Calibration of a Head-Mounted Projective Display for Augmented Reality Systems

Hong Hua, Chunyu Gao, and Narendra Ahuja
Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana-Champaign, Urbana, IL 61801
Email: Honghua@uiuc.edu

Abstract

In augmented reality (AR) application, registering a virtual object with its real counterpart accurately and comfortably is one of the basic and challenging issues in the sense that the size, depth, geometry, as well as physical attributes of the virtual objects have to be rendered precisely relative to a physical reference, which is well-known as the calibration or registration problem. This paper presents a systematic calibration process to address static registration issue in a custom-designed augmented reality system, which is based upon the recent advancement of head-mounted projective display (HMPD) technology. Following a concise review of the HMPD concept and system configuration, we present in detail a computational model for the system calibration, describe the calibration procedures to obtain the estimations of the unknown transformations, and include the calibration results, evaluation experiments and results.

1. 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, 20], medical training [1], repair and maintenance of complex facilities [9, 6], and tele-manipulation [23].

Physical and computer-generated entities coexist in an AR environment. One of the basic challenges is the calibration or registration problem, which requires that the size, depth, geometry, as well as physical attributes of virtual objects have to be rendered accurately with respect to their physical counterparts to create the illusion of co-existence. Calibration is a significant aspect of research in AR domain, as well as in other related fields, including computer graphics, computer vision and robotics. Deering presented the general steps that must be taken to produce accurate high-resolution head-tracked stereo display in order to achieve sub-centimeter virtual to physical registration [7]. Janin et al. described the experimental procedures to determine the calibration parameters of a see-through HMD by both manual measurement and optimization in [17]. Azuma and Bishop described experimental steps to estimate viewing parameters and predictive tracking techniques to improve both static and dynamic registrations in an optical see-through HMD system [2]. Bajura and Neumann presented the idea of dynamically measuring registration errors in synthetic images and using that information to correct 3D registration errors in a video-based AR system [4]. Tuceryan et al. described a systematic view of the calibration issues arising in a monitor-based augmented reality system [30]. Grimson et al. explored vision techniques to automate the process of registering medical data to a patient's head in a video-based system [11]. Oishi and Tachi proposed a calibration method to minimize systematic errors of projection transformation parameters for optical see-through HMDs [24]. McGarrity and Tuceryan described a camera-based online calibration method for optical see-through display [21, 31].

The subject of this paper is to present a systematic calibration method to address static registration issue in a custom-designed augmented reality system, which is based upon recent advancement of head-mounted projective display (HMPD) technology [8, 19, 12]. The proposed method can be adapted by other AR systems using a different display technology with minor modifications.

The rest of the paper is organized as follows: In section II, we first briefly review the concept and recent advancement of the HMPD technology, and concisely overview the system configuration. Section III presents a computational model for the system calibration, which gives the transformations needed to link the physical viewing system with the virtual cameras. In Section IV, we then present the calibration procedures that accomplish the calibration steps to obtain the estimation of the transformations. Finally, we present a series of experiments to evaluate the calibration results in Sec. V.

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)
0-7695-1781-1/02 $17.00 © 2002 IEEE

2. System configuration

Video and optical see-though HMDs have been the two basic approaches to combining real and virtual images [26]. In both approaches, the viewing optics typically is eyepiece-type or microscope-type compound magnifier. The HMPD is an emerging technology that can be thought to lie on the boundary of conventional HMDs and projective displays such as the CAVE systems [5]. It has been recently demonstrated to yield 3D visualization capabilities with a large-FOV, lightweight and low distortion optics, and correct occlusion of virtual objects by real objects [8, 12, 13, 16, 19]. Thus, the HMPD technology was identified as an alternative to optical see-though HMDs for augmented applications [14, 15, 18, 27].

An HMPD consists of a pair of miniature projection lenses, beam splitters, and displays mounted on the head and a supple retro-reflective screen placed strategically in the environment [8, 19, 25]. Its monocular configuration is illustrated in Fig. 1. Rays from 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, are projected through the lens and retro-reflected back to the exit pupil, where a user's eye is positioned to observe the magnified virtual image. Because of the retro-reflective property, in which the rays hitting the surface are reflected back on themselves in the opposite directions, the location and size of the perceived image are independent of the location and shape of the retro-reflective screen [12].

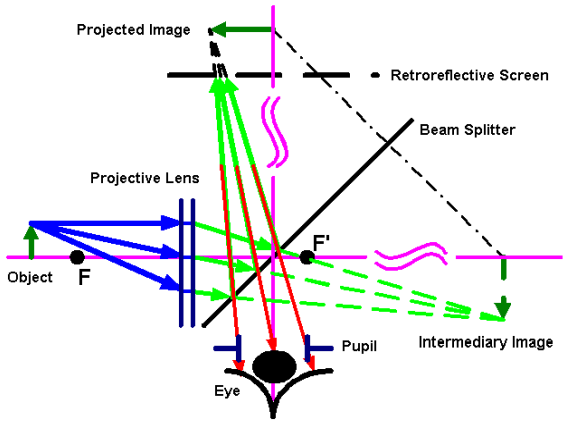

Fig.1 Imaging concept of HMPD

Our experimental system uses an HMPD prototype designed by Hua and her colleagues [13, 14, 15]. The prototype achieves 52-degree FOV and weighs about 750 grams, with a 640x480 VGA resolution. The HMPD is driven by a SGI Octane workstation with one R12000 processor and one four-channel VGA output option. The head motion is detected by a 6DOF Hiball3000 optical tracking system. Figure 2 shows the front view of the prototype with a Hiball3000 sensor attached.

Fig. 2 HMPD prototype

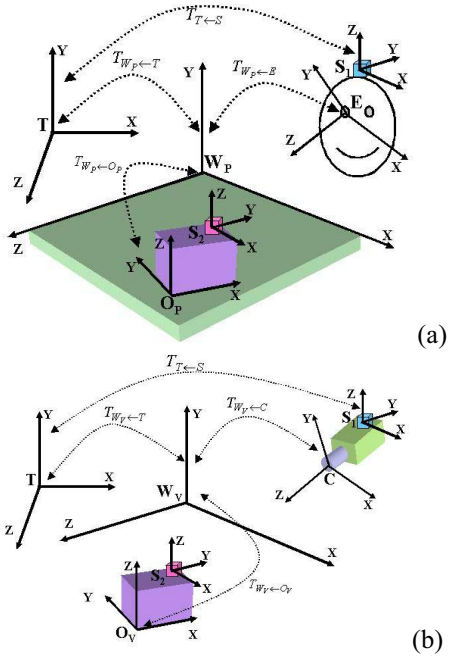

Fig. 3 Illustration of the physical and virtual world components and transformations: (a) Physical world components; (b) Virtual world components.

3. Computational model for system calibration

The calibration process is to obtain the optimal estimations of a set of transformations, which map a virtual world representation with its physical counterpart. Figures 3-a and 3-b illustrate the main components, the associated coordinate systems, and the transformations of the testbed system in the physical and virtual worlds, respectively. The key references are the physical world coordinates (PWC), W_PXYZ, and the virtual world coordinates (VWC), W_VXYZ. The PWC is fixed at a known location in the physical environment and all other physical components are specified in the unique physical framework by a set of rigid transformations. The VWC is fixed in the virtual world and all other virtual elements are determined in the unique virtual framework. The key transformations fall into two categories: viewing transformations and object transformations. The former is a set of extrinsic and intrinsic transformations, which are utilized to precisely place a virtual camera in the VWC and to generate accurate 2D mapping of the 3D virtual world. The latter is a set of extrinsic rigid transformations, which are utilized to place virtual objects in the virtual world. This paper shall focus on how to obtain the viewing

transformations. The calibration of object transformations is application-oriented, and is not included in the paper.

Illustrated in Fig 4, viewing transformations include the extrinsic viewing orientation transformation, M_{ext}, and the intrinsic viewing projection transformation, M_{int}. The former specifies how to place a virtual camera in the world space and transform a 3D model representation in the VWC to that in the camera viewing reference coordinates (VRC), *CXYZ*. The latter specifies how to generate the mapping from a 3D point in the VRC to its 2D image representation in the viewport space.

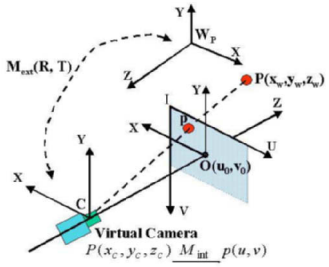

Fig. 4 Illustration of the extrinsic and intrinsic viewing transformations

3.1. Viewing orientation transformation

The extrinsic viewing orientation transformation is defined by $M_{ext} = T_{C\leftarrow W} = [R \quad T]$, or $\overrightarrow{P_C} = R \bullet \overrightarrow{P_W} + T$ where $R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$ represents the composite rotations, $T = [t_x \quad t_y \quad t_z]^T$ represents the translations, $\overrightarrow{P_C} = [x_c \quad y_c \quad z_c]^T$ and $\overrightarrow{P_W} = [x_w \quad y_w \quad z_w]^T$ are the vector representations of a point P defined in the camera and world coordinates, respectively. The viewing orientation transformation can be further decomposed as:

$$M_{ext} = T_{C\leftarrow W} = T_{E\leftarrow W} = T_{E\leftarrow S} T_{S\leftarrow T} T_{T\leftarrow W} \qquad (1)$$

where $T_{T\leftarrow W}$, referred as tracker transformation, gives the position and orientation of the tracker transmitter in the world space, $T_{S\leftarrow T}$, referred as sensor transformation, is the position and orientation of the moving sensor in the transmitter space, and $T_{E\leftarrow S}$, referred as eye transformation, specifies the eyepoint or camera position and viewing direction relative to the sensor coordinates.

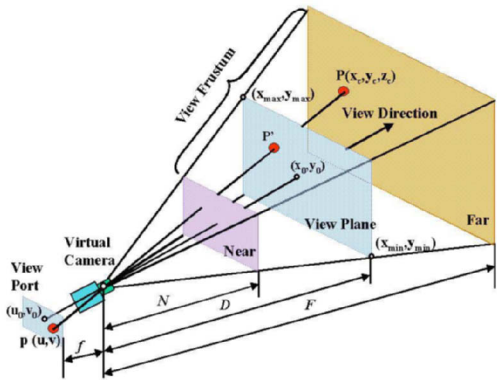

Fig. 5 Illustration of the HMPD calibration method

The sensor transformation is measured explicitly by the motion tracker, and dynamically changes when a user moves around. Both the tracker and eye transformations are determined by a calibration procedure. The 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 (same user) due to the potential slippage of the helmet. In this paper, we assume the eye transformation remains fixed in the context of static registration. Furthermore, the left and right viewing optics are not necessarily parallel. For example, in some HMD designs, the left/right optics are purposely diverged or converged to achieve larger overall FOV. Therefore, the viewing orientation transformations corresponding to the left/right virtual cameras have to be calibrated separately.

3.2. Viewing projection transformation

The viewing projection transformation specifies how a virtual camera captures a 3D scene within its view frustum and maps it onto a 2D representation in the viewport. A generic model of the virtual camera used by most of the graphics package such as OpenGL is illustrated in Fig. 5 [10]. In particular, the virtual camera parameters fall into three categories of operation: projection, warping and clipping. The projection parameters specify a projection transformation that maps a 3D scene point in the VRC space onto its 2D representation in the viewport. 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 by the virtual camera. Among them, only the projection and warping parameters affect the 3D-to-2D mapping. They together form the intrinsic viewing projection transformation, M_{int}. The clipping parameters, however, determine which scene point is captured by a virtual camera, and do not affect the mapping. The clipping parameters will be discussed in the next sub-section.

We denote the effective focal length (EFL) of a virtual camera as f (mm), the viewport reference point as (u_0, v_0) (pixels), and the pixel scale factors as (S_u, S_v) (mm/pixel) (Fig. 5). Given a 3D point in the VRC space, $\overrightarrow{P_c} = [x_c \quad y_c \quad z_c]^T$, its 2D projection, $p(x_1, x_2, x_3)$, is given by:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -f/S_u & 0 & u_0 \\ 0 & -f/S_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \qquad (2)$$

The pixel coordinates (u, v) of the projection in the viewport or device coordinates is given by:

$$\begin{cases} u = x_1/x_3 = \dfrac{-f * x_c}{S_u * z_c} + u_0 \\ v = x_2/x_3 = \dfrac{-f * y_c}{S_v * z_c} + v_0 \end{cases} \quad (3)$$

The lens distortion can be modeled rather accurately as simple radial distortions:

$$\begin{cases} x = x_d(1 + k_1 * r^2) \\ y = y_d(1 + k_1 * r^2) \end{cases} \quad (4)$$

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

3.3. Clipping parameters

A view frustum is defined by a projection plane and a viewing window in the plane, a far clipping plane, and a near clipping plane (Fig. 5). The projection plane distance d(mm), the viewing window dimensions (W, H) (mm), and view window reference point (x_0, y_0) (mm) are conjugate to those of the view port parameters. They specify the FOV of the virtual camera, which corresponds to visual angle of the HMPD virtual image subtended to the eyepoint and may vary each time when the display configuration is changed (e.g. adjusting image distance or IPD). They can be computed from the intrinsic parameters obtained along with the display calibration. The far and near clipping planes of the view frustum, given by (F, N) (mm), are related to the accommodation capability of the human visual system.

In nature, humans have become accustomed to focusing and converging their eyes in a coordinated fashion: changing the distance from an observed object leads to not only changing the convergence, but also changing lens power. Stereoscopic displays, however, require the eyes be accommodated on the image plane, while they converge on a point of interest whose apparent depth may be different from the image plane. Valyus reports that most people can comfortably tolerate a change in convergence angle of up to $\pm 1.6^\circ$ [28, 32], or recommends a change in convergence distance of up to 0.75 diopters. Consequently, the general guideline by Valyus on parallax is:

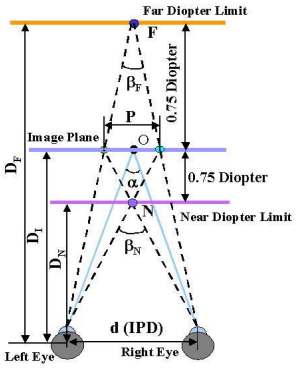

Fig. 6 Illustration of the HMPD calibration method

$$|P| \le 0.03 D_I \quad (5)$$

Where P is the parallax of a stereo-pair point measured on the image plane, and D_I is the virtual image distance. This parallax restriction places a limit on the depth range for a given viewing distance. Excessive parallax leads to eyestrain or failure of viewability, which is referred as the accommodation/convergence conflict.

Setting the near and far clipping planes of the viewing frustum to the limits specified by the convergence tolerance can possibly eliminate the possibility of excessive parallax. As illustrated in Figure 6, we denote the image plane distance as D_I, the IPD as d, and the change in convergence angle as θ, and then the distance of the point of interest is given by:

$$D_{F|N} = \frac{d}{2\tan(\arctan(\frac{d}{2D_I}) - \frac{\theta}{2})} \quad (6)$$

where $|\theta = \alpha - \beta_{F|N}| \le 1.6^\circ$, in which α is the convergence angle of the point A with zero parallax and $\alpha = 2\arctan(\frac{d}{2D_I})$, and β is the convergence angle of the point of interest, and $\beta_{F|N} = 2\arctan(\frac{d}{2D_{F|N}})$.

When we set θ to be the maximum and minimum convergence angles ($\theta = +1.6^\circ$ or -1.6°), equation (6) gives the recommended far and near clipping plane distances. For example, by using the $\pm 1.6^\circ$ change of convergence angle, and 65mm of IPD, Southward recommended a viewing distance of at least 2.3m for viewing distant objects stereoscopically, which provides a comfortable stereoscopic depth range from about 1.2m to infinity [28]. Decreasing the viewing distance can bring the comfortable viewing range closer to the viewer in arm-length applications, but this implies the viewer can tolerate a much smaller comfortable viewing range.

4. Calibration procedures and method

Assuming $f_u = f/S_u$ and $f_v = f/S_v$, we obtain

$$\begin{cases} u - u_0 = \dfrac{-f_u(r_{11}x_w + r_{12}y_w + r_{13}z_w + t_x)}{r_{31}x_w + r_{32}y_w + r_{33}z_w + t_z} \\ v - v_0 = \dfrac{-f_v(r_{21}x_w + r_{22}y_w + r_{23}z_w + t_y)}{r_{31}x_w + r_{32}y_w + r_{33}z_w + t_z} \end{cases} \quad (7)$$

This equation explicitly links a 3D point in the world space $\overrightarrow{P_W} = [x_w \quad y_w \quad z_w]^T$ to a pixel (u, v) in the display coordinates. Given sufficient number of world-image correspondences, through the well-known numerical fitting methods, we shall obtain the estimation of the extrinsic viewing orientation transformation (R, T) and projection parameters (f, S_u, S_v, u_0, v_0, k_1), which is referred as the display calibration.

The calibration steps to obtain a complete estimation of the viewing transformations include: 1) Depth range

calibration; 2) Tracker transformation calibration; and 3) Display calibration.

As discussed in section 3.3, the average viewing distance and depth range of a particular application scenario has to be determined empirically to avoid excessive parallax. Therefore, the first step determines the proper depth range using equation (6), and adjusts the display hardware to comply with specific conditions, which should be done in the first place. We will omit the details on how we make adjustment in our prototype. The tracker transformation only needs to be calibrated once and the result will be used for display calibration. After making all the display adjustment, display calibration follows.

4.1. Tracker transformation calibration

While the head-tracker measurement gives the position/orientation of the sensor with respect to the tracker transmitter space, $T_{T\leftarrow S}$, it is necessary to calibrate the position/orientation of the transmitter coordinates in world space, $T_{W\leftarrow T}$. If the tracker manufacturer has specified the transmitter coordinates with sufficient accuracy, it is possible to directly measure the offset from the given world coordinates and compute $T_{W\leftarrow T}$. Typically, the actual origin and orientation of the transmitter coordinates are not accurately aligned with those specified by the manufacturer, and therefore it is not possible to measure them directly. For example, our head-tracker is the HiBall 3000 from 3rd Tech. It is unclear where the actual origin and axis are. The following method is used to compute the $T_{W\leftarrow T}$. In a designated world coordinate system, when aligning the tip of the stylus with the origin, a selected point along the X axis, and a point along Y axis of the world coordinate system, we recorded the position measurements of the sensor, $\vec{W}$, $\vec{X}$, and $\vec{Y}$, respectively. The normalized vectors of the X, Y, and Z axes in the tracker coordinates are given by

$$\vec{X'} = \frac{\vec{X} - \vec{W}}{\left\| \vec{X} - \vec{W} \right\|}, \quad \vec{Y'} = \frac{\vec{Y} - \vec{W}}{\left\| \vec{Y} - \vec{W} \right\|}, \quad \vec{Z'} = \frac{\vec{X'} \times \vec{Y'}}{\left\| \vec{X'} \times \vec{Y'} \right\|}$$

Therefore,

$$T_{T\leftarrow W} = \begin{bmatrix} \vec{X'} & \vec{Y'} & \vec{Z'} & W \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

To achieve good results, the separation of X and Y from W should be large enough. To improve single-point measurement accuracy of the HiBall stylus, at each position, a large number of samples (e.g. 50) were taken at multiple different orientations and their average is used to compute the normalized vectors. We also took measurements at multiple positions along the axis and averaging the results of normalized vectors. Furthermore, this method does not ensure that $T_{W\leftarrow T}$ is a valid homogeneous transformation matrix and a post-processing step is necessary to make the rotation components orthogonal. The actual transformation measured in our system is:

$$T_{W_P\leftarrow T} = \begin{bmatrix} 0.9998 & 0.0175 & -0.0140 & -1.83 \\ -0.0176 & 0.9998 & -0.0085 & 0 \\ 0.0138 & 0.0087 & 0.9998 & 1.45 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Which shows about 0.5, 0.8, and 1 degrees twist around the XYZ axis as well as the offsets.

4.2. Display calibration

The computational model for the display calibration described in the section III is similar to the model for camera calibration used in computer vision [29], thus similar fitting algorithm can be used and we will omit the derivation when they are the same. However, the key difference of the display calibration from camera calibration lies in the fact that a camera involves a single and direct mapping process from a 3D scene to its 2D image while a HMD or HMPD display is not. The display involves two mapping processes, from the 2D display device to the 2D virtual image plane and from the image plane to a 3D world, and an observer (e.g. eye) is needed to determine world-image correspondences, which severely challenges the accuracy of the correspondence matching. Furthermore, in HMPDs, the necessity for a retro-reflective screen to view a virtual image requires a special-made calibration target. Consequently, a physical target should either be printed on a transparent sheet or directly on the retro-reflective screen. This paper describes a manual correspondence match (MCM) strategy to locate a large number of world-image correspondences.

Figure 7 illustrates the setup used for the MCM calibration process. The calibration target is an MxN grid pattern precisely drawn on a flat retro-reflective screen. The calibration task is to move a cross stimuli displayed through the HMPD, align it with each of the grid

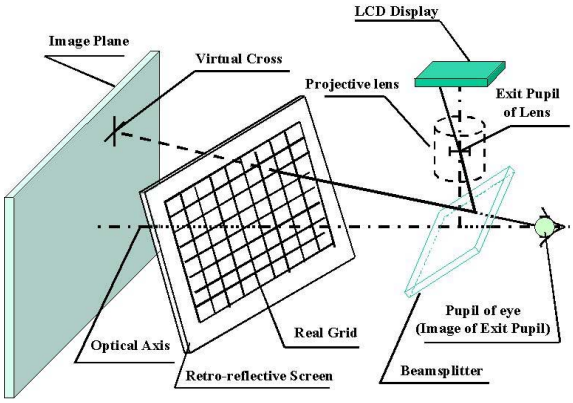

Fig. 7 Illustration of the HMPD calibration method

intersections, and record the corresponding stimuli pixel coordinates. The operator can either directly observe and make the alignment by properly positioning his/her eye at the exit pupil of the HMPD, or make the observation task through a calibration camera that is properly positioned at the exit pupil.

The calibration procedure consists of the following steps:

1) Mount the HMPD on a fixed platform, record the measurement of the head sensor $(\vec{S}, \overrightarrow{\Theta_S})$, and obtain the corresponding sensor transformation $T_{S\leftarrow T}$. Empirically, when only one sensor is used for both headtracking and stylus, this measurement should be taken at the end of the experiments. Whenever the sensor is re-installed on the helmet, a full calibration procedure is needed;
2) Define a calibration world coordinates $W_C XYZ$, which can be different from the application world coordinates, and obtain the tracker transformation $T_{W_C\leftarrow T}$;
3) Position the calibration target at a fixed attitude, record the MxN grid coordinates of the target in the tracker coordinates using the hiball stylus, $\left[P_{i,j}^k\right]_T$, where k represents the kth target attitude, i and j represent the row and column index of the grid pattern, respectively, and compute the corresponding world coordinates ($\left[P_{i,j}^k\right]_{W_C}$) from the transformation $T_{W_C\leftarrow T}$;
4) Present a cross stimuli on the display under calibration. The operator moves the virtual cross using a mouse to align it with each of the grid intersections on the target, and clicks the left button of the mouse to record the corresponding pixel coordinates, $Q_{i,j}^k(u,v)$;
5) Change the target attitude, and repeat the steps 3 and 4 at least 3 times ($k \geq 3$). Empirically, the angular difference between two target attitudes is recommended to large enough to ensure good convergence;
6) Form the calibration matrix (Equations 1, 2, 3, 7) from the P-Q correspondences, and apply the fitting methods [29] to solve the intrinsic and extrinsic parameters, which gives the eyepoint position and orientation $T_{E\leftarrow W_C}$, the focal length (*f*), aspect ratio (S_u,S_v), view window offsets (u_0,v_0), and distortion coefficients (k_1);
7) Insert $T_{S\leftarrow T}$ from step 1, $T_{W_C\leftarrow T}$ from step 2 and $T_{E\leftarrow W_C}$ from step 7 to Equation (1) and obtain the eye transformation $T_{S\leftarrow E}$;
8) Insert the intrinsic parameters to equation (2) and obtain the viewing projection transformation M_{int}.

We apply this method to the left and right arms of the HMPD separately, and their intrinsic and extrinsic parameters are listed in Table I. The eye transformations for the left and right arms are:

$$[T_{S\leftarrow E}]_{Left} = \begin{bmatrix} 0.9996 & 0.0044 & 0.0264 & -36.1664 \\ -0.0049 & 0.99981 & 0.0208 & -147.2725 \\ -0.0263 & -0.0209 & 0.9994 & -21.4114 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$[T_{S\leftarrow E}]_{Right} = \begin{bmatrix} 0.9999 & 0.0102 & 0.0069 & 32.4025 \\ -0.0097 & 0.9981 & -0.0616 & -135.3587 \\ -0.0075 & 0.0615 & 0.9981 & -7.7901 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Table I Intrinsic and extrinsic parameters of the HMPD

Parameters	Left	Right
Focal length (mm)	35.1473	35.09793
FOV (degrees)	41.17(H) 31.76(V)	41.22 (H) 31.81(V)
Distortion coefficient (1/mm^2)	8.134e-005	1.310e-004
Pupil position relative to calibration pattern (mm)	X=-178.716 Y=-171.653 Z=773.7375	X=-214.9796 Y=-105.5128 Z=688.77537
Orientation of the optical axis (degree)	Yaw=143.35 Pitch=-4.73 Roll=90.37	Yaw=142.299 Pitch=0.0061 Roll=90.1573
Display offsets(pixel)	H=77 V=-55	H=14 V=21

5. Evaluation experiments and results

In this section, we will describe the calibration setup specification and discuss a set of experiments and methods to evaluate how the number of correspondence samples and attitudes of calibration target affect the calibration accuracy.

5.1. Experiment setup specification

In the experiments, the calibration target is a 14x13 grid pattern with a pitch of 40mm. The grid line is 2mm wide, which corresponds to an angle of 7 arc minutes in the eye space. The stimulus presented through the HMPD is a cross with 1-pixle line-thickness when making direct observation or with 2-pixle line-thickness when observing through a video camera. 1-pixle line in the display space corresponds to 1.1mm on the virtual image plane of 1m away, which corresponds to an angle of 3.5 arc minutes in the eye space. The correspondence matching for each eye is conducted at up to 10 different target attitudes, through both direct observation and camera observation approaches.

5.2. Evaluation method

A particular problem associated with AR system calibration is how to evaluate the accuracy. Generally,

there have been two approaches to accuracy verification for optical see-through displays. The first approach is to have a user to subjectively report the qualitative accuracy of the registration of a computer-generated object with its physical counterpart. The second approach replaces human eye with a camera and conduct image-based measurements to assess registration accuracy [22]. In order to study the relationship of the display calibration accuracy with different configuration parameters such as number of samples and variation of target attitudes, using the same display platform as the one we conduct the calibration, the 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 evaluation target and task are the same as those used in the calibration procedure, but direct observation approach is always chosen for evaluation. Furthermore, the same sample of data is used for all evaluation purposes.

The evaluation algorithm includes the following steps:

1) Position the evaluation target at a fixed attitude, record its grid coordinates in the tracker coordinates using the hiball stylus, $\left[P_{i,j}^{k}\right]_{T}$, and compute their coordinates in the world coordinates $\left[P_{i,j}^{k}\right]_{W}$ using the transformation $T_{W\leftarrow T}$ obtained in tracker calibration;
2) Compute the projection of each sampled point in the display space, $\hat{Q}_{i,j}^{k}(u,v)$, from the calibrated intrinsic and extrinsic viewing transformations ($\hat{M}_{\text{int}}(\hat{f},\hat{S}_u,\hat{S}_v,\hat{u}_0,\hat{v}_0,\hat{k}_1)$, and $\hat{M}_{ext}(\hat{R},\hat{T})$);
3) Ask the operator to align a virtual cross with each of the grid intersections on the target, and record the corresponding pixel coordinates, $Q_{i,j}^{k}(u,v)$;
4) Compute the difference between the computed and the true projections $e_{i,j}^{k}(u,v)=\left|Q_{i,j}^{k}(u,v)-\hat{Q}_{i,j}^{k}(u,v)\right|$;
5) Change the target attitude, and repeat the steps 1 through 4;
6) Evaluate the means and STDs of all the sample points at each attitude and at all different attitudes.

5.3. Calibration accuracy analysis

The accuracy of the display calibration and convergence relies on the accuracy of the world coordinate measurements, the accuracy of the correspondence matching, and the number and distribution of samples. In our experiments, the accuracy of the world coordinate measurements relies on that of the Hiball stylus. Due to the highly reflective environments in our system, the stylus has limited accuracy, between 2mm and 5mm, which corresponds to approximately 2- and 5-pixel error in the display space. The stylus measurements are lack of

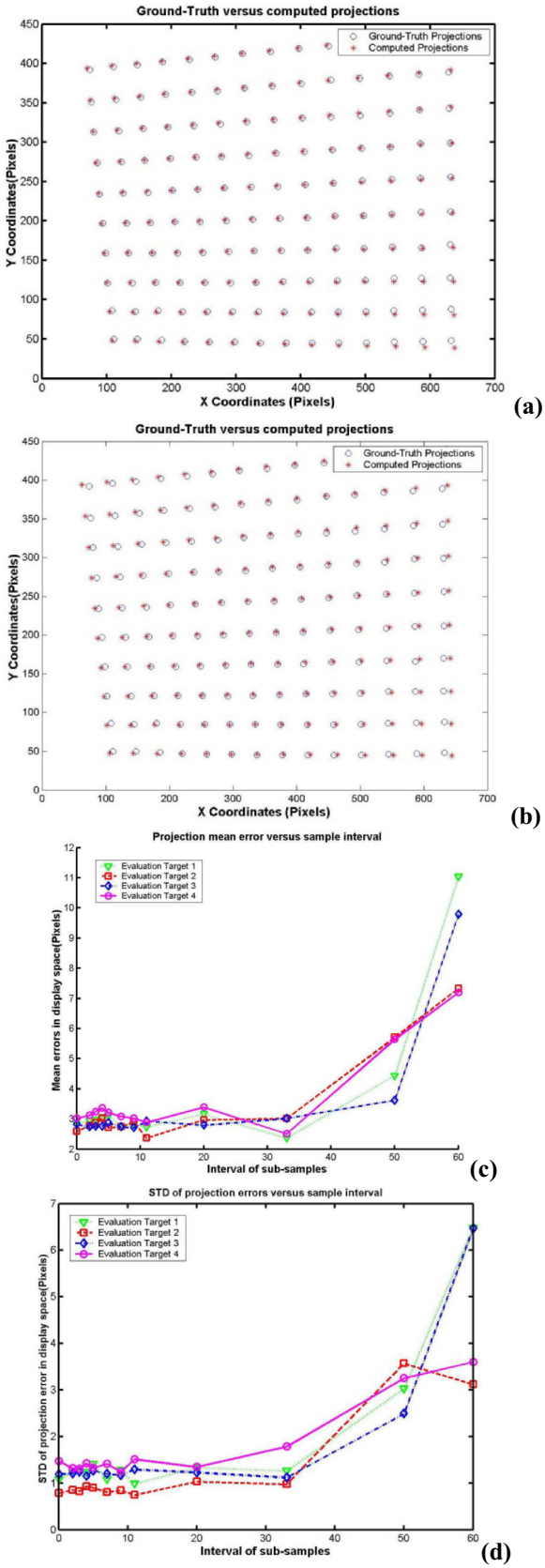

Fig. 8 Number of samples affects registration accuracy and convergence

consistency at different sensor orientations, therefore, instead of measuring world coordinates for every grid directly, we utilize the calibration approach described the Section 4.1 to obtain a target-tracker transformation, and then compute the grid coordinates in tracker reference based on their local measurements. Experiments show that this approach can improve the accuracy and signal-noise-ratio of the raw data.

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 about 1mm, which corresponds to less than 1-pixel resolution in the display space, or 3.5 arc minutes in the eye space. Assuming the virtual image is about 1m away and the display pixel is 40um, the angular resolution of the virtual image is about 3.5 arc minutes in the eye space. Therefore, the accuracy of the target matches with the resolution of the display. The matching accuracy would mainly rely on the observer, while it is limited to the level of the display resolution.

Therefore, the raw data of world coordinate measurements (P) and correspondence matching (Q) are much noisy and less accurate than the sub-pixel accuracy that can be typically achieved in camera calibration process. Consequently, we anticipate large number of correspondence samples is necessary to achieve usable accuracy and good convergence. Thus the number of samples is an important parameter to study.

5.4. Experiments and observations

5.4.1. Number of samples

Manually determining large number of correspondences is a difficult and time-consuming task. Minimizing the number of samples while preserving reasonable accuracy and convergence is helpful. Furthermore, studying the relationship of sample number and calibration accuracy provides an informative guidance for experiment designs. In the calibration experiments, using image-based observation, we sampled 12x10 points at each target attitudes, repeated the task at 9 different target locations, obtained total 1080 correspondences. We ensured that the 9 targets are different in both position and orientation. Selecting different sample intervals, we formed 12 groups of data by down-sampling the number of points at each attitude. The sample intervals are: 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, 3and 2 samples at each target attitude, respectively. Then the extrinsic and intrinsic transformations corresponding to each group of samples were computed, and were denoted as $(\hat{M}_{ext}, \hat{M}_{\text{int}})_{i=1..12}$.

In the evaluation experiments, correspondence matching was done by direct observation at the exit pupil. Measurements are taken at 4 different target attitudes. At each orientation, 14x10 points were sampled. Based on

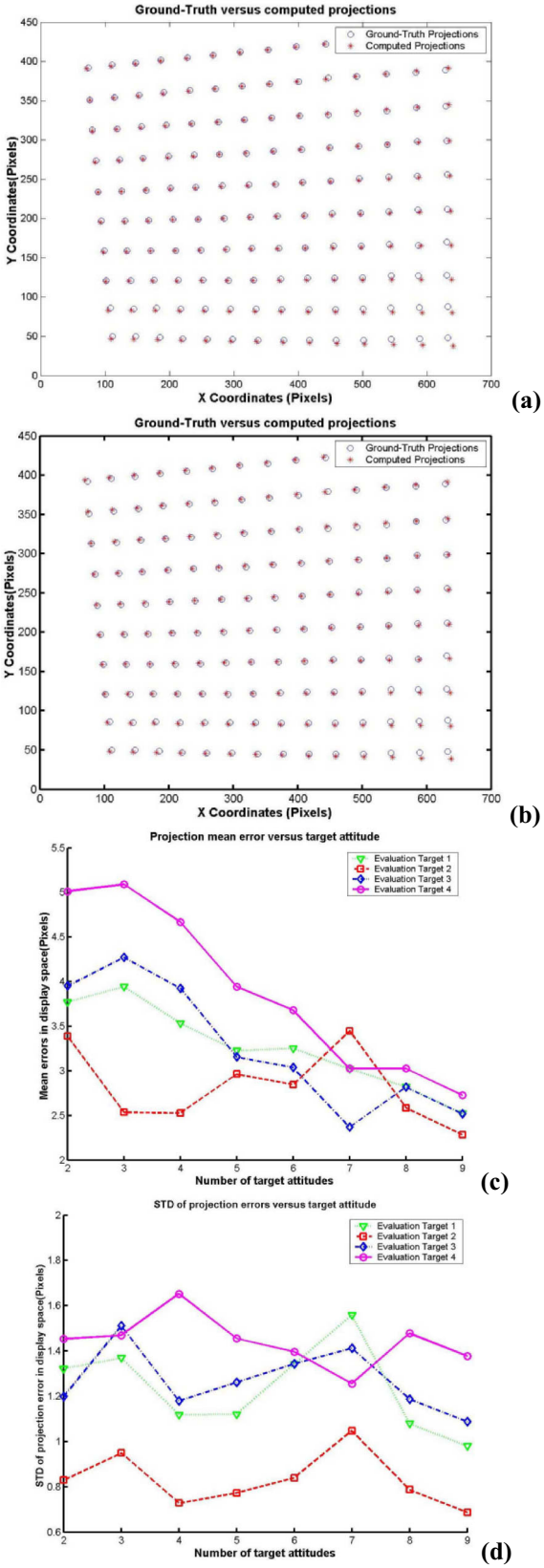

Fig. 9 The relationship of registration accuracy and convergence with calibration target attitudes

the world-coordinates of the points measured with the stylus, the projections of these samples in the display space were computed from the calibrated extrinsic and intrinsic transformations $(\hat{M}_{ext}, \hat{M}_{\text{int}})_i$ (i=1..12). Figures 8-a and 8-b show the error distributions corresponding to i=2 and i=12, respectively, for evaluation target 1. Figures 8-c and 8-d show the mean and standard deviation of the projection errors corresponding to the 4 different evaluation targets. The graphs illustrates that dense samples achieved better accuracy than sparse samples, and about 40 samples widely-distributed on 9 calibration targets can achieve convergence. It is important to note that the small number of samples has to maintain wide distribution in the world space.

5.4.2. Number of target attitudes

The number of calibration target attitudes may affect accuracy and convergence. Using the same calibration samples, instead of down-sampling the points on each target at certain intervals, we utilized the full samples at each selected target and computed the extrinsic and intrinsic transformations at 8 different number of target combinations, $(\hat{M}_{ext}, \hat{M}_{\text{int}})_{i=2..9}$.

Using the same evaluation target obtained in the last experiment, we computed the projections of the evaluation samples in the display space from the calibrated extrinsic and intrinsic transformations $(\hat{M}_{ext}, \hat{M}_{\text{int}})_i$ (i=2..9), based on their world-coordinates of the points measured with the stylus. Figures 9-a and 9-b show the error distributions corresponding to 2 targets and 9 targets, respectively, for evaluation target 1. Figures 9-c and 9-d show the mean and standard deviation of the projection errors corresponding to the 4 different evaluation targets. The graphs illustrate that large number of targets achieved better accuracy, while target number has less impact on accuracy than sample number.

The graphs in Figures 9 indicate that increasing the number of target attitudes can improve calibration accuracy and convergence at some extent, while 2 calibration targets with total 240 samples can achieve convergence. These graphs further indicate that the sample accuracy is more critical than sample numbers or variation of target attitudes. For example, in Figure 9, the good convergence at target number 6 was deteriorated at target number 7 due to the usage of a poor group of samples.

6. Conclusion

The head-mounted projective display (HMPD) has been proposed as an alternative solution to optical see-through devices. This paper presents a systematic calibration process to address static registration issue in a custom-designed augmented reality system, which is based upon the recent advancement of head-mounted projective display (HMPD) technology. Following a concise review of the HMPD concept and system configuration, we present in detail a computation model for the system calibration, describe the calibration procedures to obtain the estimations of the unknown transformations, and include the calibration results, evaluation experiments and results. In future work, more accurate calibration and less time-consuming methods will be investigated and comprehensive evaluation on the perception performance will be carried out.

Acknowledgements

We would like to specially acknowledge our collaborators Dr. Jannick Rolland of the ODALab at the University of Central Florida for her stimulating discussion, and Dr. Ken Smith and the 3M Inc. for supplying retro-reflective samples. We would also like to thank Leonard D. Brown for his work on developing the testbed program. This paper is based on work supported by National Science Foundation Grant IIS 00-83037 ITR and Beckman Institute Equipment Allocations.

References

[1] Y. Argotti, L. Davis, V. Outters, and J. P. Rolland, **"**Dynamic superimposition of synthetic objects on rigid and simple-deformable real objects", *Proceedings of IEEE International Symposium on Augmented Reality 2001*, pp. 5-10, October 29-30, 2001, New York, NY.

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

[3] M. Bajura, H. Fuchs, and R. Ohbuchi, "Merging virtual objects with the real world: Seeing ultrasound imagery within the patient," *Computer Graphics (Proc. SIGGRAPH Conf.)*, pp203-210, Chicago, IL, July 1992.

[4] M. Bajura and U. Neumann. "Dynamic registration correction in augmented-reality systems," *Proc. Virtual Reality Ann. Int'l Symp. (VRAIS'95)*, pp189-196, Research Triangle Park, N.C., Mar. 1995.

[5] D. Buxton, G W. Fitzmaurice. "HMDs, caves and chameleon: a human-centric analysis of interaction in virtual space". *Computer Graphics (ACM)*, Vol. 32, No. 4, Nov. (1998), 69-74.

[6] T. Caudell and D. Mizell, "Augmented reality: An application of heads-up display technology to manual manufacturing processes," *Proc. Hawaii Intl Conf. Systems Sciences*, pp659-669, Jan 1992.

[7] Michael Deering, High resolution virtual reality, Computer Graphics, Prof SIGGRAPH'92 Conf., 26(2), 195-202, July 1992

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

[9] S. Feiner, B. MacIntyre, and D. Seligmann, "Knowledge-based augmented reality," *Comm. ACM*, 36(7): 53-62, July 1993.

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

[11] 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 sterertaxy, image guided surgery, and enhanced reality visualization," *IEEE Trans. On Medical Imaging*, 15(2), 129-140, April 1996.

[12] Hong Hua, A. Girardot, Chunyu Gao, and J. P. Rolland. "Engineering of head-mounted projective displays". *Applied Optics*, 39 (22), 2000, 3814-3824.

[13] Hong Hua, Chunyu Gao, Frank Biocca, and Jannick P. Rolland, "An Ultra-light and Compact Design and Implementation of Head-Mounted Projective Displays", *Proceedings of IEEE-VR 2001*, p. 175-182, March 12-17, 2001, Yokohama, Japan.

[14] Hong Hua, Chunyu Gao, Leonard Brown, Narendra Ahuja, and Jannick P. Rolland, "Using a head-mounted projective display in interactive augmented environments", in *Proceedings of IEEE International Symposium on Augmented Reality 2001*, pp. 217-223, October 29-30, 2001, New York, NY.

[15] Hong Hua, Chunyu Gao, Leonard Brown, Narendra 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 *IEEE VR 2002 Proceedings*, 81-89 Orlando, FL, March 22-28, 2002

[16] Masahiko Inami, Naoki Kawakami, Dairoki Sekiguchi, Yasuyuki Yanagida, Taro Maeda, and Susumu Tachi, "Visuo-haptic display using head-mounted projector", *Proceedings IEEE Virtual Reality 2000*, IEEE Comput. Soc. 2000, pp.233-40. Los Alamitos, CA, USA.

[17] A. Janin, D. Mizell, and T. Caudell, "Calibration of head-mounted displays for augmented reality applications," *Proc Virtual Reality Ann. Int'l Symp, (VRAIS'93),* pp246-255, Sept. 1993.

[18] Naoki Kawakami, Masahiko Inami, Dairoku Sekiguchi, Yasuyuki Yangagida, Taro Maeda, and Susumu Tachi, "Object-oriented displays: a new type of display systems—from immersive display to object-oriented displays", *IEEE SMC'99 Conference Proceedings*, 1999 IEEE International Conference on Systems, Man, and Cybernetics, Vol.5, 1999, pp.1066-9 vol.5. Piscataway, NJ, USA.

[19] Ryugo Kijima and Takeo Ojika, "Transition between virtual environment and workstation environment with projective head-mounted display", *Proceedings of IEEE 1997 Virtual Reality Annual International Symposium,* IEEE Comput. Soc. Press. 1997, pp.130-7. Los Alamitos, CA, USA.

[20] W. Lorensen, H. Cline, C. Nafis, R. Kikinis, D. Altobelli, and L. Gleason, "Enhancing reality in the operating room," *Proc. Visualization'93 Conf.*, pp. 410-415, Los Alamitos, CA., Oct. 1993.

[21] Erin McGarrity and Mihran Tuceryan, "A method for calibrating see-through head-mounted displays for AR," *2nd IEEE and ACM International Workshop on Augmented Reality*, pp75-84, 1999.

[22] Erin McGarrity Y. Genc, M. Tuceryan, C. Owen, and N. Navab, "A new system for online quantitative evaluation of optical see-through augmentation," In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality*, pp. 157-66, New York, NY, October 2001.

[23] P. Milgram, S. Zhai, D. Drascic, and J. J. Grodski, "Applications of augmented reality for human-robot communication," *Proc. IROS'93: Int'l Conf. Intelligent Robots and Systems*, pp1467-1472, Yokohama, July 1993.

[24] Takashi Oishi and Susumu Tachi, "Methods to calibrate projection transformation parameters for see-through head-mounted displays," Presence: *Teleoperators and Virtual Environments (MIT Press), 5(1), 122-135, 1996.*

[25] J. Parsons, and J. P. Rolland, "A non-intrusive display technique for providing real-time data within a surgeons critical area of interest," *Proceedings of Medicine Meets Virtual Reality98*, 1998, 246-251.

[26] J. P. Rolland, and H. Fuchs, "Optical versus video see-through head-mounted displays in medical visualization," *Presence: Teleoperators and Virtual Environments (MIT Press),* 9(3), (2000), 287-309.

[27] J. P. Rolland, J. Parsons, D. Poizat, and D. Hancock, "Conformal optics for 3D visualization," *SPIE-Int. Soc. Opt. Eng. Proceedings of SPIE-the International Society for Optical Engineering*, vol.3482, 1998, pp.760-4.

[28] David A. Southard, Transformations for stereoscopic visual simulation, *Computer & Graphics* Vol.16, No.4, 401-410, 1992.

[29] Roger Y. Tsai. "A versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", *IEEE Journal of Robotics and Automation*, Vol. RA-3, No. 4, August 1987, 323-344.

[30] Mihran Tuceryan, et al. "Calibration requirements and procedures for a monitor-based augmented reality system," *IEEE Trans. Visualization and Computer Graphics*, 1(3): 255-273, Sep. 1995.

[31] M. Tuceryan and N. Navab. "Single point active alignment method (SPAAM) for optical see-through HMD calibration for AR," In *International Symposium for Augmented Reality*, pp. 149–158, Munich, Germany, October 2000.

[32] N. A. Valyus, Stereoscopy, Focal Press, London, 1996.