Modeling Physical Structure as Additional Constraints for Stereoscopic Optical See-Through Head-Mounted Display Calibration

Long Qian* Alexander Winkler† Bernhard Fuerst‡ Peter Kazanzides§ Nassir Navab¶

Johns Hopkins University, Baltimore, MD, United States
Technische Universität München, Munich, Germany

Abstract

For stereoscopic optical see-through head-mounted display calibration, existing methods that calibrate both eyes at the same time highly depend on the HMD user's unreliable depth perception. On the other hand, treating both eyes separately requires the user to perform twice the number of alignment tasks, and does not satisfy the physical structure of the system. This paper introduces a novel method that models physical structure as additional constraints and explicitly solves for intrinsic and extrinsic parameters of the stereoscopic system by optimizing a unified cost function. The calibration does not involve the unreliable depth alignment of the user, and lessens the burden for user interaction.

Index Terms: Augmented Reality, SPAAM, OST-HMD Calibration and Human Factors

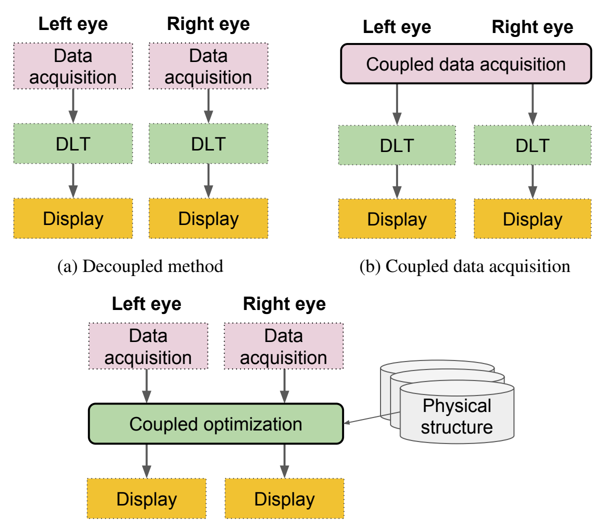
1 Introduction

Augmented Reality (AR) integrates virtual objects into the user's perception of reality by graphics technologies. Unlike a video see-through head-mounted display (VST-HMD), which generates virtual objects on top of video capture, an optical see-through head-mounted display (OST-HMD) renders only virtual objects and requires a user-specific calibration procedure to correctly align virtual and real objects. Since the user's vision is not accessible, the calibration procedure usually requires user interaction. Even for interaction-free calibration procedures [4], the initialization still requires user-specific information.

Single Point Active Alignment Method (SPAAM) [6] is widely used for the calibration of optical see-through devices due to its simplicity and accuracy. In the traditional SPAAM calibration procedure, the user changes his or her head pose to make the alignment between a virtual crosshair and real world target, whereas in the stylus-marker method [1], the user aligns a finger with the crosshair. In SPAAM-based methods, 3D-2D corresponding points are collected and the Direct Linear Transformation (DLT) algorithm is applied to solve for the projection matrix from the 3D spatial coordinates to 2D screen coordinates.

For stereoscopic augmented reality platforms, there are two categories of calibration methods. **Decoupled methods** treat both eyes individually: 3D-2D point pairs are collected separately for each eye. In this case, there is no coupling between the two eyes. Decoupled methods are based on the assumption that if the virtual object is aligned with the real-world target for both eyes, then humans can perceive the virtual display at the correct depth. **Coupled data acquisition** methods display the virtual object at a certain depth, and require the user to align it with its real counterpart at the correct depth. In each alignment task, the 3D-2D point pairs are collected for both left and right eyes [2].

In this poster, the **coupled optimization** method is proposed for calibrating stereoscopic OST-HMDs. The data acquisition stage of the coupled optimization method is the same as for the decoupled method. The mapping between the 3D-2D points are calculated by optimizing a cost function under various physical constraints. The introduction of these constraints in the optimization stage improves the consistency with the physical property of the stereoscopic system and, at the same time, reduces the parameter space, thus lessening the user interaction burden. Figure 1 presents the workflow of the *decoupled method*, *coupled data acquisition method*, and *coupled optimization method*.



(a) Decoupled method (b) Coupled data acquisition

(c) Coupled optimization

Figure 1: The workflow for different categories of calibration methods for stereoscopic OST-HMDs: (a) Decoupled methods treat left and right eye separately and do not consider the physical constraints of the two eyes. (b) Coupled data acquisition is difficult for the user to perform due to the requirement for accurate depth alignment. (c) Coupled optimization (proposed method) does not require depth alignment and optimizes under the physical constraints.

2 Method

The coupled optimization method takes advantage of the effectiveness of SPAAM interaction [5], and overcomes the inaccuracy by adding physical constraints to the optimization problem.

2.1 Parameters and Constraints

In order to model physical properties explicitly in the state space, the parameters of the coupled optimization method are chosen as

*e-mail: long.qian@jhu.edu
†e-mail: alexander.winkler@tum.de
‡e-mail: be.fuerst@jhu.edu
§e-mail: pkaz@jhu.edu
¶e-mail: nassir.navab@tum.de

the intrinsic and extrinsic parameters of the projection matrix.

$$\begin{bmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & g_{23} & g_{24} \\ g_{31} & g_{32} & g_{33} & 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & s & d_x \\ 0 & \alpha_y & d_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} R & T \end{bmatrix} \quad (1)$$

where $R = R(q_x, q_y, q_z, q_w)$ is the extrinsic rotation matrix represented by quaternions and $T = [t_x, t_y, t_z]'$ is the extrinsic translation vector. The full parameter space of a stereoscopic OST-HMD is

$$\begin{aligned} \Psi = \{ & \alpha_x^l, \alpha_y^l, s^l, d_x^l, d_y^l, q_x^l, q_y^l, q_z^l, q_w^l, t_x^l, t_y^l, t_z^l, \\ & \alpha_x^r, \alpha_y^r, s^r, d_x^r, d_y^r, q_x^r, q_y^r, q_z^r, q_w^r, t_x^r, t_y^r, t_z^r \} \end{aligned} \quad (2)$$

with additional constraints that the quaternions are unit quaternions.

Constraints on the parameter space are related to physical properties of the eyes and the OST-HMD screens:

- Pixel density of x and y axis on the screen is same, i.e. $\alpha_x^l = \alpha_y^l$
- Pixel density of the two virtual cameras is same, i.e. $\alpha_x^l = \alpha_x^r$
- There is no skew in user perceived image, i.e. $s^l = s^r = 0$
- Both eyes have the same viewing direction, i.e. $q^l = q^r$
- The line between both eyes is parallel to the horizontal image axis, i.e. $t_y^l = t_y^r$, $t_z^l = t_z^r$
- The interpupillary distance is given, i.e. $\left| t_x^l - t_x^r \right| = IPD$

These constraints can be used to reduce the state space directly, instead of handling constraints in the optimization stage via penalty functions.

The choice of constraints is dependent on the system and the application. With more constraints applied to the calibration, fewer degrees of freedom remain in the system. The benefit of this is that the requirement for the number of alignment tasks is reduced.

2.2 Optimization

With intrinsic and extrinsic parameters explicitly expressed, the optimization problem is nonlinear, as shown in Eq. 1.

Problem Statement: Given 3D target positions in world coordinates $\{P_{Left}\}_i$ and $\{P_{Right}\}_j$, with corresponding 2D screen crosshair positions in pixel coordinates $\{I_{Left}\}_i$ and $\{I_{Right}\}_j$, $\Gamma(p, \theta)$ computes the projection of the 3D point on the screen with a projection matrix parameterized by $\theta \in \Psi$. The cost function $F(\theta)$ is the total reprojection error:

$$F(\theta) = \sum_i \left\| I_{Left}^i - \Gamma_L(P_{Left}^i, \theta) \right\| + \sum_j \left\| I_{Right}^j - \Gamma_R(P_{Right}^j, \theta) \right\|$$

The optimization problem is defined as:

$$\arg\min_{\theta} F(\theta), \theta \in \Psi \quad (3)$$

which is a nonlinear problem where the physical structure is incorporated by reducing the dimension of the parameter space. Iterative methods, e.g. gradient descent, Newton's method, Levenberg-Marquardt algorithm, can be used to calculate the parameters that result in the minimum reprojection error locally. Special attention should be paid to the quaternion parameters, which should be normalized after each iteration.

3 Experiment and Discussion

A pilot study comparing the decoupled method and coupled optimization method is conducted. The user acquired 100 3D-2D point alignments for both eyes in the data acquisition stage, using Moverio BT-200. The 200 corresponding points are utilized in the separate DLT calculation and the coupled optimization.

For the **decoupled method**, the decomposed intrinsic and extrinsic parameters are not consistent with the physical structure of the two eyes: (i) pixel density for both screens and both axes is different ($\alpha_x^l = 2637.88$, $\alpha_x^r = 2797.33$, $\alpha_y^l = 2506.21$, $\alpha_y^l = 2608.20$), (ii) skew factor is non-zero ($s^l = -95.69$, $s^r = 22.39$), (iii) rotation between the two virtual cameras is not identity ($q_x = -0.088$, $q_y = -0.066$, $q_z = -0.004$, $q_w = 0.994$), (iv) translation between the two eyes is not parallel to the horizontal image axis. The average reprojection error of the decoupled method is 6.211 pixels.

In the **coupled optimization method**, the physical constraints are strictly observed, as described in section 2.1. The average reprojection error is 8.34 pixels. The reprojection error is larger than for the decoupled method because the parameters in the decoupled method are overfitting the experiment data without considering the coupling between the projection matrices of the two eyes.

An initial state vector $\theta_0 \in \Psi$ is required in coupled optimization. Due to the fact that the gradient descent method can find local minima, the initial value should be close enough to the actual value. The virtual camera formed by the user's eye and HMD screen is similar between different people, so it should be possible to use a nominal initial value for all calibrations based on coupled optimization. Coupled optimization takes more time than DLT. However, since OST-HMD calibration is separate from the actual application, the time consumed by an iterative method (several seconds in the experiment) is not critical.

4 Conclusion

In this poster, a new method is proposed that models physical structure as additional constraints for stereoscopic OST-HMD calibration. The **coupled optimization method** provides the advantage of the decoupled data acquisition and, at the same time, explicitly follows the physical requirement of the stereoscopic system. Intrinsic and extrinsic parameters form the state vector, and the physical constraints directly reduce parameter dimension. Different constraints can be chosen according to different HMD platforms and applications. The cost function to be optimized is the total reprojection error, also known as geometric distance, which is more straightforward compared to the algebraic distance in the DLT algorithm [3]. A pilot experiment comparing the decoupled method and the coupled optimization demonstrated that the decoupled method violated the physical structure of the stereoscopic system, whereas the coupled optimization method successfully represented this structure.

Acknowledgments

This work was supported in part by DoD W81XWH-15-C-0156, awarded to Juxtopia LLC.

References

[1] A. Fuhrmann, D. Schmalstieg, and W. Purgathofer. Fast calibration for augmented reality. In *ACM Symp. on Virtual Reality Software and Technology*, pages 166–167, 1999.

[2] Y. Genc, F. Sauer, F. Wenzel, M. Tuceryan, and N. Navab. Optical see-through HMD calibration: A stereo method validated with a video see-through system. In *IEEE/ACM Intl. Symp. on Augmented Reality (ISAR)*, pages 165–174, 2000.

[3] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[4] Y. Itoh and G. Klinker. Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization. In *IEEE Symp. on 3D User Interfaces (3DUI)*, pages 75–82, 2014.

[5] A. Tang, J. Zhou, and C. Owen. Evaluation of calibration procedures for optical see-through head-mounted displays. In *IEEE/ACM Intl. Symp. on Mixed and Augmented Reality (ISMAR)*, page 161, 2003.

[6] M. Tuceryan and N. Navab. Single point active alignment method (SPAAM) for optical see-through HMD calibration for AR. In *IEEE/ACM Intl. Symp. on Augmented Reality (ISAR)*, pages 149–158, 2000.