, there are icons adapted for menu navigation, such as for scrolling up and down on the list of icons on the same level of depth.

Three different types of menu alignments are designed to cover a wide range of accessibility options with regard to the hand interaction with the menu interface. These are (a) screen alignment, (b) 2D object alignment, and (c) 3D world coordinate alignment (Figure 7). A similar approach has been reported in Xu et al. (2007).

View alignment (Figure 7a) requires that the menu is rendered in a fixed position in the AR view, during a whole work session, disregarding the actual position of either hand or physical object. View alignment has the advantage that the menu is visible, independent to the position and orientation of the hand or of the physical object. 2D alignment (Figure 7b) assumes that the menu follows the location of the hand or of the physical object in the AR view. 3D world alignment (Figure 7c) synchronizes the menu both with the 3D position and 3D orientation of the user's hand or of the physical object.

2D and 3D alignments are designed in particular for situations in which a user needs to access menu items located at deep levels in a menu hierarchy. The repositioning is done to support access to elements of interest in the interface like navigation arrows and the menu item to be selected. The user can reposition the whole menu using the nondominant hand and can proceed with the icon selection using the dominant hand.

Hypothesis 3 is tested by comparing the fixed menu UI layout (the view alignment condition in Figure 7a) with the nonfixed UI layout conditions depicted in Figure 7b (2D alignment) and Figure 7c (3D alignment).

5. STUDY DESIGN

To compare the five different interaction types just explained, with respect to usability and effectiveness and to test the three hypotheses, the following instruments are used:

1. A series of experiments to compare the performance of subjects in selecting specific menu items, measuring the time necessary to accomplish a specific task and checking the mistakes made while performing a task.

TABLE 1
Augmented Reality Interface Elements: Icon Representation and Examples of Tasks

Examples of Selection Tasks per Test Session		Home, Up/Down Icons	On-Screen Cursor
Instruction (“SELECT . . .”)	Sequence of icons to select		
“Bullet Trajectory Analysis”			
“Blood Spatter Analysis”			
“Hand Scanner”			
“Add Virtual Post-it Text to Scene”			

2. NASA Task Load Index (TLX) method (Hart, 2006) to determine user appreciation based on a written questionnaire. This method assesses the task load on the following different scales: mental demand, physical demand, temporal demand, performance, effort, and frustration. Each scale has 7 points with low, medium, and high increments. The NASA TLX form is extended with several questions to assess the usability as well as the interaction effectiveness.
3. Open text fields in the questionnaire to acquire qualitative feedback from subjects after completing the experiment. Comments provided by subjects while interacting with the system are collected and analyzed.

5.1. Experimental Setup

The subjects chosen for the comparative study are employees of the Delft University of Technology. The 25 subjects are between 24 and 55 years old. The percentage of female subjects in the experiments is 35%. Of the subjects, 68.42% have no experience related to user interfaces in AR, 21.05% have some experience, and only 5.26% are advanced AR users. In a preliminary session, subjects become acquainted with the AR system: equipment and controls including menu selection for the five types of interaction.

Each subject participates in one experiment. An experiment consists of a series of system interaction tests. Each test is limited to a 30-min session. Within this session each subject is assigned one or two tasks of selecting a specific function from the menu for one of the interaction types shown in Figure 4 and one of the menu alignments shown in Figure 7. Accomplishing one task requires activation of the menu interface, navigation to the specific option in the hierarchy of menu items, and selection of the indicated menu item. The menu interface consists of 21 different items, each having assigned an icon. The icons are organized on several levels, according to their function similarity. For instance, the selection of the Toolbox icon leads to expanding a second level, which contains icons regarding functionalities for Bullet Trajectory Analysis, Blood Spatter Analysis, and Scanner Analysis. Further, the Scanner Analysis icon leads to another level, which contains icons for Hand Scanner and Full Range Scanner. The depth of the menu does not exceed four levels.

Table 1 shows task examples for an experiment session. Navigation through the menu interface requires the use of the Home, Up, and Down icons (in the second column of Table 1). The virtual cursor representation is depicted in the last column of Table 1. Activating the Home icon by the on-screen cursor leads to the expansion of the menu hierarchy of icons on the screen. Similarly, deactivating the Home icon by holding the on-screen cursor for 3 s over the surface of the icon, leads to the collapse of the menu hierarchy of icons. The deactivation of the Home icon is the equivalent of partially hiding the menu. The Home icon disappears after a predetermined (3 s) of no interaction activity. From the completely hidden menu state, the Home icon is displayed again when interaction is perceived by the system (when the hand is detected). The Up and Down icons are used in order to scroll through the options.

Subjects are notified about the current task by a message projected on the screen with information on the shortest path to the target menu item.[3] Menu items and the type of interaction are randomly generated at the beginning of each experiment session.

[3]This resembles real-life situations for which users already know the structure of a menu.

TABLE 2
Distribution of Sessions During the Experiment, Given the Interaction Type

Interaction	*#train sessions*	*#test sessions*	*#tests/subject (Mdn)*	*#tests/subject (M ± SD)*	*#subjects Not Testing Condition*
Two hands	7	106	3	4.2 ± 3.2	1
Two hands and object	5	93	3	3.7 ± 3.9	7
Hand and object	9	47	2	1.9 ± 2.6	7
One hand	9	29	1	1.2 ± 0.8	3
One hand and two objects	12	71	2	2.8 ± 3.5	6
Total	42	346	11	13.8 ± 11.4	

TABLE 3
Distribution of Sessions During the Experiment, Given the Alignment Type of Menu UI Layouts

Type of Alignment	*#train sessions*	*#test sessions*	*#tests/subject (Mdn)*	*#tests/subject (M ± SD)*	*#subjects Not Testing Condition*
Screen (fixed)	8	39	1	1.6 ± 1.5	6
Nonfixed	34	307	9	12.3 ± 10.6	0
• 2D	19	176	5	7.0 ± 5.4	0
• 3D	15	131	4	5.2 ± 5.5	2

TABLE 4
Distribution of NASA TLX Questionnaire Inputs, Given the Interaction Type

	Two Hands	*Two Hands and Object*	*Hand and Object*	*One Hand*	*One Hand and Two Objects*
Q1: Mental demand	19	16	15	17	16
Q2: Physical demand	19	16	15	17	16
Q3: Temporal demand	16	13	12	14	13
Q4: Performance	19	15	14	16	15
Q5: Effort	19	16	15	17	16
Q6: Frustration	19	16	15	17	16

Table 2 illustrates the distribution of sessions during the experiment, with respect to the type of interaction. Regarding the investigation of Hypothesis 1, there are 16 training sessions and 135 testing sessions operating the free-hands condition. In addition, there are 26 training sessions and 211 testing sessions operating the condition related to the interaction by physical objects. Regarding the investigation of Hypothesis 2, there are 30 training sessions and 147 testing sessions operating the one-hand condition. Furthermore, there are 12 training sessions and 199 testing sessions operating the two-hand condition. Table 3 illustrates the distribution of sessions during the experiment, with respect to the alignment type of menu UI layouts. Regarding the validation of Hypothesis 3, there are eight training sessions and 39 testing sessions operating fixed menu condition. Furthermore, there were 34 training sessions and 307 testing sessions operating the nonfixed menu condition.

After the experiment, 20 subjects filled in the NASA TLX questionnaire.[4] The feedback collected is compared to the actual data from the system logs. Only the data related to the conditions assigned during the training and testing steps are considered, subject input for other experiment conditions is automatically discarded. Tables 4 and 5 show the number of questionnaire inputs per condition, for each of the six standard questions of the NASA TLX test.

The technical setup for the experiments consists of a laptop and Vuzix Wrap 920AR HMD (Figure 8). The graphical user interface is based on Ogre3D open-source graphics rendering engine library (http://www.ogre3d.org/) and developed using C++ programming language. To implement the

[4]Due to time restrictions, the other subjects could not fill in the questionnaire.

TABLE 5
Distribution of NASA TLX Questionnaire Inputs, Given the Alignment Type of Menu User Interface Layouts

	Screen (Fixed)	*2D*	*3D*
Q1: Mental demand	13	19	17
Q2: Physical demand	13	19	17
Q3: Temporal demand	11	17	15
Q4: Performance	13	19	17
Q5: Effort	13	19	17
Q6: Frustration	13	19	17

FIG. 8. Technical setup: Subject at the experiment during the test session.

models for detection, recognition and tracking the hand and physical objects, the following systems are used: C++ programming language, Boost::Thread library (Boost C++ Libraries, 2014) for parallel computing and the open computer vision library OpenCV (http://opencv.org/). A robust algorithm for real-time pose estimation from a planar target (Schweighofer & Pinz, 2006) is used for 3D world menu alignment.

A single software application handles the execution of the experiments. During each experiment, subject interaction events are logged automatically. Such logged events relate to hand gestures, cursor movements, menu activation and deactivation, menu navigation, and menu item selection.

6. RESULTS

As just described, the interaction types are evaluated based on the information collected from the three different instruments. First, the system logs collected during the experiments are analyzed to extract objective indicators on the performance of interaction. Then the questionnaires filled in by the subjects and the additional comments are analyzed. Along with the questionnaires, the participant feedback is correlated with the results from oral examination.

6.1. Quantitative Interaction Analysis

Using the logged data on the interaction, this section evaluates the different interaction types with regard to the time needed to complete a task and the path of the cursor in relation to the menu. The upper part of Figure 9 shows an example of a subject's activity while using two hands for interaction (Figure 4a) and a 2D menu alignment (Figure 7b).

Figure 9 shows only the user interaction activity while the menu is visible on the screen. According to the graph on the left side of Figure 9, the menu disappeared from the AR view for a few seconds when the hands or object are not visible anymore. Red circles in the graph indicate selections of menu items. Based on the data collected during the experiments, the learning rate of subjects with respect to the time required for completing an assigned task is computed. The right side of Figure 9 shows an example of the learning curve of a subject completing a sequence of tasks during the test session. The vertical axis shows the menu selection times in seconds, and the horizontal axis indicates time-based selection tasks. The decrease of the task completion time indicates that the subject is able to learn to adjust to and to control the interface during the test session.

Figure 10 (for learning) and Figure 11 (for testing) show graphs with the average times per task for all subjects. In the figures, the Parameter Interaction Type 1 relates to two-hands interaction (Figure 4a), 2 to two-hand and object interaction (Figure 4b), 3 to hand and object interaction (Figure 4c), 4 to one-hand interaction (Figure 4d), and 5 to one-hand two-objects interaction (Figure 4e). The parameter type of Alignment 1 denotes the screen alignment (Figure 7a), 2 the 2D alignment (Figure 7b), and 3 the 3D World alignment (Figure 7c).

To remove the bias of the number of depth levels to the menu items to be selected, the right part of Figure 10 and Figure 11 show the full range of data, outliers, and standard deviation for the selection times corrected with the depth level of each menu item.

The figures show that the average times for completing a task during the experiments are lower than the times within the learning sessions held before the experiments. Second, the graphs in Figure 10 and Figure 11 indicate the flat surface with physical object interaction (Figure 4e) as being the best in terms of average time necessary to fulfill the tasks, for both learning ($M =$ 24.81 s) and testing ($M = 18.09$ s) sessions. The figures suggest that outliers in the test session vary more than outliers in the training session.

As suggested in Figure 11, during testing, two-hand interaction (Figure 4a) is slower than two-hand and object interaction (Figure 4b) with 2.9 s on average, and with 11.9 s on average than hand and two objects condition (Figure 4e). The interaction type relying on one-hand interaction (Figure 4d) has the longest average times for completion, followed by hand and object interaction, for both learning ($M = 90.8$ s vs. $M = 83.8$ s) and testing ($M = 63$ s vs. $M = 50.7$ s). This result correlates with the levels of frustration and the performance rankings from

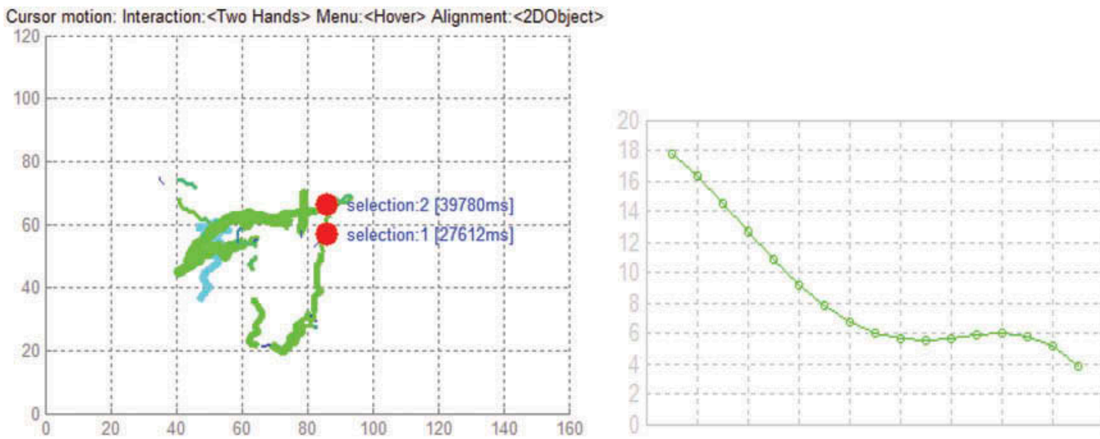


FIG. 9. Left: example of tracking menu movement (cyan), cursor movement (green), and menu item selections (red) during a task. *Note.* Thicker green lines indicate that the subject keeps the cursor in the area for longer time. Right: sample graph showing the improvement of the task completion time per task (in seconds), for a subject during the test session.

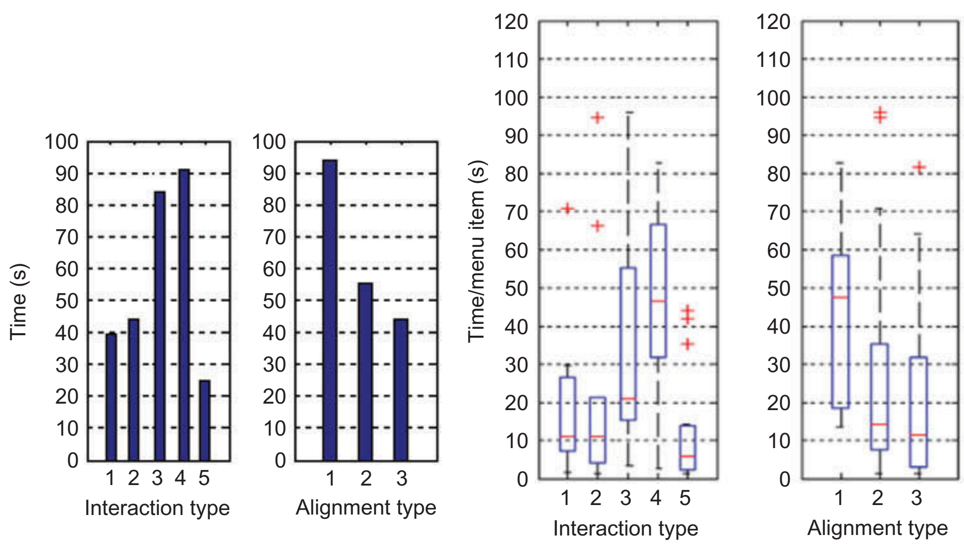


FIG. 10. Learning time per task (in seconds): average total time (left) and time per depth level of menu item (right).

the questionnaire. The free-hand interaction is not necessarily the fastest among the five types of interaction in AR. The results indicate that the completion time for one-hand interaction (Figure 4d) is more than 10 s longer than the second worst condition, that is, hand and object interaction (Figure 4c).

A Mann–Whitney test indicated that the task completion time was greater for experiment sessions on free-hand interaction (Mdn = 16.07 s) than for experiment sessions on tangible interaction (Mdn = 15.12 s, p = .06). The results at this stage do thus not support Hypothesis 1 that assumes free-hand interaction is more effective than tangible interaction.

A Mann–Whitney test indicated that the task completion time was significantly greater for experiment sessions on one-hand interaction (Mdn = 18.56 s) than for experiment sessions on two-hand interaction (Mdn = 12.71 s, p = .024).

Figure 12 indicates that the interaction by two hands and one object (Figure 4b) has on average the smallest number of errors ($M = 1.42$) during testing session. Two-hand interaction (Figure 4a) proves the next best in terms of menu item selections ($M = 1.45$).

A Mann–Whitney test indicated that the number of errors was significantly greater for experiment sessions on one-hand interaction ($Mdn = 2.10$) than for experiment sessions on two-hand interaction ($Mdn = 1.44$, $p = .002$). These aspects do not support Hypothesis 2 that assumes one-hand interaction is more effective.

Screen alignment (Figure 7a) has the most negative impact on the average task completion time for test ($M = 55.6$ s, $p < .0001$) and learning ($M = 93.9$ s, $p = .0182$) sessions, whereas 2D and 3D menu alignments give similar completion times during testing.

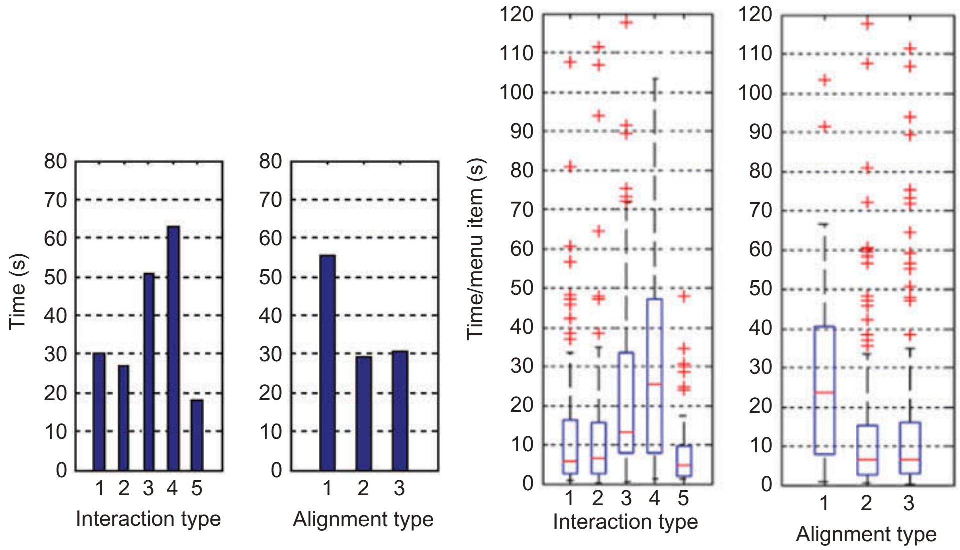


FIG. 11. Left: average total time per task during testing (in seconds). Right: task completion time per depth level (in seconds).

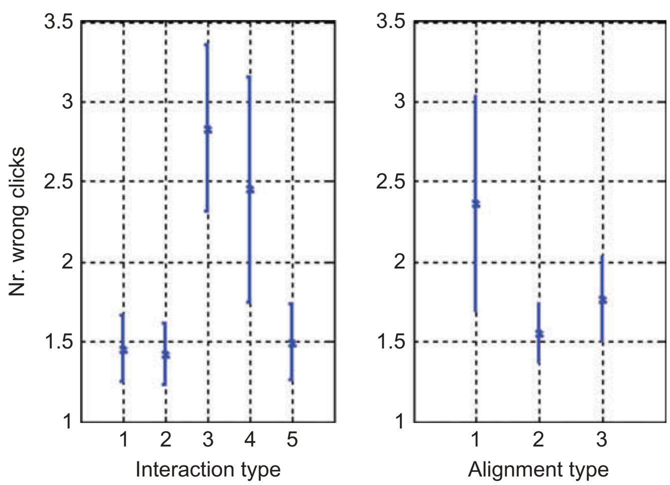


FIG. 12. Number of errors per task during testing.

A Mann–Whitney test indicated that the task completion time was significantly greater for experiment sessions using fixed menu UI layouts (*Mdn* $= 50.59$ s) than for experiment sessions on nonfixed UI layouts (*Mdn* $= 13.46$ s, $p < .0001$). Similarly, a Mann–Whitney test indicated that the number of errors was greater for experiment sessions using fixed menu UI layouts (*Mdn* $= 2$) than for experiment sessions on nonfixed UI layouts (*Mdn* $= 1$, $p = .051$).

The indicators of average task completion time and number of errors per task, lead to the rejection of Hypothesis 3 that assumes fixed menu UI layout (screen menu alignment) allows for better interaction than nonfixed alignment (2D and 3D menu alignment).

6.2. NASA Task Load Index Questionnaires

The NASA task load index (TLX) scales for the TLX questionnaires originally contain 7 points with low, medium, and high increments clustered in groups. For analysis, each scale has three groups, and each group has seven increments.[5]

According to the evaluation of the NASA TLX forms, most of the subjects give low and medium ranks on the question of how hurried or rushed was the pace of the tasks they had to accomplish (Q3), for all five interaction types. Only 7.69% of the subjects testing one-hand and two-objects condition (Figure 4e) gave high ranks to this question.

The NASA TLX questionnaire reveals that two-hand interaction (Figure 4a) has the worst scores compared to the other interaction types with regard to mental demand (Q1; 73.68%) and performance (Q4; 63.16%, vs. rest: $p = .1669$) for low and medium scores. The two hands with object interaction type (Figure 4b) shows the best scores at effort (Q5; 75%) and frustration (Q6; 87.5%, vs. rest: $p = .095$). One hand and object (Figure 4c) shows the best result at mental demand (Q1; 93.33%). One-hand interaction (Figure 4d) shows the worst result at physical demand (Q2; 47.06%, vs. rest: $p = .084$). Interaction by one hand and two objects (Figure 4e) shows the best results at physical demand (Q2; 81.25%, vs. rest: $p = .1314$) and frustration (Q6; 87.5%). 2D menus show the best results at performance (Q4; 78.95%) and effort (Q5; 84.21%)

[5]This distinction makes it possible to derive statistics on the number of inputs that fall in each group, pointing to indicators of low, medium, and high rankings or scores. Because some data sets (about 15%) for the different interaction types show non-normal distributions, (according to Anderson–Darling test), Mann–Whitney U test is used for checking populations prior to reporting findings.

and the worst results at physical demand (Q2; 73.68%) and temporal demand (Q3; 76.47%). 3D menus show the best result at mental demand (Q1; 82.35%) and the worst results at performance (Q4; 64.71%) and effort (Q5; 70.59%, vs. rest: $p =$ .0697).

For each of the data sets, Anderson–Darling test is used to determine if a data set is from a normal distribution population. For testing Hypothesis 1, only the data sets for Q1, Q3, and Q5 on the free-hand interaction types and the data sets for Q4 and Q5 on tangible interaction types show populations with normal distributions. For testing Hypothesis 2, only the data set for Q4 on one-hand interaction type and the data sets Q2 and Q3 on the two-hand interaction types show populations with normal distribution. For testing Hypothesis 3, no data set shows a population with normal distribution.

Because several data sets are from non-normal distributions, Mann–Whitney *U* test is used to subsequently test if there is any compelling evidence that data sets of distinct conditions differ. Table 6 depicts the results of using Mann–Whitney *U* test on different data sets related to the three hypotheses under investigation. In the table, fixed menu refers to the screen condition, and nonfixed menu refers to the 2D and 3D conditions. From the table, there is only one case for which data sets are from distinct populations (for Q6 on one-hand vs. two-hand conditions).

Given the percentage of low and medium scores given by the subjects, the questionnaire analysis indicates that tangible interaction is better than free-hand interaction with regard to mental demand (Q1; 87.23% vs. 77.78%), physical demand (Q2; 74.47% vs. 55.56%), effort (Q5; 63.83% vs. 55.56%), frustration (Q6; 80.85% vs. 66.67%). For performance (Q4), tangible condition (72.73%) is lower than free-hand condition (80%). For the temporal demand (Q3), the tangible condition (97.37%) is comparable to free-hand condition (100%) at the previously mentioned scores. These findings lead to the rejection of Hypothesis 1.

In case of testing Hypothesis 2, two-hand interaction scores higher than one-hand interaction, given effort (Q5; 74.29% vs. 50%) and temporal demand (100% vs. 97.44%), at the previously mentioned scores. A Mann–Whitney test indicated that the frustration (Q6) was significantly higher for two-hand interaction than during experiment sessions on one-hand interaction (82.86% vs. 68.75%, $p = .018$).

For testing Hypothesis 3, there is evidence that nonfixed menu condition has better results than fixed menu condition with regard to the mental demand (83.33% vs. 76.92%), performance (72.22% vs. 69.23%), and effort (77.78% vs. 76.92%) at the previously mentioned scores. Fixed menus score better than nonfixed menus at physical demand (92.31% vs. 75%), temporal demand (100% vs. 84.38%), and frustration (88.89% vs. 92.31%).

In addition to the NASA TLX questions, the questionnaire requested separate feedback on various interaction aspects. From the answers to additional questions targeting user perception, the results show that 69.23% subjects consider the tangible interaction to be more natural and 75% consider physical objects to be comfortable for interaction in AR. Of the subjects, 91.67% indicate that physical objects help to finish the assigned tasks in AR in shorter time and 66.67% mention that tangible interaction supports them in shifting the focus from artificial content to the physical world. Furthermore, 63.64% of the subjects mention that their tangible interaction is more natural than using hands only, because it involves tactile perception that is typical for physical world bringing the digital world closer to the physical world. Of the subjects, 81.82% indicate that the interaction using physical objects is more natural because it leads to better motoric coordination, whereas 54.55% indicate that the interaction is more natural due to immersion in AR.

All subjects specify that accurate detection and tracking of hands and physical objects, better hardware optics, and small delay of rendering AR content are essential prerequisites for a natural interaction with physical objects. Moreover, 88.24% of the subjects indicate that the system should assign useful functionality to physical objects, and 62.5% indicate that such systems should even consider measuring the affective state of the subject during the interaction. Of the subjects, 94.44% mention that physical objects should have intuitive functionality for subject's task. The system capabilities triggered by interacting with physical objects should be dependent on the physical context (94.12%). Of the subjects, 66.67% suggest that physical

TABLE 6
Results (*p* Value) of Mann–Whitney *U* Test on Different Conditions

	Free Hands vs. Tangible	*One Hand vs. Two Hands*	*Fixed vs. Nonfixed Menus*
Q1: Mental demand	0.3834	0.6856	0.5837
Q2: Physical demand	0.0800	0.5717	0.3134
Q3: Temporal demand	0.7546	0.4589	0.4946
Q4: Performance	0.2869	0.1229	0.4143
Q5: Effort	0.4146	0.1546	0.6475
Q6: Frustration	0.3683	**0.0180**	0.5974

Note. Bold indicates compared data sets are from distinct populations.

objects for AR should be chosen from the ones that already exist in the physical environment.

6.3. Qualitative Feedback

Along with the collected questionnaire data, the results from the previous data sources are correlated with the oral feedback from the participants. This type of data is especially important in order to increase the confidence on findings from the questionnaire analysis, compensating for the rather low number of participants during the experiment. The analysis of the individual comments shows that subjects access the menu items more effectively when they can control the positioning of both cursor and menu in the AR view (Figure 4a, b, e).

Both the 2D and the 3D alignments are perceived as being more effective than the fixed screen alignment, rejecting Hypothesis 3. The 3D alignment is perceived as just a little bit better than the 2D alignment. Similar findings are reported in Billinghurst and Thomas (2011). A few subjects indicate that they could learn how to control the positioning of the menu or cursor by moving their head instead, holding the hands or physical object still. Also, subjects indicate that it is possible to interact by hands and physical objects even after short learning sessions, despite having little or no previous experience with AR.

Several subjects indicate that technological limitations affect the interaction in AR due to uncomfortable hand postures, the arm fully stretched provoking tiredness even after a short time. Some subjects report that the interaction is often affected by the lack of synchronization between their hand gestures and the gestures perceived by the system. This is attributed to the occasional failure of hand tracking, hand posture detection, and gesture recognition models.

6.4. Discussion

The study reveals that physical objects can be transformed into valuable means for interaction in AR. Although the study was carried on in the context of the CSI domain, the results are also applicable to other domains, as the experiments did not require CSI background knowledge.

The results are derived considering system logs and questionnaire data. To compensate for the rather low number of participants at the experiments, the questionnaire data are correlated to the results from the oral examination.

The findings indicate that the interaction by hand and two objects (Figure 4e) is the best option to navigate and manipulate the menu system in AR. The second best interaction type also uses a physical object (two hands and object, Figure 4b). Following the evaluation, interaction with a physical object is the preferred type of interaction for naturally pointing and navigating the menu interface. The results suggest that the natural feedback provided by tangible interaction closes the gap between the virtual and physical worlds, providing synchronization between the real world and the augmented view. Although slightly shorter interaction times are obtained for the 2D menu alignment, the 3D world alignment is preferred due to the smaller error for selecting menu items.

Quantitative analysis indicates that free-hand interaction is not more effective than tangible interaction (Hypothesis 1), that one-hand interaction is not more effective than two hands interaction (Hypothesis 2), and that fixed menu UI layout does not allow for better interaction (Hypothesis 3). More, the questionnaire analysis clearly indicates the rejection of Hypothesis 1 and the analysis of the qualitative feedback indicates the rejection of Hypothesis 3.

Hard limitations mainly associated with the technology lead to less natural interaction in AR. This bias is due to a small working volume that implies rather uncomfortable hand postures, arm almost fully stretched, giving tiredness after a relatively short time. More, occasional failure of tracking algorithms, hand posture detection, and gesture recognition biases the synchronization between the user's intention/real action and that perceived by the system. Higher rates of false positives and the desynchronization may have the consequence that the user gets frustration and that the AR system does not provide immersive interaction experience. The user should be able to naturally sense, interact with elements, and get feedback from both worlds. There should be temporal and spatial synchronization between the physical world and virtual representation in terms of at least the dynamics of the user's actions and physical objects with which the user is interacting. The user interaction is regulated by means of AR technology designed to support users' tasks within the physical environment. The physical objects should have their utility proven in the given work context.

7. CONCLUSIONS AND FUTURE WORK

Interaction in AR should consider users' natural body gestures and should follow the similar approach of interaction with everyday objects. In certain work contexts like CSI, physical objects can be effectively turned into handles of interaction for augmented interfaces. Being stimulated to use physical objects, possibly from the ones traditionally used to accomplish various tasks for work, the user benefits from the natural tactile feedback. The natural feedback enhances the AR experience of the user wearing the HMD by synchronizing the augmented content with the sensed physical world.

This article is the first to report results on interaction in AR based on computer vision methods for hand tracking without visible markers. Three hypotheses targeting different characteristics of hand interaction with and without physical objects in AR are rejected based on a study comparing the usability and effectiveness of five different interaction types. The outcome of this investigation shows that free-hand interaction is not more effective than tangible interaction, one-hand interaction is not more effective, and fixed menus UI layout does not lead to better interaction in AR. This hand-tracking approach, illustrated for CSI in this article, has the potential to be deployed in AR applications in domains such as medicine, space, construction and assembly, scientific simulations, and training.

The study results further show that tangible interaction allows users to perform better when they navigate and manipulate the menu system in AR. The results of the study with 25 subjects indicate that users prefer the ability to determine the position of the user interface and to physically point to a preferred option for navigation by using physical objects. Future research will investigate to which extent everyday objects can be used for interaction in AR, how can such objects be automatically detected, and how can the functionality be assigned to similar clusters of objects.

REFERENCES

Bertelsen, O. W., & Nielsen, C. (2000). Augmented reality as a design tool for mobile interfaces. *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, 185–192.

Billinghurst, M., & Thomas, B. H. (2011). Mobile collaborative augmented reality. In L. Alem & W. Huang (Eds.), *Recent trends of mobile collaborative augmented reality systems* (pp. 1–19). New York, NY: Springer.

Bowman, D., Kruijff, E., LaViola, J. J., Jr., & Poupyrev, I. (2004). *3D user interfaces: Theory and practice*. Reading, MA: Addison-Wesley.

Broll, W., Lindt, I., Ohlenburg, J., Wittkämper, M., Yuan, C., Novotny, T., . . . Strothmann, A. (2004). ARTHUR: A collaborative augmented environment for architectural design and urban planning. *Journal of Virtual Reality and Broadcasting, 1*(1). Retrieved from http://www.jvrb.org/past-issues/1.2004/34

Chen, Y.-C., Chi, H.-L., Hung, W.-H., & Kang, S.-C. (2011). Use of tangible and augmented reality, models in engineering graphics courses. *Journal of Professional Issues in Engineering Education & Practice*, *137*, 267–276.

Cheng, K.-Y., Liang, R.-H., Chen, B.-Y., Liang, R.-H., & Kuo, S.-Y. (2010). iCon: Utilizing everyday objects as additional, auxiliary and instant tabletop controllers. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI'10*, 1155–1164.

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*, 681.

CSI The Hague (2012). Digitise the crime scene: Fiction becomes reality. Retrieved from http://www.csithehague.com/

Dachselt, R., & Hübner, A. (2007). Three-dimensional menus: A survey and taxonomy. *Computers & Graphics*, *31*, 53–65.

Datcu, D., & Lukosch, S. (2013, July). *Free hands interaction in augmented reality*. Paper presented at the ACM Symposium on Spatial User Interaction (SUI), Los Angeles, CA.

Datcu, D., Swart, T., Lukosch, S., & Rusak, Z. (2012). Multimodal Collaboration for Crime Scene Investigation in Mediated Reality. *14th ACM International Conference on Multimodal Interaction – ICMI*, 299–300.

Ehnes, J. (2009). A tangible interface for the AMI content linking device—The Automated Meeting Assistant. *IEEE HSI* 2009, 306–313.

Gu, N., Kim, M. J., & Maher, M. L. (2011). Technological advancements in synchronous collaboration: The effect of 3D virtual worlds and tangible user interfaces on architectural design. *Automation in Construction*, *20*, 270–278.

Hart, S. (2006). Nasa-Task Load Index (NASA-TLX); 20 years later. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *50*, 904–908.

Henderson, S. J., & Feiner, S. (2008). Opportunistic controls: Leveraging natural affordances as tangible user interfaces for augmented reality. *Proceedings of ACM Symposium on Virtual Reality Software Technology*, 211–218.

Jones, B. R., Sodhi, R., Campbell, R., H., Garnett, G., & Bailey, B., P. (2010). Build your world and play in it: Interacting with surface particles on complex objects. *International Symposium on Mixed and Augmented Reality – ISMAR'10*, 165–174.

Kanade, T., & Hebert, M. (2012). First-person vision. *Proceedings of the IEEE*, *100*, 2442–2453.

Kinect. (n.d.). In *Wikipedia*. Retrieved August 28, 2014, from http://en.wikipedia.org/wiki/Kinect

Kölsch, M., Turk, M., & Höllerer, T. (2004). Vision-Based Interfaces for Mobility. *International Conference on Mobile and Ubiquitous Systems (MobiQuitous)*, August 22–26, Boston, Massachusetts, USA.

Lee, J. Y., Rhee, G., W., & Seo, D., W. (2010). Hand gesture-based tangible interactions for manipulating virtual objects in a mixed reality environment. *International Journal of Advanced Manufacturing Technology, 51*, 1069–1082.

Lee, M., Green, R., & Billinghurst, M. (2008). 3D natural hand interaction for AR applications. *23rd International Conference Image and Vision Computing*, 26–28.

Livingston, M. A., Ai, Z., Karsch, K., & Gibson, G. O. (2011). User interface design for military AR applications. *Virtual Reality*, *15*, 175–184.

López, G., López, M., Guerrero, L. A., & Bravo, J. (2014). Human–objects interaction: A framework for designing, developing and evaluating augmented objects. *International Journal of Human–Computer Interaction*, *30*, 787–801. doi:10.1080/10447318.2014.927281

Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *Proceedings of Imaging Understanding Workshop*, 121–130.

Maier, P., Tonnis, M., Klinker, G., Raith, A., Drees, M., & Kuhn, F. (2010). What do you do when two hands are not enough? Interactive selection of bonds between pairs of tangible molecules, *IEEE Symposium on 3D User Interfaces - 3DUI'10*, 83–90.

Mair, E., Hager, G. D., Burschka, D., Suppa, M., & Hirzinger, G. (2010). Adaptive and generic corner detection based on the accelerated segment test. *Computer Vision – ECCV'10, Lecture Notes in Computer Science, 6312*, 183–196.

Marco, J., Cerezo, E., & Baldassarri, S. (2012). Bringing tabletop technology to all: evaluating a tangible farm game with kindergarten and special needs children. *Personal and Ubiquitous Computing*, *17*, 1577–1591.

Mateu, J., Lasala, M. J., & Alamán, X. (2014). VirtualTouch: A tool for developing mixed reality educational applications and an example of use for inclusive education. *International Journal of Human–Computer Interaction*, *30*, 815–828. doi:10.1080/10447318.2014.927278

Merrill, D., & Maes, P. (2007). Augmenting Looking, Pointing and Reaching Gestures to Enhance the Searching and Browsing of Physical Objects, *In: Proceedings of the 5th international conference on Pervasive computing, PERVASIVE'07*, ISBN: 978-3-540-72036-2, Springer-Verlag Berlin, Heidelberg, 1–18.

bin Mohd Sidik, M. K., bin Sunar, M. S., bin Ismail, I., bin Mokhtar, M. K., & Jusoh, N. B. M. (2011). A study on natural interaction for human body motion using depth image data. *Workshop on Digital Media and Digital Content Management - DMDCM*, 97–102.

Nagel, T., & Heidmann, F. (2011). Exploring faceted geo-spatial data with tangible interaction. *GeoViz 2011*, March 10–11, 2011, Hamburg, Germany.

Ness, S., R., Reimer, P., Krell, N., Odowichuck, G., Schloss, W., A., & Tzanetakis, G. (2010). Sonophenology: A tangible interface for sonification of geo-spatial phonological data at multiple time scales. *The 16th International Conference on Auditory Display (ICAD-2010)*, 335–341.

Newton-Dunn, H., Nakano, H., & Gibson, J. (2003). Block Jam: A tangible interface for interactive music. *Proceedings of the 2003 conference on New Interfaces for Musical Expression NIME'03*, 170–177.

Piumsomboon, T., Clark, A., & Billinghurst, M. (2011). Physically-based interaction for tabletop augmented reality using a depth-sensing camera for environment mapping. *Proceeedings of Image and Vision Computing New Zealand (IVCNZ-2011)*, 161–166.

Poelman, R., Akman, O., Lukosch, S., & Jonker, P. (2012). As if being there: mediated reality for crime scene investigation. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, CSCW'12*, 1267–1276.

Radkowski, R., & Stritzke, C. (2012). Interactive hand gesture-based assembly for augmented reality applications. *The Fifth International Conference on Advances in Computer-Human Interactions ACHI'12*, 303–308.

Reifinger, S., Wallhoff, F., Ablassmeier, M., Poitschke, T., & Rigoll, G. (2007). Static and dynamic hand-gesture recognition for augmented reality applications. *Proceedings of the 12th International Conference on Human-Computer Interaction: Intelligent Multimodal Interaction Environments*, 728–737.

Schiettecatte, B., & Vanderdonckt, J. (2008). AudioCubes: A distributed cube tangible interface based on interaction range for sound design. *Proceedings of the Second International Conference on Tangible and Embedded Interaction (TEI'08)*, 3–10.

Schweighofer, G., & Pinz, A. (2006). Robust pose estimation from a planar target. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*, 2024–2030.

Shen, Y., Ong, S. K., & Nee, A. Y. C. (2011). Vision-based hand interaction in augmented reality environment. *International Journal of Human–Computer Interaction*, *27*, 523–544. doi:10.1080/10447318.2011.555297

Smith, C. G. (1995), May/June. The hand that rocks the cradle. *I.D.*, pp. 60–65.

Swart, T. (2012). *Design of a gesture controlled graphic interface for Head Mounted Displays for CSI The Hague* (Unpublished master's thesis). Delft University of Technology, Delft, the Netherlands.

Viola, P., & Jones, P. (2002). Robust real-time object detection. *International Journal of Computer Vision*, *57*, 137–154.

White, S., Feng, D., & Feiner, S. (2009). Interaction and presentation techniques for shake menus in tangible augmented reality. *8th IEEE International Symposium on Mixed and Augmented Reality, ISMAR'09*, 39–48.

Wiedenmaier, S., Oehme, O., Schmidt, L., & Luczak, H. (2003). Augmented Reality (AR) for assembly processes design and experimental evaluation. *International Journal of Human-Computer Interaction*, *16*, 497–514. doi:10.1207/S15327590IJHC1603_7

Williams, A., & Botet Escriba, V. J. (n.d.). Boost C++ Libraries. *Chapter 28. Thread 3.1.0*. Retrieved from http://www.boost.org/doc/libs/1_52_0/doc/html/thread.html

Xu, D., Read, J., C., Mazzone, E., & Brown, M. (2007). Designing and testing a tangible interface prototype. *IDC 2007 Proceedings: Methodology*, 25–28.

ABOUT THE AUTHORS

Dragoş Datcu is Postdoc researcher at the Delft University of Technology. His current research is on complex problem solving in new interaction spaces such as augmented reality. This includes the design of novel augmented reality-driven solutions enabling collaboration among distributed users, also by using affective computing technology.

Stephan Lukosch is associate professor at the Delft University of Technology. His current research focuses on presence, interaction, and coordination issues around virtual colocation using augmented reality. His research builds upon his recent research results on context-adaptive collaboration, collaborative storytelling, and design patterns for computer-mediated interaction.

Frances Brazier is a full professor at the Delft University of Technology. Her research focuses on the design of complex, emergent social technical participatory systems in merging networked realities. The design of witnessed presence, agency and awareness, self-management, and distributed coordination are core challenges within her research and the Participatory Systems Initiative.