BeamStellar: Low-Latency, 6-DoF Head Tracking for Beaming Displays with Spatio-Temporal LED Encoding

Jonas Weigand *
Technical University of Munich

Yuichi Hiroi †
Cluster Metaverse Lab

Yuta Itoh ‡
The University of Tokyo

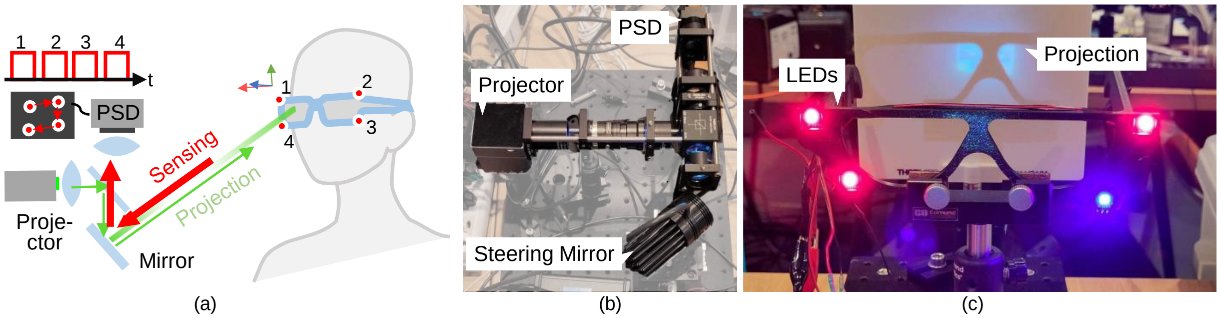

Figure 1: An overview of our BeamStellar concept, a low-latency 6DoF tracking design for beaming display systems. (a) a schematic diagram of the projection–sensing unit with a PSD (position sensing detector) and the light-receiving glasses with a coded LED array. (b) A benchtop system of the projection unit. (c) a mockup of glasses with an LED array. The colors of LED are irrelevant to the algorithm. A projected light from the projection–sensing unit is visible on a screen behind the glasses.

Abstract

Conventional augmented reality (AR) head-mounted displays (HMDs) require the integration of all components, including optics and processors, which presents significant barriers to practical implementation due to weight and heat generation concerns. While Beaming Displays address these challenges by offloading computing and projection functions to the environment, existing solutions leaving 6-degrees of freedom (6DoF) image projection with low motion-to-photon (M2P) latency. In this position paper, we introduce our trial BeamStellar, a Beaming Display system that integrates 6 DoF head tracking by combining spatio-temporally encoded LED patterns with a position-sensing detector. Through digital signal processing using an FPGA, the system aims a theoretical latency of less than 200 microseconds. This implementation alleviates the hardware burden on users, facilitating seamless AR experiences that seamlessly blend virtual content with the physical world.

Keywords: Beaming Displays, Optical See-through Display, Low-latency head tracking, Spatio-temporal Light pattern

1 Introduction

Optical see-through head-mounted displays (OST-HMDs) overlay digital content directly into the user's field of view (FoV), seamlessly integrating virtual and physical environments. Despite their potential as a key technology for augmented reality (AR) experiences, conventional OST-HMDs face significant trade-offs among critical performance factors such as weight, power consumption, and computational capacity. The integration of essential components, including optics, sensors, batteries, and processors, into a compact form factor often results in compromises, such as reduced battery life, increased heat load, and limited FoV and computing power.

To overcome these challenges, Itoh et al. [5] proposed the Beaming Display approach, which offloads computational and optical processing to the surrounding environment. In this concept, projection units are placed externally, while users wear lightweight, passive headsets to view AR images. This reduces the user's hardware burden while enabling high computational power. However, practical implementation of Beaming Display systems requires addressing a critical technical hurdle: achieving low motion-to-photon (M2P) latency [6] while tracking freely moving viewpoints. While Hiroi et al. [4] demonstrated a beaming display system with a latency of 133 µs using a 2D position-sensing detector (PSD) and a single LED, their approach was limited to tracking 3D position (i.e., xyz) only, lacking the full 6 degrees-of-freedom (DoF) pose estimation necessary for dynamic AR environments.

In this position paper, we present BeamStellar, an ongoing project to achieve full 6-DoF headset tracking in Beaming Display systems. Our approach leverages spatiotemporally encoded LED patterns and a PSD to ensure precise alignment between projected images and headset position. By integrating digital signal processing on an FPGA, our system can minimize processing latency while maintaining accurate headset pose tracking. This work outlines BeamStellar's design principles with current hardware design and analyzes the architecture's theoretical latency performance.

2 BeamStellar Concept

BeamStellar's goal is to achieve low-latency 6 DoF headset tracking in Beaming Displays by combining spatio-temporally encoded LED patterns with a single PSD.

2.1 System Architecture

Figure 1 shows an overview of the BeamStellar system architecture. The system consists of two main units: a projection–sensing module and light-receiving glasses with an LED array (Fig. 1(b)). The tracking works as follows: the LEDs on the glasses turn on sequentially, and then the module captures each LED and detects their 2D positions, allowing the reconstruction of the headset's 6DoF pose (Fig. 1(a)).

The projection–sensing module has a PSD, steering mirror, and projector. The sensing and the projection paths align optically to keep the tracking and projection in the same FoV. A PSD is an analog 2D sensor that can precisely detect the center of projected light at a fast analog response. For example, our PSD (Thorlabs 2D Lateral Effect Position Sensor PDP90A) has a position resolution

*e-mail: jonas.weigand@tum.de
†e-mail: y.hiroi@cluster.mu
‡e-mail: yuta.itoh@iii.u-tokyo.ac.jp

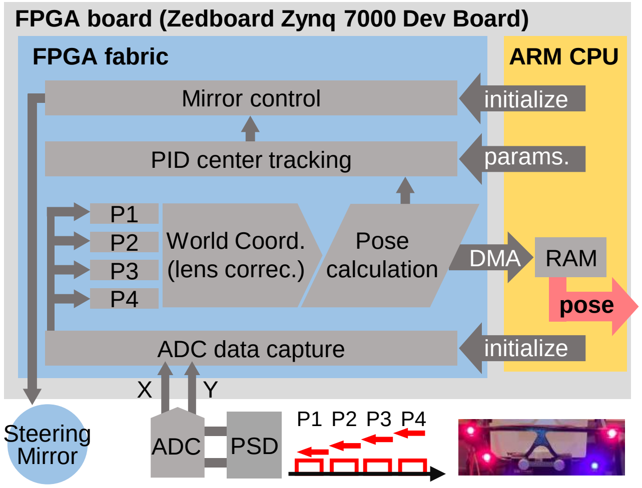

Figure 2: A system diagram of the current processing architecture design (Sec. 2.1).

of 0.75 um. The PSD sensor equipped with an image-forming lens connects to a controller unit that gives a fast data bandwidth. Our prototype uses a Thorlabs KPA101 with 15 kHz in the open-loop mode. The signal drives a fast steering mirror. We use Optotune MR-E-2, which operates at a 10 kHz sampling rate. For projection, we used a commodity DLP projector (Felicross PicoCube) with a custom projection lens array for glasses-size image projection.

The light from each LED on the headset forms a point distribution on the PSD sensor, which calculates the centroid. The controller outputs this as an XY voltage signal. Using the signal, imaging lens parameters, and LED array geometry, a processing board reconstructs the array's 6DoF pose with the Perspective-n-Point (PnP) method (n=4 ensures uniqueness). Analog and embedded processing on such a board minimizes M2P latency. The following outlines our progress toward this goal.

2.2 Headset Tracking and Mirror Control

Our tracking architecture is similar to Blate et al.'s low-latency multi-LED head tracking [3] where two sets of PSD sensors observe an LED array on a headset. Our sensing unit uses one PSD module integrated into a steering projector path, and the headset uses a temporary modulated LED array. The difference between the two approaches stems from the need for beaming projection.

We develop the tracking approach with the projection–sensing unit above and a 3D-printed glasses frame with four LEDs (OSRAM GH CSSRM6.24 Hyper Red LEDs) positioned at the frame's four corners (Fig. 1(c)). Although the glasses ideally should include light-receiving optics [1], we do not implement it since the focus of this project is tracking. Nonetheless, the projection–sensing module includes a projection capability, demonstrating its potential for future applications.

For the data processing of the PSD's output signals, we employ a combination of a successive approximation register analog-to-digital converter (SAR-ADC) and a field-programmable gate array (FPGA) to achieve high-speed signal processing. The SAR-ADC digitizes the analog signals from the PSD, and the FPGA processes this data to estimate the 6-DoF pose of the headset. This combination ensures real-time performance while maintaining low M2P latency.

Figure 2 illustrates the BeamSteller concept, where multiple light flashes from the glasses strike the PSD. A 16-bit SAR-ADC (Analog Devices, AD4630-16) operating at 2 million samples per second (MSPS) across two channels digitizes the sequential signals. An 8-lane SPI interface (configured with four lanes per channel, host clock control, and DMA mode) then sends these digitized outputs to an FPGA, Diligent Zedboard Zynq 7000), which processes them through a custom digital signal processing (DSP) pipeline. To ensure correct point correspondences, the FPGA uses temporal gaps to segment the incoming data into four "buckets," one for each corner LED on the glasses. After sampling all four signals, the system converts the 2D LED coordinates on the PSD into 3D world coordinates, applies lens distortion correction, and uses a 4PnP homography to estimate the glasses' 3D pose. Because this computation primarily involves a small number of matrix multiplications (four 3×3 and one 4×4), hardware acceleration greatly improves performance. Finally, a PID controller adjusts the mirror to align the PSD's center with the glasses' virtual center, enabling real-time tracking.

The system subsequently steers the mirror to center the detected four-point pattern on the PSD. This control mechanism establishes a direct link between the PSD signals and the mirror control, enabling low-latency tracking similar to the method described by Hiroi et al. [4]. Next, the system estimates a 6-DoF pose and calculates a homography matrix from the four detected points using the PnP method. This matrix computation is performed in parallel on the FPGA to ensure fast pose estimation. By combining the homography transformation matrix with the steering mirror's rotation angle, the system computes the user's 6-DoF head pose. This information enables real-time keystone correction for projected images.

2.3 M2P Latency Estimation

Our current prototype under development integrates the described components. With the current hardware specification, we calculate the theoretical motion-to-pose (M2P) latency of such a system.

The M2P latency consists of:

- **LED-to-PSD Detection:** 66 µs, including 60 ns LED drive delay and 66 µs PSD response time.
- **Signal Processing:** 121 µs, with 500 ns ADC conversion and 0.5 µs FPGA processing.
- **Pose Control:** 20 µs for homography computation and 100 µs for mirror updates.

The total latency expected is 187 µs, improving by three orders of magnitude over camera-based systems (10–20 ms).

3 Discussion and Future Work

This position paper presented the vision of BeamStellar, a beaming display system that achieves 6-DoF tracking with 187 µM2P latency through spatio-temporal LED patterns and PSD integration. The system's low latency and lightweight design enable seamless integration of physical and cyber spaces.

By overcoming the weight and latency challenges of traditional OST-HMDs, BeamStellar potentially enables AR applications in dynamic scenarios such as sports training. The system can also advance research in human visual cognition by enabling ultra-low latency visual stimulus presentation, contributing to the understanding of temporal perception in visual processing and perceptual characteristics in AR environments.

We plan to evaluate the actual performance with the prototype. Future work includes reducing latency by optimizing FPGA processing and improving the steering mirror. We also aim to extend the tracking range and support multi-user environments through coordinated control of multiple projection modules [2]. These improvements will contribute to a more natural AR experience.

Acknowledgements

This work was partially supported by JST ASPIRE Grant Number JPMJAP2327, JST FOREST Grant Number JPMJFR206E, and JSPS KAKENHI Grant Number JP20H05958, JP23H03430 (JP23K28120), and JP23K16920 Japan.

References

[1] K. Akşit and Y. Itoh. Holobeam: Paper-thin near-eye displays. In *Proc. of IEEE VR 2023*, pages 581–591. IEEE, 2023. 2

[2] H. Aoki, T. Tochimoto, Y. Hiroi, and Y. Itoh. Towards co-operative beaming displays: Dual steering projectors for extended projection volume and head orientation range. *IEEE TVCG*, 30(5):2309–2318, May 2024. 2

[3] A. Blate, M. Whitton, M. Singh, G. Welch, A. State, T. Whitted, and H. Fuchs. Implementation and evaluation of a 50 khz, 28us motion-to-pose latency head tracking instrument. *IEEE TVCG*, 25(5):1970–1980, 2019. 2

[4] Y. Hiroi, A. Watanabe, Y. Mikawa, and Y. Itoh. Low-latency beaming display: Implementation of wearable, 133 μs motion-to-photon latency near-eye display. *IEEE TVCG*, 29(11):4761–4771, Nov. 2023. 1, 2

[5] Y. Itoh, T. Kaminokado, and K. Akşit. Beaming displays. *IEEE TVCG*, 27(5):2659–2668, 2021. 1

[6] P. Lincoln, A. Blate, M. Singh, T. Whitted, A. State, A. Lastra, and H. Fuchs. From motion to photons in 80 microseconds: Towards minimal latency for virtual and augmented reality. *IEEE TVCG*, 22(4):1367–1376, 2016. 1