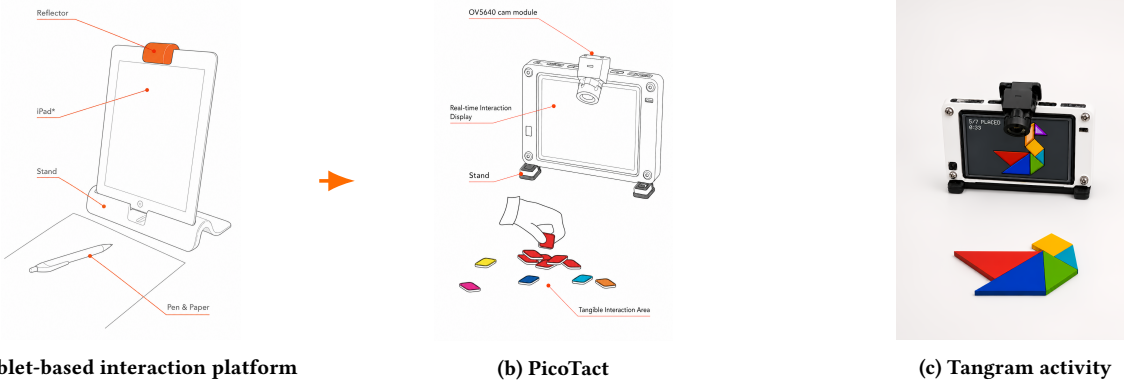


# PicoTact: Low-Cost Embedded Vision for Tangible Learning Experiences

Anonymous Author(s)



(a) Tablet-based interaction platform

(b) PicoTact

(c) Tangram activity

**Figure 1: PicoTact replaces an Osmo-style tablet-and-reflector setup with a standalone ESP32-based camera-and-display module; a tangram activity illustrates its physical-to-digital feedback loop.**

## Abstract

PicoTact is a low-cost, standalone embedded-vision platform for screen-guided interaction with physical learning materials. An ESP32-class microcontroller, OV5640 camera, SPI display, and lightweight vision pipeline recognize constrained tangible pieces and provide immediate local feedback without a tablet, external host, network, or cloud service. We present the prototype, its classical vision pipeline, and interfaces spanning puzzles, maps, spelling, counting, and tangible coding.

## CCS Concepts

• **Human-centered computing** → **Tangible user interfaces**; *Interaction devices*; • **Applied computing** → *Interactive learning environments*.

## Keywords

tangible interaction, embedded vision, microcontrollers, educational interfaces, physical manipulatives, on-device computing

## ACM Reference Format:

Anonymous Author(s). 2026. PicoTact: Low-Cost Embedded Vision for Tangible Learning Experiences. In *Adjunct Proceedings of the 39th Annual ACM Symposium on User Interface Software and Technology (UIST '26 Adjunct)*, November 2–5, 2026, Detroit, MI, USA. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
UIST '26 Adjunct, Detroit, MI, USA

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM  
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## 1 Introduction

Camera-based tangible interfaces make ordinary manipulatives responsive: a learner places a tile or puzzle piece and a nearby screen reacts. Osmo popularized this pattern across tangrams, spelling, arithmetic, drawing, and coding, while research systems have connected physical objects to digital feedback for programming and learning [1, 5, 8].

While Osmo requires a compatible tablet, reflector and base, app, and activity-specific tangible kits, PicoTact integrates the camera, recognition logic, activity state, and feedback display into one self-contained device. It also uses designed manipulatives, but requires neither an external app nor network connectivity during use.

This work-in-progress poster contributes (1) a standalone architecture for camera-mediated tangible interaction, (2) a low-compute vision pipeline for known tabletop materials, and (3) reusable activity patterns for puzzles, shapes, colors, letters, numbers, and tangible code. We do not yet claim learning gains or general-purpose object recognition.

## 2 Related Work

Tangible user interfaces couple digital information with graspable objects and physical action [2]. Educational systems have used augmented manipulatives for construction, experimentation, and abstract reasoning [5, 8]; prior comparisons and meta-analyses show broad application domains, with benefits shaped by task and design [3, 7].

PicoTact is closest to camera-mediated systems that recognize objects near a display. Strawbies used an Osmo attachment and tagged wooden tiles for tangible programming [1], while Teachable Reality supported vision-based authoring with everyday objects using more capable computing and machine-learning infrastructure

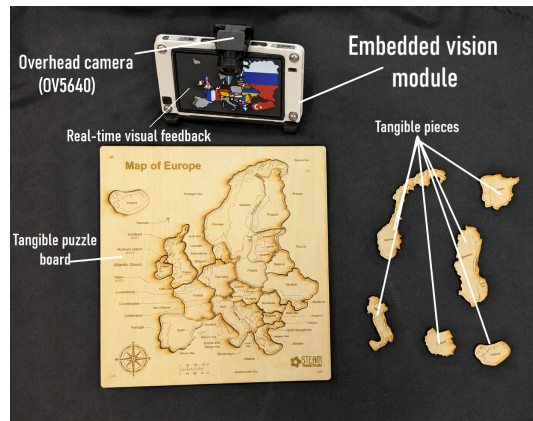


**Figure 2: The current PicoTact prototype combines an overhead OV5640 camera, embedded vision module, and real-time display.**



**Figure 3: Sample interfaces support (a) tangram assembly, (b) a geography puzzle, and (c) spelling, updating as students place tangible pieces within the camera frame.**

[4]. Paulo Blikstein et al.’s Mechanix automatically captured children’s tangible constructions to support documentation, reflection, and collaboration during engineering-design activities [6]. PicoTact instead targets fixed work surfaces, known object sets, classical vision, and immediate on-device feedback.

### 3 PicoTact

#### 3.1 Hardware and Interaction Loop

The prototype combines an ESP32-class microcontroller, OV5640 camera, SPI display, local storage, and a compact enclosure above the activity area (Figure 2). Its current off-the-shelf component cost is approximately \$25, with substantial room for reduction through volume purchasing, board integration, and a lower-cost enclosure. The downward-facing camera observes a board or mat while the learner-facing display provides prompts, progress, errors, and completion feedback. Processing remains local.

**physical action → camera frame → on-device detection → activity state → SPI feedback**

Activities reuse this loop by changing expected pieces, detection rules, and display content rather than the hardware. Figure 3 shows how the geography interface tracks recognized countries on the Europe board and leaves missing regions—including Spain, Italy, Norway, and Sweden—visually muted. The tangram interface reports that five of seven expected pieces for the swan are placed and renders the partial construction.

#### 3.2 Low-Compute Vision Pipeline

PicoTact uses a lightweight classical vision pipeline instead of a large neural model. Each frame is thresholded, denoised, and processed to extract contours and connected regions. The system then estimates shape, position, and orientation before confirming detections across multiple frames. A calibrated board area converts image coordinates into activity coordinates, while piece identity is inferred from color, geometry, size, templates, or fiducial markers. In preliminary bench tests under controlled lighting, PicoTact achieved an average end-to-end response time of 90 ms and correctly recognized 82% of piece placements across 72 trials.

Because PicoTact operates with known workspaces and object sets, each activity can use the simplest reliable cues. Tangrams rely on color and polygon geometry, counting activities use component counts, and coding blocks use printed markers or fixed symbols. This approach supports maps, rods, letter tiles, memory tasks, and physical programming blocks on the same low-cost platform.

### 4 Limitations and Future Work

The prototype is sensitive to lighting, shadows, texture, and low contrast; bright, even illumination and distinct pieces are currently most reliable. Large boards can exceed the OV5640 field of view and may require a wider lens, higher mount, tiling, or multiple cameras. Future work will measure latency and accuracy, study children’s interpretation of feedback, refine robustness and UI behavior, and test cheaper enclosures such as laser-cut acrylic.

## References

- [1] Felix Hu, Ariel Zekelman, Michael S. Horn, and Frances Judd. 2015. Strawbies: Explorations in Tangible Programming. In *Proceedings of the 14th International Conference on Interaction Design and Children (IDC '15)*. Association for Computing Machinery, 410–413. doi:10.1145/2771839.2771866
- [2] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '97)*. Association for Computing Machinery, 234–241. doi:10.1145/258549.258715
- [3] Yanhong Li, Meng Liang, Julian Preissing, Nadine Bachl, Michelle Melina Dutoit, Thomas Weber, Sven Mayer, and Heinrich Hussmann. 2022. A Meta-Analysis of Tangible Learning Studies from the TEI Conference. In *Proceedings of the Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '22)*. Association for Computing Machinery, Article 7, 17 pages. doi:10.1145/3490149.3501313
- [4] Kyzyl Monteiro, Ritik Vatsal, Neil Chulpongsatorn, Aman Parnami, and Ryo Suzuki. 2023. Teachable Reality: Prototyping Tangible Augmented Reality with Everyday Objects by Leveraging Interactive Machine Teaching. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, Article 459, 15 pages. doi:10.1145/3544548.
- [5] Mitchel Resnick, Fred Martin, Robert Berg, Richard Borovoy, Vanessa Colella, Kwin Kramer, and Brian Silverman. 1998. Digital Manipulatives: New Toys to Think With. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '98)*. Association for Computing Machinery, 281–287. doi:10.1145/274644.274684
- [6] Tiffany Tseng, Coram Bryant, and Paulo Blikstein. 2011. Collaboration Through Documentation: Automated Capturing of Tangible Constructions to Support Engineering Design. In *Proceedings of the 10th International Conference on Interaction Design and Children*. Association for Computing Machinery. doi:10.1145/1999030.1999044
- [7] Lesley Xie, Alissa N. Antle, and Nima Motamedi. 2008. Are Tangibles More Fun? Comparing Children's Enjoyment and Engagement Using Physical, Graphical and Tangible User Interfaces. In *Proceedings of the 2nd International Conference on Tangible and Embedded Interaction (TEI '08)*. Association for Computing Machinery, 191–198. doi:10.1145/1347390.1347433
- [8] Oren Zuckerman, Saeed Arida, and Mitchel Resnick. 2005. Extending Tangible Interfaces for Education: Digital Montessori-Inspired Manipulatives. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. Association for Computing Machinery, 859–868. doi:10.1145/1054972.1055093