Calibration of an HMPD-Based Augmented Reality System

Hong Hua, *Member, IEEE*, Chunyu Gao, and Narendra Ahuja, *Fellow, IEEE*

***Abstract*—In augmented reality (AR) applications, accurately registering a virtual object with its real counterpart is a challenging problem. The size, depth, and geometry, as well as the physical attributes of the virtual object, have to be rendered precisely relative to a physical reference. This paper presents a systematic calibration process to address the registration challenge in a custom-designed AR system, which is based upon recent head-mounted projective display (HMPD) technology. Following a concise review of the HMPD concept and our system configuration, we first present a computational model of the HMPD viewing system and requirements for the system calibration. Then, we describe, in detail, the calibration procedures to obtain estimates of unknown transformations, summarize the application of the estimates in a customized graphics rendering toolkit, and discuss the evaluation experiments and observations. Finally, the implementation of a testbed to demonstrate a successful registration is briefly described, and experimental results are presented.**

***Index Terms*—Augmented reality (AR), camera calibration, display calibration, head-mounted display (HMD), head-mounted projective display (HMPD).**

I. Introduction

AUGMENTED reality (AR) is a paradigm in which a user's sensory perception of the physical world (e.g., visual, auditory, tactile, smell, or taste) is enhanced, rather than replaced, with computer-generated supplemental information. Visual augmentation is a well-known example, and several research groups have been exploring potential applications such as computer-aided surgery [3], [29], [42], medical training [1], repair and maintenance of complex facilities [5], [10], and telemanipulation [32].

A successful AR application demands accurate registration of a virtual object with its physical counterpart to create the illusion of coexistence, which requires that the size, depth, and geometry, as well as the physical attributes of virtual objects, be rendered accurately with respect to their physical counterparts. The registration accuracy depends mainly on the following factors [2], [14]: 1) the modeling of the virtual cameras, through which 2-D images of the 3-D virtual world are generated; 2) the modeling of the optical viewing system such as a head-mounted display (HMD), through which both the physical and virtual worlds are viewed; 3) the resolution and accuracy of the motion tracking systems, through which users' heads and objects of interest are tracked to update their corresponding transformations; 4) the end-to-end system latency, which refers to the time delay from the moment an interactive event takes place to the moment the corresponding properties of the synthetic world are updated; and 5) the rendering quality, which refers to the modeling of the geometric and physical attributes (i.e., shape, resolution, reflectance, etc.) of virtual objects with respect to their physical counterparts.

The first three factors are identified as the sources of static registration errors because they result in misregistration even if a user is still. The end-to-end latency is typically identified as the source of dynamic registration error because it only plays a role in a dynamic environment in which either the user moves or the physical objects of interest are manipulated. The last factor is identified as the modeling error related to rendering techniques.

Accurately modeling the viewing process and strategically calibrating the viewing system are essentially critical in achieving a precise registration of a virtual world with its physical counterpart. Several efforts have been made to study computational models and parameters of HMD-based 3-D display systems, which are needed to accurately generate 2-D images from 3-D virtual environments. Robinett and Rolland provided a detailed analysis of the computational process in a traditional HMD, taking into account the optics, tracking, and geometry of the HMD [37]. This model was further upgraded by Robinett and Holloway to mathematically expose the sequence of transformations in a virtual-reality system from object to screen coordinates [38]. Deering presented the general steps that must be taken to produce an accurate high-resolution head-tracked stereo display in order to achieve a subcentimeter virtual-to-physical registration [8].

A variety of calibration methods and procedures have been investigated to calibrate AR systems based on various display technologies. Video and optical see-through HMDs are two typical approaches for combining real and virtual images [36]. In both approaches, the viewing optics is typically an eyepiece-type compound magnifier. A comprehensive review of HMD technologies for AR systems can be found in the study in [37]. Due to the fact that the real world is simply imaged through digital cameras and can be analyzed directly, the calibration of video-based AR systems is relatively straightforward, and it can be done accurately using traditional camera calibration procedures [44]. For example, Tuceryan *et al.* described the

Manuscript received February 15, 2005; revised January 18, 2006. This work was supported by the National Science Foundation under Grants 0083037, 0417598, and 0411578. This paper was recommended by Associate Editor Y. Xiao.

H. Hua is with the College of Optical Sciences, The University of Arizona, Tucson, AZ 85721 USA (e-mail: hhua@optics.arizona.edu).

C. Gao and N. Ahuja are with the Beckman Institute for Advanced Science and Technology and the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: cgao@uiuc.edu; ahuja@vision.ai.uiuc.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCA.2007.893471

calibration issues arising in a video-based AR system [45]. Bajura and Neumann presented the idea of dynamically measuring registration errors in synthetic images to correct 3-D registration errors in a video-based system [4]. Grimson *et al.* explored vision techniques used to automate the process of registering medical data to a patient's head in a video-based system [13]. The calibration of AR systems based on optical see-through HMDs has demonstrated much more challenges, and researchers have been continually investigating strategic procedures to achieve more accurate, reliable, and automated calibration. For instance, Janin *et al.* described the procedures used to determine the calibration parameters of a see-through HMD via both manual measurement and optimization [25]. Azuma and Bishop described experimental steps used to estimate viewing parameters and presented predictive tracking techniques used to improve both static and dynamic registrations in an optical see-through HMD system [2]. Oishi and Tachi proposed a calibration method used to minimize systematic errors in projection transformation parameters for optical see-through HMDs [33]. Tuceryan and Navab described a single point active alignment method for optical see-through HMDs [46]. McGarrity and Tuceryan described a camera-based online calibration method for optical see-through display [30], [31].

Head-mounted projective display (HMPD), pioneered by Fisher [11] and Kijima and Ojika [27], is an emerging optical see-through display technology which lies on the boundary between conventional HMDs and CAVE Automatic Virtual Environment (CAVE)-like projection displays [6]. The HMPD concept and the technical advancements will be reviewed in Section II. The focus of this paper is to present a computational mechanism that accurately models the HMPD viewing system with an equivalent viewing device and to present a manual correspondence matching (MCM) calibration method that estimates both intrinsic and extrinsic parameters of the display system and thus establishes the viewing and projection transformations for the equivalent viewing devices, which will then be utilized to configure the virtual cameras for image rendering. This paper will also describe a set of experiments used to evaluate the calibration results and include a testbed example used to visually assess the registration quality.

The rest of this paper is organized as follows. In Section II, we briefly review the basic concept and recent developments in HMPD technology and describe our system configuration. Section III presents a computational model and the requirements for the system calibration, which gives the transformations needed to link the physical viewing system with the virtual cameras. Section IV describes the calibration procedures used to estimate the transformations and includes calibration results. In Section V, we present evaluation methods and experiments and discuss our observations. Finally, in Section VI, we demonstrate the results of a successful calibration as part of a testbed that involves an augmented "GO" game.

II. System Configuration

A. Overview of the HMPD Technology

A monocular configuration of an HMPD [11], [27] is conceptually illustrated in Fig. 1. Unlike a conventional optical

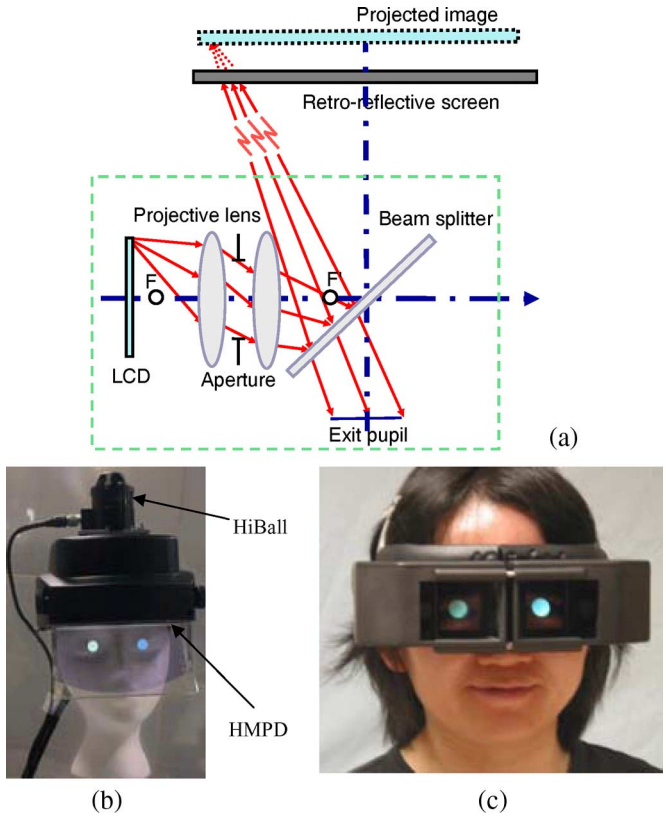

Fig. 1. Advancement of the HMPD. (a) Basic imaging concept of the HMPD. (b) First-generation prototype designed at the University of Central Florida and built at the University of Illinois. (c) New prototype designed and built at the University of Arizona, using the same projection lens designed for the first prototype.

see-through HMD, an HMPD replaces eyepiece-type optics with a projective lens [9], [11], [16], [27]. An image on the miniature display, which is located beyond the focal point of the lens, rather than between the lens and the focal point as in a conventional HMD, is projected through the lens and thus forms a magnified real image. A beamsplitter then reflects the real image outward into the object space, where it is reflected toward the observer by a retro-reflective screen. The uniqueness of a retro-reflective screen lies in the fact that a ray hitting the surface at any angle is ideally reflected back on itself in the opposite direction. Therefore, the projected real image is retro-reflected back to the exit pupil of the optics, where the eye is positioned to observe the magnified image. Owing to the essence of retro-reflection, the location and the size of the perceived image are theoretically independent of the location and shape of the retro-reflective screen [16].

The HMPD concept has recently been explored extensively by several researchers [16], [17], [21], [26], [27] and has been recognized as an alternative solution for a wide range of 3-D visualization applications [18], [22], [24], [28], [34], [39]. Pioneered by Rolland, large field-of-view (FOV), lightweight, and low distortion optics for HMPD systems have been designed [21], and a custom-designed ultralight compact prototype was developed [17], [18], [21]. Several attempts have been made thereafter to explore more compact and brighter display designs. For instance, Ha *et al.* have designed a 70° wide FOV projection lens [15], and Rolland has been developing an

organic-light-emitting-display-based 42° prototype, collaborating with NVIS Inc. [40]. Anterior views of the first-generation HMPD prototype and a recent prototype are shown in Fig. 1(b) and (c), respectively. Both prototypes achieved a 52° FOV with a 640 × 480 video graphics array resolution. While the first-generation prototype weighs about 750 g, the newer prototype, with a folding mechanism, weighs about 400 g.

B. Overview of System Implementation

The HMPD concept intrinsically enables a shared workspace with an arbitrary number of individual viewpoints rather than the leader-privileged viewing mode of a traditional CAVE-like projection environment. While such HMPD-based collaborative space can potentially take many forms, for instance, a deployable room coated with retro-reflective material, demonstrated by the study in [7], [39], we have developed a collaborative infrastructure, referred to as stereoscopic collaboration in augmented and projective environments (SCAPE), which combines a retro-reflective workbench, allowing exocentric viewing of an augmented dataset with a room display allowing egocentric viewing of life-size virtual environments [22]. A computer-generated low-detailed microscene is registered with the workbench, and physical objects are placed on it, while a corresponding high-detailed, life-size, and immersive walk-through or macroscene is visualized in the surrounding room. Hence, SCAPE allows a seamless blending of dual-scales, dual-perspectives, and virtual and augmented components with which multiple users can concurrently interact from their individual viewpoints [23]. A schematic simulation of the SCAPE conceptual design, a prototype implementation, and sample views captured from the display are shown in Fig. 2. Particularly, Fig. 2(c) shows an augmented view of a 3-D virtual map with two physical user ID markers (e.g., #1 and #2) aligned with the virtual "blue" and "red" arrows, respectively, and it also shows an augmented view of a physical device with a high-resolution "pop-up" view of a portion of the low-resolution 3-D map which is physically blocked by the device.

In our experimental system, the two displays are driven separately by two Dell Precision Graphics workstations. The head motion of each user is detected by a 6DOF HiBall 3000 optical tracking system from the 3rdTech Inc. Fig. 1(b) shows the front view of the helmet with a HiBall 3000 sensor attached. A $6' \times 3.5'$ workbench and a $12' \times 12' \times 9'$ immersive wall display coated with retro-reflective film serve as the screens for the HMPDs, which provides a collaborative platform for multiple participants to seamlessly interact with an augmented virtual environment [22].

Fig. 3(a) and (b) illustrates the main components, the associated coordinate systems, and the extrinsic transformations of the SCAPE system in the physical and virtual worlds, respectively. All the coordinate systems used in this paper are right-handed. Each coordinate system is denoted as the combination of its origin and its axes. In the physical world, the key references are the physical world coordinate (PWC) associated with the workbench $W_P XYZ$, the HiBall tracker coordinate $TXYZ$, the HiBall sensor coordinate $S_1 XYZ$ for the head tracker, the eye coordinate $EXYZ$, the object coordinate $O_P XYZ$, and the object-sensor coordinate $S_2 XYZ$. As shown in Fig. 3(b), these physical references also have their corresponding counterparts in the virtual world coordinate

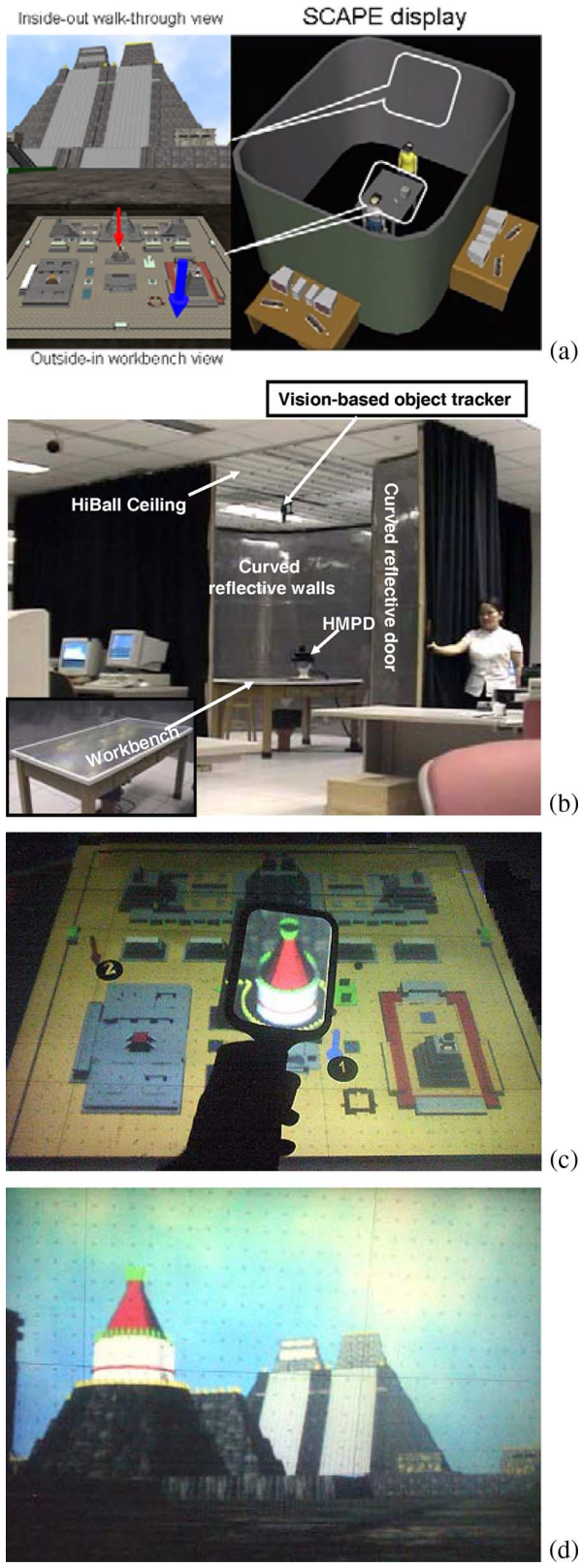

Fig. 2. SCAPE: A collaborative infrastructure. (a) Conceptual simulation. (b) Prototype implementation. (c) Sample view through the workbench. (d) Sample view through the surrounding wall display.

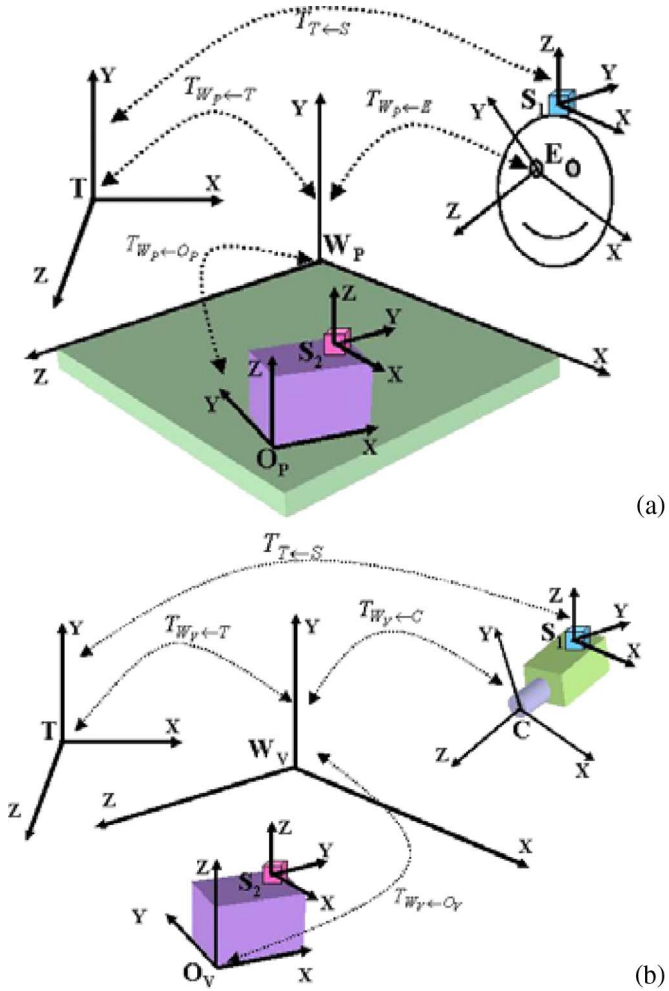

Fig. 3. Illustration of the physical and virtual world components and transformations. (a) Physical world illustration. (b) Virtual world illustration.

(VWC) $W_V XYZ$. Furthermore, the homogeneous coordinates of a point V in a given 3-D reference $AXYZ$ are expressed as $V_A(x_a, y_a, z_a, w_a)^T$. A rigid transformation from coordinate system A to coordinate B is expressed as a 4×4 homogeneous matrix $T_{B \leftarrow A}$. Therefore, a matrix transforming a vector V from A to B coordinates is expressed as $V_B = T_{B \leftarrow A} V_A$.

III. Computational Model and System Calibration Requirements

In an augmented environment, a user wearing an HMPD device observes the superposition of a virtual object and its physical counterpart. The requirements for achieving a visually accurate superposition include the following: 1) The VWC system $W_V XYZ$ should be aligned with its counterpart in the physical world $W_P XYZ$; 2) the virtual objects should be placed precisely in the VWC in the same way as their physical counterparts in the PWC; 3) the virtual cameras utilized to generate 2-D virtual images from 3-D virtual objects should have the same position and orientation transformation relative to the VWC as those of the eye reference to the PWC; and 4) the imaging properties of the virtual cameras should match with those of the display system. The key to an AR system calibration is to model and estimate these transformations and viewing parameters from the real-world setup for accurately configuring the virtual counterparts.

A. Modeling the HMPD Viewing System

As indicated by Fig. 1(a), the nature of the HMPD viewing system presents a two-step projection process: 1) an image displayed on the LCD screen is projected through the projection lens to form a real image in the physical world space; and 2) the projected image is viewed by a user through the exit pupil of the display. This two-step projection process is different from the viewing mechanism in a classical nonpupil-forming eyepiece-type HMD system, in which the exit pupil location of the display system is defined by a user's eye location rather than by a definite pupil position characterized by the optical design of a display system, like an HMPD. Without loss of generality, we model both of the projection processes through the projection lens and the eye viewing system, respectively, as pinhole imaging systems; each of which is further characterized by a projection center and a set of extrinsic geometric and intrinsic imaging transformations. For the eye viewing system, the projection center is usually the entrance pupil of the eye optics, which is referred to as the eyepoint.

In an HMPD system, the projection center of the display optics and the eyepoint are theoretically overlapped, and the optical axis of the projection optics is perpendicular to the projection image plane and is aligned with the viewing direction of the eye [Fig. 1(a)]. In practice, however, as shown in Fig. 4(a), the eyepoint E is often slightly displaced from the projection center O of the projection lens by $(\Delta x, \Delta y, \Delta z)$. The eye viewing direction, which is typically measured by a head-tracking or eye-tracking system attached to the helmet, is neither aligned with the optical axis nor perpendicular to the projection image plane. The optical axis is slightly deviated from the eye viewing direction by an angle of ϕ and is tilted relative to the projection image plane by an angle of φ. To make a 3-D virtual object $Q_{W\nu}$ (denoted with a blue circle) apparently superimposed upon its real counterpart P_{Wp} (denoted with a red rectangle), it is required that the 2-D projection of the virtual object $q_{I\nu}$ on the projection image plane should be collinear with the eyepoint and the real object. This indicates that, to maintain an accurate superposition, the 2-D projection varies as the position of the eyepoint or the viewing direction changes. Consequently, any position or orientation misalignments of the eye viewing subsystem from a perfect overlapping with the projection optical system, along with any tilting of the optical axis relative to the projection image plane, can cause a change in the imaging properties of the compound viewing system and demand an accurate system calibration to achieve visually accurate superposition. It thus becomes critical to take these practical displacements into account and to establish an accurate computational camera model, which is utilized to generate 2-D projection images from 3-D virtual objects. We can then estimate the transformations and imaging parameters of the two-step-projection viewing system and establish a mapping of these parameters with the computational camera model.

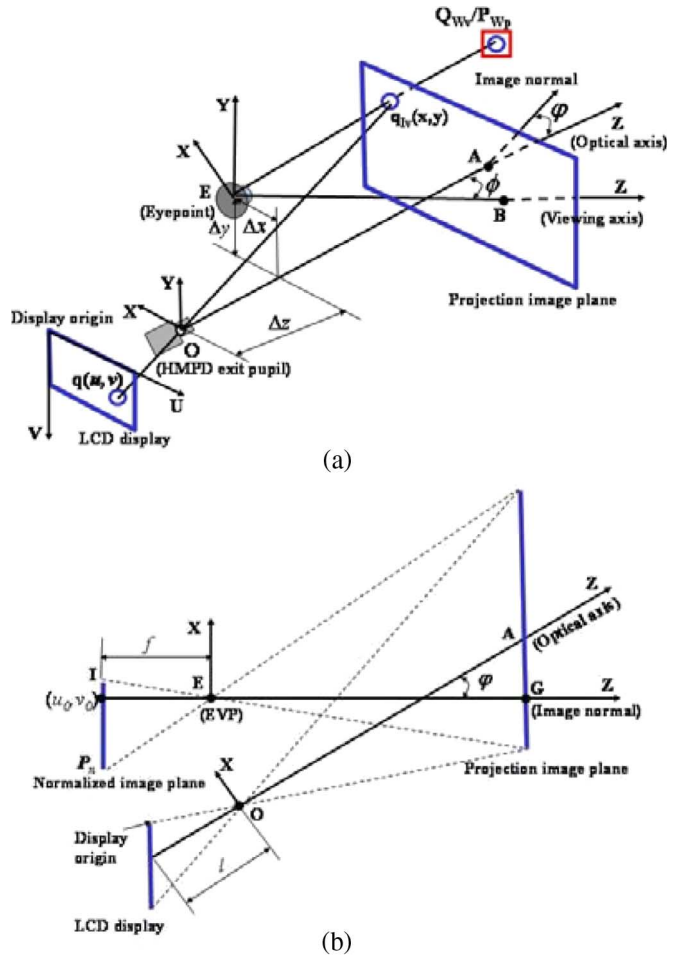

Fig. 4. Modeling the HMPD viewing system. (a) Illustration of the two-step projections in an HMPD system. (b) Illustration of an EVP system.

For implementation considerations, while an adequate camera model needs to take into account the practical misalignments, it also needs to be compatible with the established camera model in standard graphics libraries such as OpenGL, which is essentially a single-step projection system. We define an equivalent viewing projection (EVP) system which encompasses the two-step projections described above. The projection center of the EVP coincides with the user's eyepoint. Considering the fact that, in most graphics libraries, the image plane of a virtual camera is often assumably normal to the camera's viewing direction, the Z axis of the EVP reference is selected to be normal to the projection image plane of the LCD display, and its X and Y axes are aligned with the horizontal and vertical edges of the image plane. We further introduce a normalized image plane P_n. It is perpendicular to the Z axis of the EVP reference, with an intersection G. Its window size is the same as that of the LCD panel used in the HMPD system, measured by $m \cdot n$ pixels along the horizontal and vertical directions, respectively. The pixel scale factors of the normalized image plane in the X and Y directions are defined as (S_u, S_ν) mm/pixel, in which the ratio α, which is equal to S_u/S_ν, is defined as the aspect ratio used to compensate for a nonsquare pixel shape. The normalized image plane is indeed conjugated with the HMPD projection image plane or the LCD screen through the projection center E, and the 2-D projection relationship is illustrated in Fig. 4(b). Referenced to the image origin I, which corresponds to the projection of the LCD panel origin (usually the upper-left corner pixel), the pixel distance (u_0, ν_0) (in pixels) of the Z-axis intersection point is defined as the center offset of the EVP system. Finally, the distance f (in millimeter) from the eye position to the normalized plane P_n is defined as the equivalent focal length (EFL) of the EVP system. For a more accurate optical modeling, the EVP system can also take into account lens distortions present in the HMPD optical system. For instance, lens distortion can be modeled rather accurately as a simple radial distortion with a first-order distortion coefficient k_1.

The EVP is an abstract viewing device which is compatible with the camera model in most graphics libraries, and thus, it was utilized to configure a virtual camera for 2-D image rendering from 3-D virtual objects. Given a 3-D virtual point $Q_{W_V}(x_{w\nu}, y_{w\nu}, z_{w\nu}, 1)$ in the VWC, its 2-D projection $q_{I\nu}$ on the normalized image plane is given by

$$q_{I\nu} = M_{\text{EVP}} T_{C \leftarrow W\nu} Q_{W\nu} \tag{1}$$

where $T_{C \leftarrow W\nu}$ consists of a set of extrinsic rigid-body transformations that place the EVP reference in the VWC and M_{EVP} represents the intrinsic imaging transformation of the EVP system.

The imaging parameters of the EVP system fall into three categories of operation: projection, warping, and clipping [12]. The projection parameters specify a projection transformation that maps a 3-D scene point in the world space onto its 2-D representation in the normalized image plane. The warping parameters specify a distortion transformation that imitates the lens distortion. The clipping parameters specify a view frustum within which a scene point is captured inside a viewing window. Among them, only the projection and warping parameters affect the 3-D-to-2-D mapping. The clipping parameters, however, do not affect the mapping, but affect the comfortable depth range of the viewing system. The focused viewing distance of the display system should be adjusted accordingly based on application requirements [41]. Therefore, the intrinsic imaging transformation of the EVP M_{EVP} is further composed of two components: an intrinsic projection transformation and a warping transformation. The projection transformation M_{int} is given by

$$M_{\text{int}} = \begin{bmatrix} -f/S_u & 0 & u_0 \\ 0 & -f/S_\nu & \nu_0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2}$$

The distortion transformation of the EVP M_{dist} can be modeled rather accurately as simple radial distortions by

$$\begin{cases} x_d = x(1 + k_1 * r^2) \\ y_d = y(1 + k_1 * r^2) \end{cases} \tag{3}$$

where (x_d, y_d) and (x, y) are the distorted and undistorted projections on the normalized image plane, respectively, and $r^2 = x^2 + y^2$.

TABLE I
TRANSFORMATIONS IN THE PHYSICAL AND VIRTUAL WORLDS

Transformations		Components: Physical World ($W_P XYZ$)	Components: Virtual World ($W_V XYZ$)	Attributes	Calibration Category
Viewing Transf.	Position & Orientation $T_{W_P \leftarrow E} = T_{W_V \leftarrow C}$	$T_{W_P \leftarrow T}$: Tracker transformation	$T_{W_V \leftarrow T}$: Tracker transformation	Extrinsic, Unknown, Fixed	Tracker calibration
		$T_{T \leftarrow S1}$: Head sensor transformation	$T_{T \leftarrow S1}$: Head sensor transformation	Extrinsic, Known, Varying	N/A
		$T_{S1 \leftarrow E}$: Eye transformation	$T_{S1 \leftarrow C}$: Camera transformation	Extrinsic, Unknown, Varying	Display calibration
	Imaging Transf. M_{EVP}	M_{int} : Projection transformation	M_{int} : Projection transformation	Intrinsic, Unknown, Partially Fixed	
		k_1: Optical distortion	k_1: Warping		
		Field-of-view and depth range	Viewing frustum clipping parameters		

On the other hand, the intrinsic and extrinsic transformations of the EVP system are equivalent to those of the two-step viewing system. The position of the EVP reference is the same as the translation component of the eye reference, and the orientation of the EVP relates to the orientation of the projection image plane or the LCD screen of the HMPD viewing system relative to the eye viewing direction. While the window size, pixel scale factors, and aspect ratio, as well as the first-order distortion coefficient, are the same as their equivalents in an HMPD system, it is important to note that the EFL, f, and the center offset (u_0, ν_0) of the EVP system are related to but not the same as their counterparts of the HMPD projection optics. Given a 3-D point $P_{W_P}(x_{\text{wp}}, y_{\text{wp}}, z_{\text{wp}}, 1)$ in the PWC and observed from the pupil position, its 2-D projection p_{I_P} on the normalized display window of the EVP system is given by

$$p_{I_p} = M_{\text{int}} T_{E \leftarrow W_p} P_{W_p} \tag{4}$$

where $T_{E \leftarrow W_p}$ is a viewing orientation transformation that defines the EVP reference in the PWC. Note that the distortion component of the EVP imaging transformation is excluded in (4) due to the fact that the real-world is viewed directly through a beamsplitter in HMPD, rather than through the projection optics like the virtual viewing path. In our current system setup, the viewing orientation transformation $T_{E \leftarrow Wp}$ can be further expressed with its correspondence in the PWC by

$$T_{E \leftarrow Wp} = T^{-1}_{S1 \leftarrow E} T^{-1}_{T \leftarrow S1} T^{-1}_{Wp \leftarrow T} \tag{5}$$

where $T_{Wp \leftarrow T}$, referred to as the tracker transformation, gives the position and orientation of the tracker reference in the world space; $T_{T \leftarrow S1}$, referred to as the head sensor transformation, is the position and orientation of the moving sensor in the transmitter space; and $T_{S1 \leftarrow E}$, referred to as the eye transformation, specifies the eyepoint position and viewing orientation in the head sensor reference.

Provided that the assumption that the virtual and real points and their 2-D projections overlap, and the PWC and VWC are aligned in an augmented environment, we have $Q_{Wp} = P_{W\nu}$ and $T_{E \leftarrow Wp} = T_{C \leftarrow W\nu}$. By combining (1) through (5), the pixel coordinates (u, ν) of the 2-D projection $q_{I\nu}$ on the normalized image window are given by

$$\begin{cases} u = u_0 - \frac{f}{S_u} x - \frac{k_1 f}{S_u} x(x^2 + y^2) \\ \nu = \nu_0 - \frac{f}{S_\nu} y - \frac{k_1 f}{S_\nu} y(x^2 + y^2) \end{cases} \tag{6}$$

where $x = (r_{11}x_{\text{wp}} + r_{12}y_{\text{wp}} + r_{13}z_{\text{wp}} + t_x) \ / (r_{31}x_{\text{wp}} + r_{32}y_{\text{wp}} + r_{33}z_{\text{wp}} + t_z)$, $y = (r_{21}x_{\text{wp}} + r_{22}y_{\text{wp}} + r_{23}z_{\text{wp}} + t_y) \ / (r_{31}x_{\text{wp}} + r_{32}y_{\text{wp}} + r_{33}z_{\text{wp}} + t_z)$, and $T = [t_x \ t_y \ t_z]^T$ and $R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$ represent the translation and rotation components in the $T_{E \leftarrow Wp}$ transformation, respectively.

Equation (6) explicitly links a 3-D point in the world space $\overrightarrow{P_{Wp}}$ to a pixel (u, ν) in the display coordinates. Given a sufficient number of world-image correspondences, through the well-known least-square fitting methods [35], we can obtain the estimates of the extrinsic viewing orientation transformation $M_{\text{ext}}(R, T)$, the projection parameters (f, S_u, S_ν, u_0, ν_0,), and warping parameter k_1, which is referred to as the display calibration.

B. Calibration Requirements and Strategy

The viewing transformations involved in the system setup, including both the physical and virtual worlds, are summarized in Table I. Besides intrinsic and extrinsic, the attributes of each transformation are further categorized as known or unknown and fixed or varying. A subset of these transformations, marked as "known" in Table I, is obtained directly from the associated sensors. Other transformations marked "unknown" need to be estimated via the calibration process. All the unknown transformations need to be determined through a calibration procedure. Furthermore, the transformations marked "fixed" in the table rarely change once the system is set up, and thus, only a one-time calibration is needed. For example, the unknown tracker transformation remains fixed once the setup is established. The eye transformation, however, may be slightly different from user to user and may even change during each session due to slippage of the helmet. Except for the distortion coefficient, the imaging parameters of the EVP system may change due to user dependence.

The calibration of AR systems based on optical see-through HMDs is much more challenging than a video-based AR system owing to two facts: 1) There is no practical mechanism in capturing the image of the real world seen directly by the eye; and 2) there is no direct measurement for the correspondences between real objects, virtual objects presented through an HMD, and their projections on the eye retina. Therefore, such AR displays have to be calibrated with a user in the loop to establish the correspondences between real and virtual objects. This process is considered to be complex and mentally demanding. Therefore, the accuracy of calibration is highly user dependent and lacks reliability [43]. A trained user with a well-prepared knowledge of display calibration and a novice user could achieve considerably different calibration accuracy [43]. Therefore, designing easy alignment tasks and minimizing the number of required correspondences are highly desirable.

Instead of taking a one-step online calibration, which usually requires a nontrained user to perform a considerably large number of alignment tasks for calibration before proceeding with an experimental session, a two-step strategy was adopted to calibrate our HMPD system: offline and online calibration steps. The offline calibration step is performed by a trained user with a better understanding of the calibration procedures; its aim is to optimally estimate the unknown fixed transformations and initially estimate the unknown varying parameters. Since this step is only required once for a new display system, it is worthy of time-consuming methods to achieve the best possible accuracy. On the other hand, the online calibration step is performed during an application session by a regular user, and its aim is to refine the varying parameters based on initial offline estimates. Because this step might be performed each time when a user puts on the helmet or when helmet slippage occurs during a session, a fast and easy calibration method is required to minimize a user's cognitive load. Overall, the offline calibration offers best accuracy for the unknown fixed transformations and adequate accuracy for the user-dependent parameters. It meets most of the application requirements. An online calibration is only necessary when a more accurate estimation of the user-dependent parameters is demanded for a better registration quality. In this case, the initial estimates from an offline calibration make it possible to significantly simplify the fitting model and thus enable a fast online calibration procedure to only fine-tune the user-dependent parameters. This paper focuses on the offline calibration step and its registration quality evaluation; we are developing a fast online calibration method based on the offline estimates.

Finally, the left and right viewing optics are not necessarily parallel. The left and right optics may be purposely diverged or converged to achieve a larger overall FOV. Therefore, the viewing transformations corresponding to the left and right virtual cameras have to be calibrated separately.

IV. Calibration Procedures and Methods

Based on the calibration requirements summarized in Table I, the calibration steps to obtain complete estimates of the viewing transformations include: 1) Tracker transformation calibration; and 2) Display calibration.

A. Tracker Transformation Calibration

While the head-tracker measurement explicitly gives the sensor transformation $T_{T\leftarrow S1}$, it is necessary to calibrate the tracker transformation $T_{Wp\leftarrow T}$. The exact origin and axis of the HiBall tracking system are under-defined and depend on the actual installation. Therefore, the following method is used to estimate the $T_{Wp\leftarrow T}$. The manufacture supplies a HiBall stylus that enables point-by-point position measurement. In the designated PWC, when aligning the tip of the stylus with the origin, a selected point along the X axis, and a point along the Y axis of the world coordinate system, we recorded the position measurements of the sampled points, $\overrightarrow{W}$, $\overrightarrow{X}$, and $\overrightarrow{Y}$, respectively. The normalized vectors of the X, Y, and Z axes in the tracker coordinates are given by $\overrightarrow{X'} = (\overrightarrow{X} - \overrightarrow{W})/\|\overrightarrow{X} - \overrightarrow{W}\|$, $\overrightarrow{Y'} = (\overrightarrow{Y} - \overrightarrow{W})/\|\overrightarrow{Y} - \overrightarrow{W}\|$, and $\overrightarrow{Z'} = (\overrightarrow{X'} \times \overrightarrow{Y'})/\|\overrightarrow{X'} \times \overrightarrow{Y'}\|$, respectively. To ensure the axes are orthogonal, $\overrightarrow{Y'}$ is replaced by $\overrightarrow{Y'} = (\overrightarrow{Z'} \times \overrightarrow{X'})/\|\overrightarrow{Z'} \times \overrightarrow{X'}\|$. Therefore, the tracker transformation is given by

$$T_{W\leftarrow T} = \begin{bmatrix} \overrightarrow{X'} & \overrightarrow{Y'} & \overrightarrow{Z'} & W \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{7}$$

To achieve good results, the separation of X and Y from W should be large enough. To improve the accuracy of single-point measurements using the HiBall stylus, at each position, a large number of samples (e.g., 50) were taken at different orientations, and their average is used to compute the normalized vectors. We also took measurements at multiple positions along the axis and averaged the results of these normalized vectors. The actual transformation measured in our system shows about 0.5°, 0.8°, and 1° twist around the XYZ axes.

B. Display Calibration

The computational model for the display calibration described in Section III is similar to the model for camera calibration used in computer vision [44]; thus, the basic strategy for the display calibration is to determine a sufficient number of world-image correspondences and to select proper fitting algorithms from camera calibration methods to estimate the intrinsic and extrinsic parameters. However, the key differences of display calibration from camera calibration are reflected in the following aspects.

1) An HMD or HMPD display involves two mapping processes from the 2-D display device to the 2-D image plane through the display optics and from the image plane to a 3-D world through an observer, while a camera involves a single and direct mapping process from a 3-D scene to its 2-D image. As a result, the eye position of the observer affects the mapping of the image with the 3-D world.
2) In display calibration, an observer needs to manually determine world-image correspondences one by one, while in a camera calibration, the camera can capture an image of the real world and automatically determine the correspondences by established feature detection algorithms

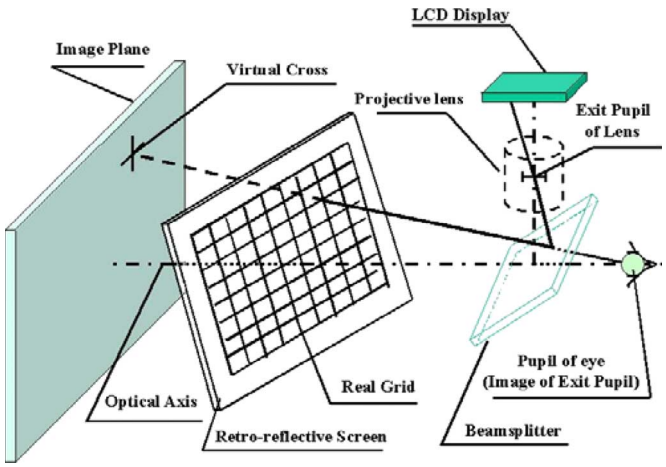

Fig. 5. Illustration of the HMPD calibration method and setup.

with subpixel accuracy [44]. Consequently, the accuracy of the correspondence matching relies on the resolvability of the observer.

3) Display calibration requires absolute measurements of calibration points relative to a predefined world reference, while in a camera calibration, the world reference is typically associated with its calibration target. As a result, the accuracy of the world coordinate measurements relies on both the sensor measurements and calibration target in display calibration, while it only depends on the target accuracy in camera calibration. Additionally, in HMPDs, the necessity for a retro-reflective screen to view the image requires an especially made calibration target. Consequently, a physical target should be printed either on a transparent sheet or directly on a retro-reflective film.

These differences reflect two major challenges in display calibration: 1) accurate measurements of the world coordinates of calibration points; and 2) precise matching of world-to-image correspondences. This paper describes an MCM strategy to locate a sufficient number of world-image correspondences [20]. Fig. 5 illustrates the setup used for the MCM calibration process. The calibration target is an $M \times N$ grid pattern drawn on a flat retro-reflective screen. The calibration task is to move a cross stimuli displayed through the HMPD such that it is aligned with each of the grid intersections and to record the corresponding stimuli pixel coordinates. To the cross-grid alignment, the operator can directly observe either through the exit pupil of the HMPD or through a secondary camera that is properly positioned at the exit pupil. The former is referred to as a direct observation approach, and the latter is referred to as an image-based observation approach. The calibration procedure consists of the following steps.

1) Mount the HMPD on a fixed platform, record the measurement of the head sensor $(\overrightarrow{S}, \overrightarrow{\Theta_S})$, and obtain the sensor transformation $T_{S1\leftarrow T}$. Empirically, when only one sensor is used for both the head and stylus trackers, this measurement should be taken at the end of the experiments. Whenever the sensor is reinstalled on the helmet, a full calibration procedure is needed.
2) Estimate the tracker transformation $T_{W_p\leftarrow T}$. Here, rather than using the application world reference W_pXYZ, we actually recommend defining a calibration world reference W_CXYZ, which is close to the calibration platform, to ensure that the scale of the world coordinate measurements is compatible with that of the calibration target. Compatible scales are expected to lead to more stable and accurate fitting estimates. The transformations and parameters estimated from the display calibration are independent of the choice of the world reference.
3) Place the calibration target at a fixed attitude, record the coordinate measurements of the $M \times N$ grid intersections using the HiBall stylus $[P^k_{i,j}]_T$, and compute their corresponding world coordinates $([P^k_{i,j}]_{Wp})$ from the transformation $T_{W_P\leftarrow T}$, where k represents the kth target attitude and i and j represent the row and column indexes of the grids, respectively.
4) Present a cross stimulus on the display under calibration. The operator moves the virtual cross to align it with each of the grid intersections on the target and records the corresponding pixel coordinates $Q^k_{i,j}(u, \nu)$.
5) Change the target attitude, and repeat steps 3) and 4) at least three times $(k \geq 3)$. Empirically, the angular difference between two target attitudes is recommended to be large enough to ensure good convergence.
6) Form the calibration matrix (6) from the $P-Q$ correspondences, and apply selected fitting methods to solve the intrinsic and extrinsic parameters [44], which give the estimates of the EVP parameters, including the extrinsic transformation $T_{E\leftarrow W_p}$, the focal length f, aspect ratio (S_u, S_ν), center offsets (u_0, ν_0), and distortion coefficients (k_1).
7) Insert $T_{S1\leftarrow T}$ from step 1), $T_{Wp\leftarrow T}$ from step 2), and $T_{E\leftarrow W_P}$ from step 7) into (5) to process in the computer the eye transformation $T_{S1\leftarrow E}$.
8) Insert the intrinsic parameters into (2) to compute the projection transformation M_{int}.

In the calibration experiments, we selected the image-based observation approach and sampled 12×10 points on each calibration target. We repeated the task at nine different target locations and attitudes and obtained a total of 1080 correspondences. We ensured that the nine targets are different in both position and orientation. This procedure is applied to the left and right arms of the HMPD separately. The intrinsic and extrinsic parameters estimated from the samples are listed in the Table II.

C. Implementation of the Computational Model

The computational model for graphics generation needs to be fully customized to adopt the EVP model described in Section III, rather than using the default functions implemented in OpenGL libraries. Particularly, instead of assuming symmetry of the left and right virtual cameras, it is essential to implement customized viewing orientation transformations, projection transformations, and viewing frustums individually for the cameras corresponding to the left and right eyes, which are obtained through the calibration procedures.

TABLE II
INTRINSIC AND EXTRINSIC PARAMETERS OF THE EVP SYSTEM

Parameters	Left	Right
Equivalent Focal length (mm)	35.1473	35.09793
FOV (degrees)	41.17(H), 31.76(V)	41.22 (H), 31.81(V)
Radial distortion	2.2% (at the marginal field) 1.08% (at 0.707 field)	3.56% (at the marginal field) 1.75% (at 0.707 field)
Pupil position w.r.t. the Hiball sensor reference (mm)	X=-23.3953 (H) Y=-121.5083(V) Z=-14.2093	X=39.51(H) Y=-127.0863(V) Z=-11.4750
Eye transformation (i.e. display orientation w.r.t. sensor reference)	$\begin{bmatrix} 0.9999 & 0.0038 & -0.0082 \\ -0.0034 & 0.9987 & -0.0511 \\ -0.083 & 0.0511 & 0.9987 \end{bmatrix}$	$\begin{bmatrix} 0.9999 & 0.0004 & -0.0037 \\ -0.0007 & 0.9975 & -0.0705 \\ 0.0037 & 0.0705 & 0.9975 \end{bmatrix}$
Angles between the axes of the display reference and that of the sensor reference (degrees)	X axis: 0.5139 Y axis: 2.9371 Z axis: 2.9665	X axis: 0.2158 Y axis: 4.042 Z axis: 4.048
Offsets from display center(pixels)	ΔH=-7, ΔV=1	ΔH=27, ΔV=-22

For each virtual camera, the corresponding eye, sensor, and tracker transformations form a composite viewing orientation transformation. This combination specifically compensates for the pupil offsets relative to the head sensor reference and the variation of the left/right image plane orientations. It further requires that the graphics software separately defines the viewing transformations for the left and right cameras to generate stereo image pairs, instead of simply applying an offset of interpupillary distance (IPD), which is the typical practice of stereo image generation.

In OpenGL, the projection and clipping parameters needed to specify imaging properties of a virtual camera include viewing plane distance, view window dimensions, viewing frustum asymmetry, and far and near clipping planes [12]. In our implementation, these parameters are customized for each individual camera based on the calibration results listed in Table II. Specifically, the viewing plane distance D is set to be proportional to the calibrated equivalent focal distance f. To ensure comfort, D is set to approximate the HMPD image distance. The viewing window dimensions (W, H) are computed from the calibrated intrinsic parameters (f, S_u, S_ν), the viewing plane distance D, and the microdisplay dimensions (w, h). The viewing frustum asymmetries are computed from (u_0, ν_0), which is a conjugate to the horizontal and vertical center offsets of the Z-axis intersection of the EVP system with the image plane. For each camera, both the horizontal and vertical asymmetries are taken into account to compensate for hardware asymmetries, instead of only having horizontal asymmetry configurable in a conventional practice. The far and near clipping planes are set up to be the farthest and nearest depth limits, respectively, specified by convergence tolerances by the study in [41]. For example, in an arm-length application, if we set up the image distance of the display as 0.6 m, the comfortable depth range is from 0.48 to 0.8 m. When lens distortion needs to be compensated, a prewarping transformation is formed from the calibrated distortion coefficient (k_1) and is preapplied before rendering the window context in the viewport.

V. EVALUATION EXPERIMENTS AND RESULTS

In this section, we will describe the experimental setup specification and discuss a set of experiments to evaluate how the number of correspondence samples, attitudes of the calibration target, and the selection of an observation approach affect the accuracy and convergence of the calibration results.

A. Experimental Setup Specification

In the experiments, the calibration target is a 14 × 13 grid pattern with a pitch of 40 mm. The grid line is 1-mm wide, which corresponds to an angle of $3.5'$ in the eye space for a 1-m viewing distance. The stimulus presented through the HMPD is a cross with 1-pixel line-thickness, when making direct observation, or with 2-pixel line-thickness, when making image-based observations. A 1-pixel line in the display space corresponds to 1.1 mm on the image plane at a distance of 1 m, which corresponds to an angle of $3.5'$ in the eye space.

B. Calibration Accuracy Analysis

The accuracy and convergence of the display calibration rely on the accuracy of the world coordinate measurements, the accuracy of the correspondence matching, and the number and the distribution of calibration samples. In our experiments, the accuracy of the world coordinate measurements relies on that of the HiBall stylus and the calibration target. Due to the highly reflective environments in our system, the stylus has a limited accuracy, between 2 and 5 mm, which corresponds to approximately 2- and 5-pixel error in the display space. Instead of directly measuring world coordinates for every grid, we utilize a similar calibration approach described in Section IV-A to obtain a target-tracker transformation and then compute the grid coordinates in the tracker reference based on their local measurements. Our experiences confirm that this approach can improve the calibration accuracy.

The accuracy of correspondence matching mainly relies on the accuracy of the target, the resolution of the display, and the resolution of the observer. The accuracy of the calibration target is less than 0.5 mm, which corresponds to less than 0.5 pixel in the display space or $1.75'$ in the eye space. Given that the image plane was set to be about 1-m away, which is measured from the pupil, and the pixel size of the display is 42 μm that approximately corresponds to 1 mm after optical magnification, the angular resolution of the magnified image is about $3.5'$

in the eye space. Therefore, the accuracy of the target matches the resolution of the display. In this configuration, the matching accuracy would mainly rely on the observer.

The raw data of world coordinate measurements (P) and correspondence matching (Q) are noisier and less accurate than the subpixel accuracy that can be typically achieved in a camera calibration process. Consequently, we anticipate that a large number of correspondence samples are necessary to achieve a high accuracy and a stable convergence.

C. Evaluation Method

In order to study the relationship of the display calibration accuracy with different configuration parameters such as the number of samples, variation of target attitudes, and selection of an observation approach, evaluation is conducted by comparing the difference between the computed projections of evaluation targets in the display space and their ground-truth projections obtained along with correspondence matching. The target and task used in the evaluation experiments are the same as those used in the calibration procedure. The evaluation procedure includes the following steps.

1) Position the evaluation target at a fixed attitude, record the head sensor measurements of the grid intersections using the HiBall stylus $[P^k_{i,j}]_T$, and compute their coordinates in the world coordinates $[P^k_{i,j}]_{Wp}$ using the transformation $T_{Wp \leftarrow T}$ obtained from the tracker calibration.
2) Compute the projection of each sampled point in the display space $\widehat{Q}^k_{i,j}\ (u, \nu)$ from the calibrated intrinsic and extrinsic viewing transformations ($\widehat{M}_{\text{int}}\ (\widehat{f}, \widehat{S_u}, \widehat{S_\nu}, \widehat{u_0}, \widehat{\nu_0}, \widehat{k_1})$ and $\widehat{M}_{\text{ext}}\ (\widehat{R}, \widehat{T})$).
3) Ask the operator to align a virtual cross with each of the grid intersections on the target, and record the corresponding pixel coordinates $Q^k_{i,j}(u, \nu)$.
4) Compute the difference between the computed and the ground-truth projections: $e^k_{i,j}(u, \nu) = |Q^k_{i,j}(u, \nu) - Q^k_{i,j}(u, \nu)|$.
5) Change the target attitude, and repeat the steps 1) through 4).
6) Evaluate the means and standard deviations (STDs) of the samples at all different attitudes.

D. Experiments and Observations

In the following sections, we describe a set of experiments and our observations.

1) Number of Calibration Samples: Manually determining a large number of correspondences is a difficult and time-consuming task. Minimizing the number of samples while preserving reasonable accuracy and convergence is desirable. Furthermore, studying the relationship between the sample number and calibration accuracy provides an informative guide for experiment design. For this purpose, we selected the 1080 left-eye correspondence samples obtained through image-based observation, as described in Section IV-B. At different sample intervals, we down-sampled the number of points on each of the nine calibration targets, and formed 12 groups of subsamples. The sample intervals are the following: 1, 2, 3, 4, 5, 7, 9, 11, 20, 33, 50, and 60, which approximately corresponds to 120, 60, 40, 30, 24, 17, 13, 11, 6, 4, 3, and 2 samples at each target attitude, respectively. Then, the extrinsic and intrinsic transformations corresponding to each group of the subsamples were computed and denoted as $(\widehat{M}_{\text{ext}}, \widehat{M}_{\text{int}})_i$, where i is the group index of the subsamples and $i = 1, \ldots, 12$.

In the evaluation experiments, correspondence matching was performed through direct observation at the exit pupil. Measurements were taken at three different target attitudes. At each attitude, 14×10 points were sampled. Based on the world coordinates of the points measured with the stylus, the projections of these samples in the display space were computed from the estimated extrinsic and intrinsic transformations $(\widehat{M}_{\text{ext}}, \widehat{M}_{\text{int}})_i$. Fig. 6(a) and (b) shows the error distributions for evaluation target 1, corresponding to $i = 1$ (1080 samples) and $i = 11$ (27 samples), respectively. Fig. 6(c)–(e) shows the error distributions for $i = 9$ (54 samples), corresponding to the three different evaluation targets, respectively. Fig. 6(f) and (g) shows the means and STDs of the projection errors corresponding to the three evaluation targets. The graphs illustrate that dense samples achieved better accuracy than sparse samples, and about 40 samples widely distributed on nine calibration targets can possibly converge. However, a stable convergence needs more than 300 samples. Our experience tells us that the small number of samples has to maintain a wide distribution in the world space and to be measured accurately in order to achieve convergence.

2) Number of Calibration Targets: To study the relationship of calibration accuracy and convergence with the number of calibration target attitudes, we did the following experiments.

Experiment 1: Instead of performing down-sampling across the nine calibration targets, as described in the last experiment, we randomly selected four targets from the nine calibration targets for the left eye and applied the down-sampling procedure on the four selected targets. The sampling intervals increase from 1 to 12 using the incremental step of one. Then, the extrinsic and intrinsic transformations corresponding to the 12 groups of subsamples were estimated, denoted as $(\widehat{M}^4_{\text{ext}}, \widehat{M}^4_{\text{int}})_i$, $(i = 1 \ldots 12)$, and the same evaluation method was applied to these 12 groups of estimates. Fig. 7(a) and (b) shows the means and STDs of the projection errors corresponding to the three evaluation targets. Compared with the down-sampling results of nine targets in Fig. 6, it shows that more targets can improve calibration accuracy. The mean of projection errors is about 3 pixels in nine targets, while it is about 5 pixels in four targets.

Experiment 2: Instead of performing down-sampling on each target at certain intervals, we utilized the full samples at each selected target and computed the extrinsic and intrinsic transformations at eight different target combinations $(\widehat{M}_{\text{ext}}, \widehat{M}_{\text{int}})_j$, where j represents the number of calibration targets and $j = 2, \ldots, 9$. The same evaluation was applied on the eight groups of estimated transformations. Fig. 7(c) and (d) shows the means and STDs of the projection errors corresponding to the three evaluation targets. The graphs further confirmed the

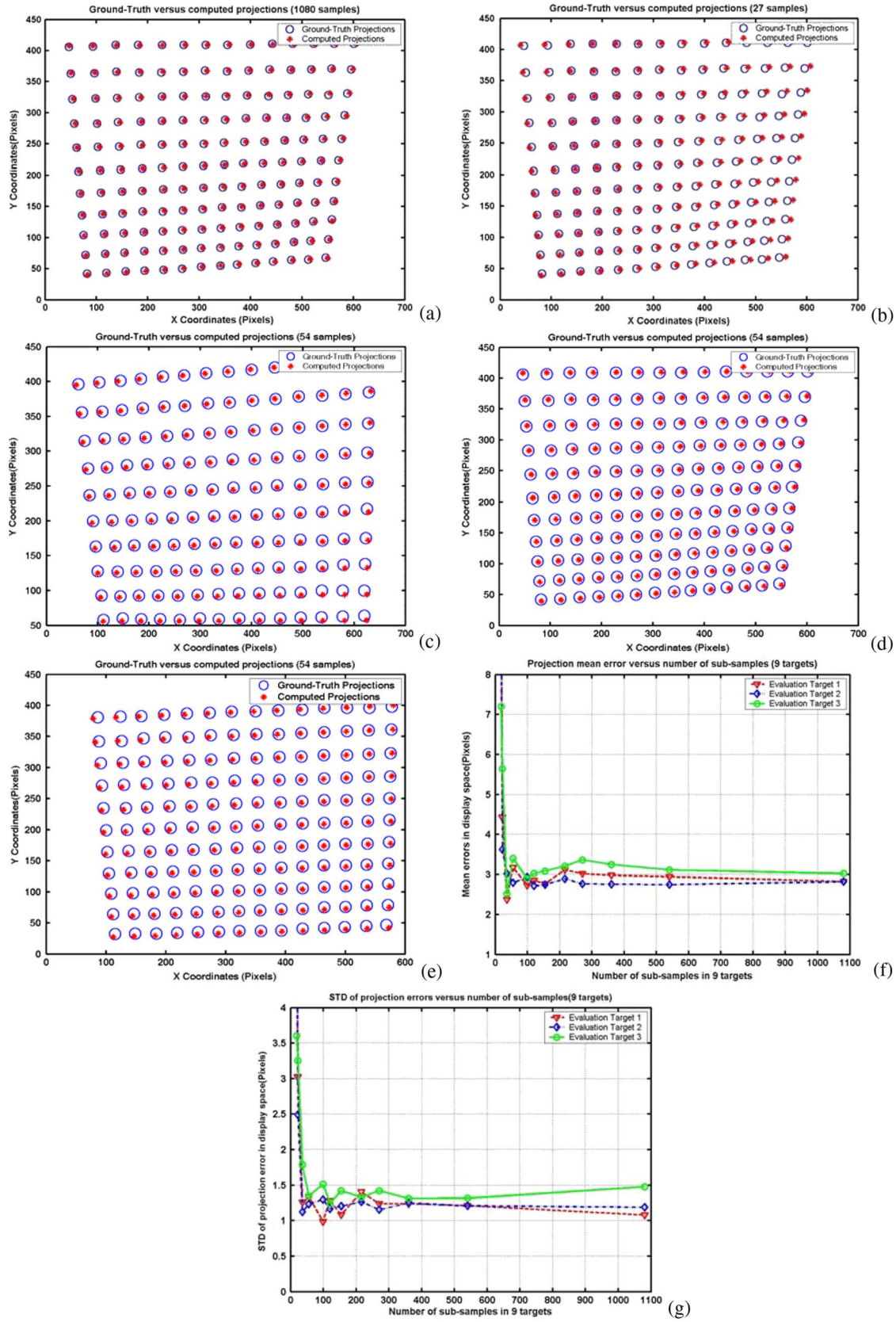

Fig. 6. Number of calibration samples affects the accuracy and convergence of the display calibration. (a) Registration error distribution for the evaluation target 1, in which the viewing transformation is estimated from 1080 samples on nine calibration targets ($i = 1$). (b) Registration error distribution for the evaluation target 1, in which the viewing transformation is estimated from 27 samples on nine calibration targets ($i = 11$). (c)–(e) Registration error distributions corresponding to the three evaluation targets, respectively, in which the viewing transformations are estimated from 54 samples on nine targets ($i = 9$). (f)–(g) Mean and STD of the projection errors for the three evaluation targets, in which the viewing transformations are estimated from 12 groups of subsamples on nine calibration targets.

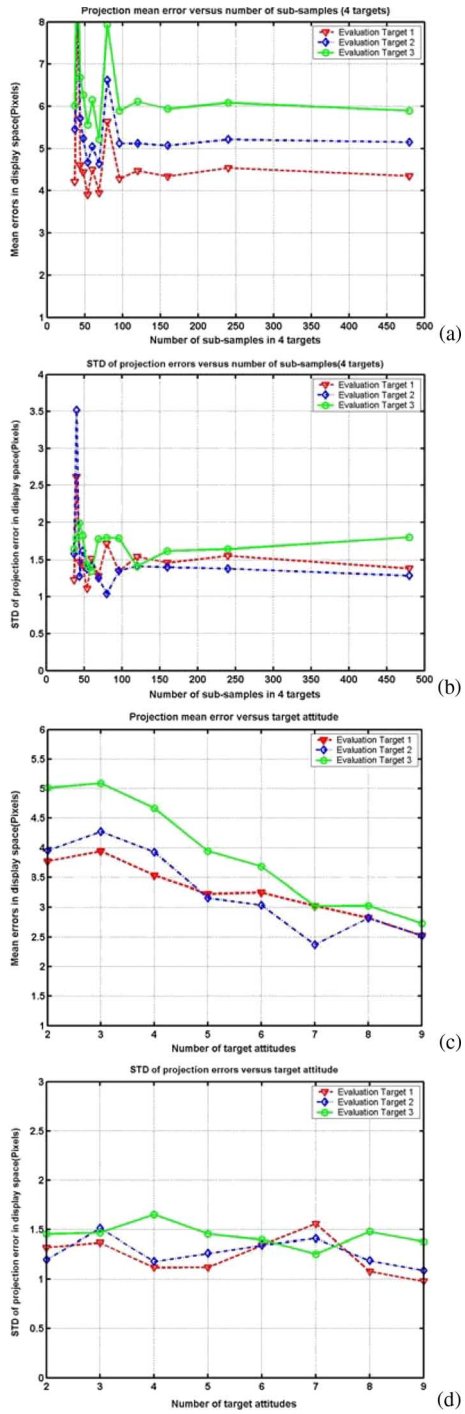

Fig. 7. Number of calibration targets affects the accuracy and convergence of the display calibration. (a)–(b) Mean and STD of projection errors for the three evaluation targets, in which the viewing transformations are estimated from 12 groups of subsamples on four selected calibration targets. (c)–(d) Mean and STD of projection errors for the three evaluation targets, in which the viewing transformations are estimated from increasing number of calibration targets.

observation that increasing the number of calibration targets can improve calibration accuracy and convergence, while two calibration targets with a total of 240 samples can possibly achieve convergence.

3) Direct Observation Versus Image-Based Observation: An operator can either directly make an observation at the exit pupil of the display or indirectly make it through a secondary calibration camera placed at the pupil. Direct observation offers the best imitation of a real application, and a human eye offers a much better resolution than a camera does. For example, in terms of resolvability, the angular resolution of human eyes is up to $1'$ at the fovea, but the resolution of a video camera ($2/3''$ charge-coupled device sensor with 640×480 pixels and 6.5-mm lens) is about $7'$. Consequently, the correspondence matching accuracy may not be as high as what is achieved with human eyes. An observer can also best position his or her eyes at the proper observing position. However, asking a subject to identify a large number of correspondences while keeping his or her head steady is a far more difficult task.

Relying on a secondary camera to perform the calibration task greatly decreases the workload on the operator and ensures a fixed observation position. As a result, it allows performing a large number of samples at more different target attitudes without overloading the operator. Experiments described in the previous sections verify that a large number of samples can improve the fitting accuracy and stability. On the other hand, the synthetic image captured by the camera suffers from low contrast, low resolution, and low-level luminance due to the limited resolvability of available video cameras as well as to the low luminance of the HMPD. Furthermore, the calibration target and the image plane are not coincident, such that it is impossible to capture their focused images simultaneously. This defocus might deteriorate alignment resolution. Finally, it is more difficult to make the camera entrance pupil fully overlap with the exit pupil of the HMPD and to align the camera axis with that of the HMPD axis.

In the comparative calibration experiments, for the left arm of the display, we conducted correspondence matching tasks by both direct and image-based observations. In each case, we obtained a total of 480 samples from four different target orientations, each with 12×10 points. The two groups of data were used to estimate the viewing transformations, denoted as $(\widehat{M}_{\text{ext}}, \widehat{M}_{\text{int}})_{\text{eye}}$ and $(\widehat{M}_{\text{ext}}, \widehat{M}_{\text{int}})_{\text{cam}}$. Then, we evaluated the results with six evaluation targets. Fig. 8(a) and (b) shows the means and STDs of the projection errors for the six evaluation targets, corresponding to direct and image-based observations, respectively. Although we did not observe a significant difference in the mean errors for the six evaluation targets, the less fluctuation of projection errors for image-based observation is coincident with the fact that a secondary camera remains at a fixed observation position and introduces less noise to raw data, while it is very difficult to achieve a steady observation for human beings. For example, the evaluation targets 1 through 4 were sampled in the same series of matching experiments as the four calibration targets used in this comparative calibration experiments through direct observation, while the subject took a brief break before he sampled targets 5 and 6. Both the mean and STD curves reflected the difference, but the STD of the image-based approach remains consistent.

VI. Testbed Example

Given that we expect augmentation in SCAPE will mostly take place on the workbench, a testbed application, namely

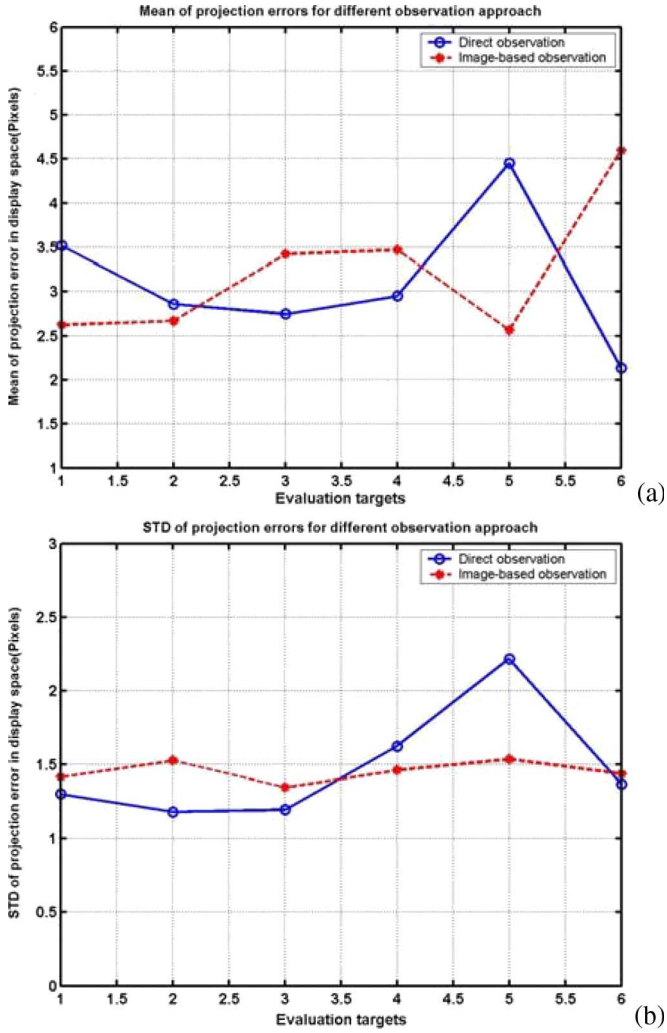

Fig. 8. Selection of observation approaches affects the accuracy and convergence of the display calibration. (a)–(b) Mean and STD of projection errors for six evaluation targets, in which the viewing transformations are estimated from about 480 samples on four calibration targets, obtained through direct and image-based observations, respectively.

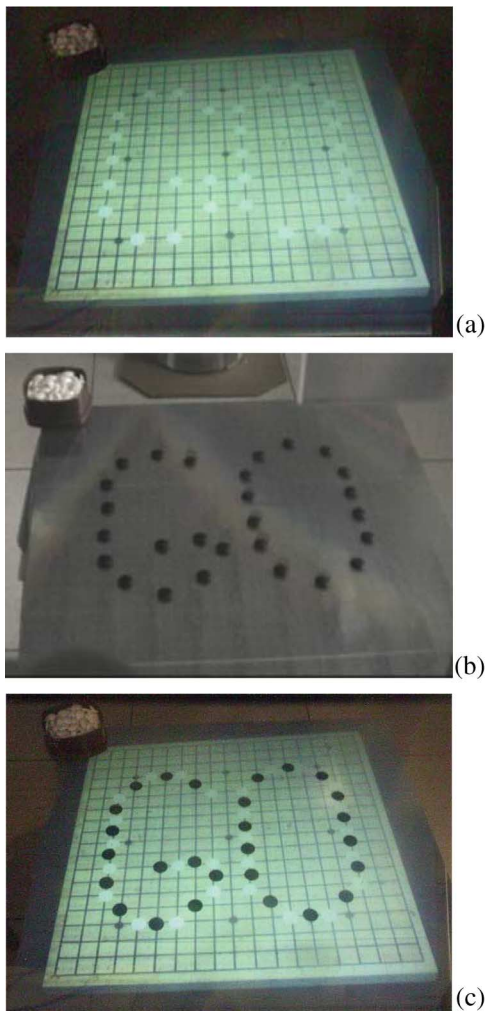

Fig. 9. Playing augmented GO game with a remote opponent. (a) HMPD player's virtual view. (b) HMPD player's direct real view. (c) HMPD player's augmented view.

"augmented GO game," has been developed to subjectively evaluate calibration methods and registration accuracy in augmented environments based on the HMPD technology [19]. In the augmented GO game simulation, a computer-generated 3-D GO board is projected onto a retro-reflective workbench through an HMPD. A local player, wearing the HMPD, perceives the virtual board as if it was a real object sitting on the tabletop and manipulates his real stone pieces on the virtual board. The locations of the pieces placed by a remote opponent are communicated to the local player via the collaborative server, and corresponding computer-generated pieces are overlaid with the virtual board. The challenges in the GO game were to ensure that the virtual board was aligned with the physical retro-reflective tabletop and physical stones placed on it and that the virtual board appeared in a fixed position and size in the real world space when viewed from arbitrary perspectives. Fig. 9(a) depicts the virtual components seen by the HMPD player, including a virtual board with the white stones placed by the remote player. Fig. 9(b) shows the HMPD player's direct view, with only his physical stones scattered on the screen. Fig. 9(c) shows the augmented view perceived through the HMPD: The virtual board, white virtual stones, black real stones, and miscellaneous elements of the physical environment are seamlessly integrated, with the black stones naturally occluding the occupied grids.

By implementing the calibration results as described in Section IV-C, we achieved much more accurate rendering than was the case without those calibration procedures. The perceived virtual GO board is fairly well aligned with real stones placed on the top of the retro-reflective screen at changing perspectives. Fig. 10(a) and (b) shows the registration viewed at two different perspectives. Results show that the HMPD and associated calibration methods enable registration of real and virtual objects within 5-mm rms error, computed with 27 stones across ten viewing perspectives. In the augmented view, the virtual board, white virtual stones, black real stones, and miscellaneous elements of the physical environments are seamlessly integrated, with the black stones naturally occluding the occupied grids.

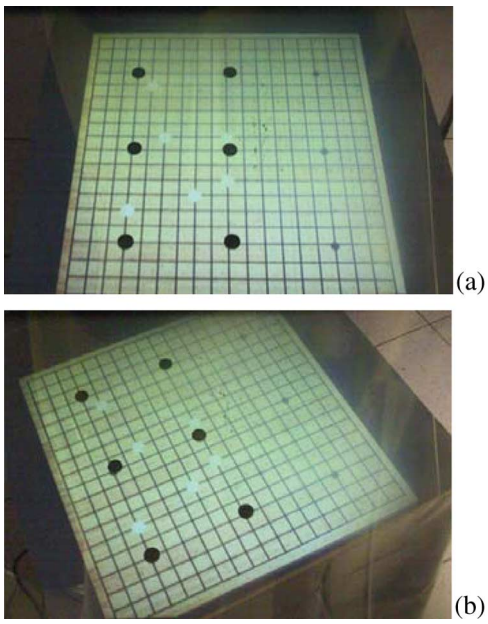

Fig. 10. Black physical stones properly register with the virtual elements (i.e., the board and white stones) in HMPD when a user was walking around the bench. (a) Registration in the front perspective. (b) Registration in a side perspective.

VII. Discussion

This paper presents a computational mechanism that accurately models the two-step projection process in an HMPD viewing system with an EVP system. The EVP system is an abstract single-step viewing device that takes into account practical misalignments in HMPD system and maintains compatibility with the camera model in most graphics libraries. To address the challenge of registration in a custom-designed HMPD-based AR system, this paper further presents a systematic calibration method that estimates both intrinsic and extrinsic parameters of the display system and thus establishes the viewing and imaging transformations for the EVP system utilized to render 2-D images from 3-D virtual environments. The procedures described in this paper ensure best estimation for the unknown fixed transformations and offer a registration accuracy that meets most of the application requirements. However, due to the offline nature of the described calibration method, an online calibration step may be necessary for demanding applications to refine the user-dependent varying parameters which are initially estimated from the offline procedures. For instance, the offline calibration method might yield a perfect registration only when a user's pupil position and viewing orientation are well aligned with the EVP system established from the offline calibration. Due to IPD variations for different users as well as the possibility of helmet slippage during an application session, the requirement for a perfect alignment is difficult to meet. Therefore, a fast and easy online refinement is required to fully compensate for all possible error sources and dynamic misregistration during an experiment session. In future work, we will develop a simplified online calibration method that takes advantage of the initial offline estimates to ensure that a minimum number of world-image correspondences are required to dynamically refine the initial estimates. We will also consider the possibility of developing a hybrid strategy for automatic online refinement.

Acknowledgment

The authors would like to thank 3M Inc. for supplying the retro-reflective films and L. D. Brown for his work on developing the testbed applications.

References

[1] Y. Argotti, L. Davis, V. Outters, and J. P. Rolland, "Dynamic superimposition of synthetic objects on rigid and simple-deformable real objects," in *Proc. IEEE Int. Symp. Augmented Reality*, New York, Oct. 29–30, 2001, pp. 5–10.

[2] R. Azuma and G. Bishop, "Improving static and dynamic registration in an optical see-through display," in *Proc. ACM SIGGRAPH (Computer Graphics)*, Jul. 1994, pp. 194–204.

[3] M. Bajura, H. Fuchs, and R. Ohbuchi, "Merging virtual objects with the real world: Seeing ultrasound imagery within the patient," in *Proc. ACM SIGGRAPH (Computer Graphics)*, Chicago, IL, Jul. 1992, pp. 203–210.

[4] M. Bajura and U. Neumann, "Dynamic registration correction in augmented-reality systems," in *Proc. IEEE VRAIS*, Mar. 1995, pp. 189–196.

[5] T. Caudell and D. Mizell, "Augmented reality: An application of heads-up display technology to manual manufacturing processes," in *Proc. Hawaii Int. Conf. Syst. Sci.*, Jan. 1992, pp. 659–669.

[6] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surround-screen projection-based virtual reality: The design and implementation of the CAVE," in *Proc. ACM SIGGRAPH (Computer Graphics)*, New York, Jul. 1993, pp. 135–142.

[7] L. Davis, J. Rolland *et al.*, "Enabling a continuum of virtual environment experiences," *IEEE Comput. Graph. Appl.*, vol. 23, no. 2, pp. 10–12, Mar./Apr. 2–4, 2003.

[8] M. Deering, "High resolution virtual reality," in *Proc. ACM SIGGRAPH (Computer Graphics)*, Jul. 1992, vol. 26, pp. 195–202.

[9] J. Fergason, "Optical system for head mounted display using retro-reflector and method of displaying an image," U.S. Patent 5 621 572, Apr. 15, 1997.

[10] S. Feiner, B. MacIntyre, and D. Seligmann, "Knowledge-based augmented reality," *Commun. ACM*, vol. 36, no. 7, pp. 53–62, Jul. 1993.

[11] R. Fisher, "Head-mounted projection display system featuring beam splitter and method of making same," U.S. Patent 5 572 229, Nov. 5, 1996.

[12] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics: Principles and Practice*. Reading, MA: Addison-Wesley, 1990.

[13] W. E. L. Grimson, G. J. Ettinger, S. J. White, T. Lozano-Perez, W. M. Wells, and R. Kikinis, "An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization," *IEEE Trans. Med. Imag.*, vol. 15, no. 2, pp. 129–140, Apr. 1996.

[14] R. L. Holloway, "Registration error analysis for augmented reality," *Presence: Teleoperators Virtual Environ. (MIT Press)*, vol. 6, no. 4, pp. 413–432, 1997.

[15] Y. Ha, H. Hua, R. Martins, and J. P. Rolland, "Design of a wearable wide-angle projection color display," in *Proc. IODC*, Tucson, AZ, Jun. 3–5, 2002, pp. 67–73.

[16] H. Hua, A. Girardot, C. Gao, and J. P. Rolland, "Engineering of head-mounted projective displays," *Appl. Opt.*, vol. 39, no. 22, pp. 3814–3824, Aug. 2000.

[17] H. Hua, C. Gao, F. Biocca, and J. P. Rolland, "An ultra-light and compact design and implementation of head-mounted projective displays," in *Proc. IEEE-VR*, Yokohama, Japan, Mar. 12–17, 2001, pp. 175–182.

[18] H. Hua, C. Gao, L. Brown, N. Ahuja, and J. P. Rolland, "Using a head-mounted projective display in interactive augmented environments," in *Proc. IEEE Int. Symp. Augmented Reality*, New York, Oct. 29–30, 2001, pp. 217–223.

[19] H. Hua, C. Gao, L. Brown, N. Ahuja, and J. P. Rolland, "A testbed for precise registration, natural occlusion and interaction in an augmented environment using a head-mounted projective display (HMPD)," in *Proc. IEEE-VR*, Orlando, FL, Mar. 22–28, 2002, pp. 81–89.

[20] H. Hua, C. Gao, and N. Ahuja, "Calibration of a head-mounted projective display for augmented reality systems," in *Proc. IEEE Int. Symp. Mixed and Augmented Reality*, Darmstadt, Germany, Sep. 30–Oct. 1, 2002, pp. 176–185.
[21] H. Hua, Y. Ha, and J. P. Rolland, "Design of an ultra-light and compact projection lens," *Appl. Opt.*, vol. 42, no. 1, pp. 97–107, Jan. 2003.
[22] H. Hua, L. Brown, and C. Gao, "SCAPE: Supporting stereoscopic collaboration in augmented and projective environments," *IEEE Comput. Graph. Appl.*, vol. 24, no. 1, pp. 66–75, Jan./Feb. 2004.
[23] H. Hua, L. Brown, and C. Gao, "System and interface framework for SCAPE as a collaborative infrastructure," *Presence: Teleoperators Virtual Environ.*, vol. 13, no. 2, pp. 234–250, Apr. 2004.
[24] M. Inami, N. Kawakami, D. Sekiguchi, Y. Yanagida, T. Maeda, and S. Tachi, "Visuo-haptic display using head-mounted projector," in *Proc. IEEE-VR*, Los Alamitos, CA, 2000, pp. 233–240.
[25] A. Janin, D. Mizell, and T. Caudell, "Calibration of head-mounted displays for augmented reality applications," in *Proc. VRAIS*, Sep. 1993, pp. 246–255.
[26] N. Kawakami, M. Inami, D. Sekiguchi, Y. Yangagida, T. Maeda, and S. Tachi, "Object-oriented displays: A new type of display systems—From immersive display to object-oriented displays," in *Proc. IEEE SMC*, Piscataway, NJ, 1999, vol. 5, pp. 1066–1069.
[27] R. Kijima and T. Ojika, "Transition between virtual environment and workstation environment with projective head-mounted display," in *Proc. IEEE Virtual Reality Annu. Int. Symp.*, Los Alamitos, CA, 1997, pp. 130–137.
[28] R. Kijima, K. Haza, Y. Tada, and T. Ojika, "Distributed display approach using PHMD with infrared camera," in *Proc. IEEE Virtual Reality Annu. Int. Symp.*, Los Alamitos, CA, Mar. 2002, pp. 33–40.
[29] W. Lorensen, H. Cline, C. Nafis, R. Kikinis, D. Altobelli, and L. Gleason, "Enhancing reality in the operating room," in *Proc. IEEE Vis.*, Los Alamitos, CA, Oct. 1993, pp. 410–415.
[30] E. McGarrity and M. Tuceryan, "A method for calibrating see-through head-mounted displays for AR," in *Proc. 2nd IEEE and ACM Int. Workshop Augmented Reality*, 1999, pp. 75–84.
[31] E. McGarrity, Y. Genc, M. Tuceryan, C. Owen, and N. Navab, "A new system for online quantitative evaluation of optical see-through augmentation," in *Proc. IEEE and ACM Int. Symp. Augmented Reality*, New York, Oct. 2001, pp. 157–166.
[32] P. Milgram, S. Zhai, D. Drascic, and J. J. Grodski, "Applications of augmented reality for human–robot communication," in *Proc. Int. Conf. Intell. Robots Syst.*, Yokohama, Japan, Jul. 1993, pp. 1467–1472.
[33] T. Oishi and S. Tachi, "Methods to calibrate projection transformation parameters for see-through head-mounted displays," *Presence: Teleoperators Virtual Environ. (MIT Press)*, vol. 5, no. 1, pp. 122–135, 1996.
[34] J. Parsons and J. P. Rolland, "A non-intrusive display technique for providing real-time data within a surgeons critical area of interest," in *Proc. Med. Meets Virtual Reality*, 1998, pp. 246–251.
[35] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computation Second Edition*. New York: Cambridge Univ. Press, 1999.
[36] J. Rolland and H. Fuchs, "Optical versus video see-through head-mounted displays," in *Fundamentals of Wearable Computers and Augmented Reality*, W. Barfield and T. P. Caudell, Eds. Mahwah, NJ: Lawrence Erlbaum Assoc., Pub., 2001, pp. 113–156.
[37] W. Robinett and J. Rolland, "A computational model for stereoscopic optics of a head-mounted display," *Presence: Teleoperators Virtual Environ. (MIT Press)*, vol. 1, no. 1, pp. 45–62, 1992.
[38] W. Robinett and R. Holloway, "The visual display transformation for virtual reality," *Presence: Teleoperators Virtual Environ. (MIT Press)*, vol. 4, no. 1, pp. 1–23, 1995.
[39] J. Rolland, L. Davis *et al.*, "3D visualization and imaging in distributed collaborative environments," *IEEE Comput. Graph. Appl.*, vol. 22, no. 1, pp. 11–13, Jan./Feb. 2002.
[40] J. P. Rolland and H. Hua, "Head-mounted display systems," in *Encyclopedia of Optical Engineering*, R. B. Johnson and R. G. Driggers, Eds. New York: Marcel Dekker, 2005, pp. 1–13.
[41] D. A. Southard, "Transformations for stereoscopic visual simulation," *Comput. Graph.*, vol. 16, no. 4, pp. 401–410, 1992.
[42] A. State, M. Livingston, W. Garret, G. Hirota, M. Whitton, E. Pisano, and H. Fuchs, "Technologies for augmented reality systems: Realizing ultrasound-guided need biopsies," in *Proc. ACM SIGGRAPH*, New Orleans, LA, Aug. 1996, pp. 439–446.
[43] A. Tang, J. Zhou, and C. Owen, "Evaluation of calibration procedures for optical see-through head-mounted displays," in *Proc. 2nd IEEE and ACM ISMAR*, 2003, pp. 161–168.
[44] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE J. Robot. Autom.*, vol. RA-3, no. 4, pp. 323–344, Aug. 1987.
[45] M. Tuceryan, D. S. Greer, R. T. Whitaker, D. E. Breen, C. Crampton, E. Rose, and K. H. Ahlers, "Calibration requirements and procedures for a monitor-based augmented reality system," *IEEE Trans. Vis. Comput. Graphics*, vol. 1, no. 3, pp. 255–273, Sep. 1995.
[46] M. Tuceryan and N. Navab, "Single point active alignment method (SPAAM) for optical see-through HMD calibration for AR," in *Proc. Int. Symp. Augmented Reality*, Munich, Germany, Oct. 2000, pp. 149–158.

Hong Hua (M'00) received the B.S.E. (with honors) and Ph.D. degrees (with distinction) all in optical engineering from the Beijing Institute of Technology, Beijing, China, in 1994 and 1999, respectively.

She was a Postdoctoral Research Associate with the School of Optics, University of Central Florida; a Beckman Fellow with the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, from 1999 to 2002; and an Assistant Professor with the Department of Information and Computer Sciences, University of Hawaii, in 2003. She has been with the University of Arizona, Tuczon, since 2003, where she is currently an Assistant Professor with the College of Optical Sciences. Her current research interests mainly include stereoscopic displays, optical engineering, virtual and augmented reality, and 3-D human computer interaction.

Chunyu Gao is currently working toward the Ph.D. degree at the Department of Electrical and Computer Engineering and the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana.

His research interests include computer vision, virtual and augmented reality, camera and imaging, and stereoscopic displays.

Narendra Ahuja (S'97–M'79–SM'85–F'92) received the B.E. degree (with honors) in electronics engineering from Birla Institute of Technology and Science, Pilani, India, in 1972, the M.E. degree (with distinction) in electrical communication engineering from Indian Institute of Science, Bangalore, India, in 1974, and the Ph.D. degree in computer science from University of Maryland, College Park, in 1979.

Since 1979, he has been with the University of Illinois at Urbana-Champaign, Urbana, where he is currently a Donald Biggar Willet Professor with the Department of Electrical and Computer Engineering, the Beckman Institute, and the Coordinated Science Laboratory. His current research emphasizes the integrated use of multiple image sources of scene information to construct 3-D and other descriptions of scenes, the use of integrated image analysis for realistic image synthesis, sensors for computer vision, extraction and representation of spatial structure, and the use of the results of image analysis for a variety of applications, including visual communication, image manipulation, robotics, and scene navigation.

Dr. Ahuja received the 1998 Technology Achievement Award of the International Society for Optical Engineering and the 1999 Emanuel R. Piore award of the IEEE.