Combining Embedded Computation and Image Tracking for Composing Tangible Augmented Reality

Tim Düwel, Nico Herbig, Denise Kahl, Antonio Krüger
German Research Center for Artificial Intelligence (DFKI)
Saarland Informatics Campus, Saarbrücken, Germany
{tim.duewel, nico.herbig, denise.kahl, krueger}@dfki.de

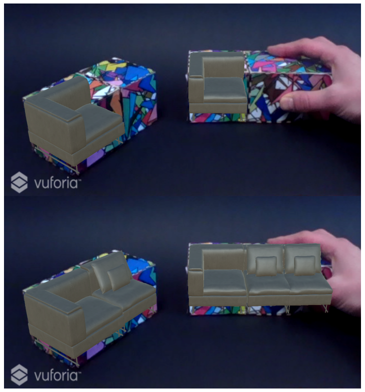

Figure 1: A system purely based on marker tracking (top) cannot augment blocks, which are occluded or perspectively distorted. The approach combining marker tracking and embedded computation (bottom) overlays both compositions correctly.

CHI'20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA
ACM ISBN 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.3383043

Abstract
This work proposes a combination of embedded computation and marker tracking to provide more robust augmentations for composed objects in Tangible Augmented Reality. By integrating conductive elements into the tangibles' sides, communication between embedded microprocessors is enabled, such that a connected composition can be computed without relying on any marker tracking information. Consequently, the virtual counterparts of the tangibles can be aligned, and this virtual composition can be attached to a single marker as a whole, increasing the tracking robustness towards occlusions and perspective distortions. A technical evaluation shows that this approach provides more robust augmentations if a tangible block in a composition is occluded by at least 50% or perspectively distorted by at least 40 to 50 degrees, depending on the block's size. Additionally, a test with users relying on the use case of a couch configuration tool shows promising results regarding usability and user experience.

Author Keywords
Tangible Augmented Reality; Augmented Reality; Tangible User Interface; Embedded Computation; Configuration; Composition; Assembling

CCS Concepts

•**Human-centered computing** → **Interactive systems and tools;** Mixed / augmented reality;

Introduction

By combining a Tangible User Interface (TUI) with an Augmented Reality (AR) view, Tangible Augmented Reality (TAR) systems provide tangible elements to manipulate visual reality augmentations [3]. By that, the direct, intuitive and natural interaction techniques of TUIs are combined with the visual flexibility of AR. To provide a TAR system, tangibles are spatially tracked and overlaid with virtual counterparts. As head-mounted displays (HMD), which often are used as AR interfaces, usually contain a camera, it is reasonable to rely on fiducial marker tracking to gather the position and rotation of a tangible relative to the rendering HMD. This also simplifies extending an application with an additional element, as only a new marker has to be printed and attached to a tangible. In the context of TAR, however, marker tracking also introduces a crucial drawback. If a marker is occluded or strongly distorted from a user's perspective, which frequently happens during direct interactions, the tracking breaks and the application is basically unusable.

This paper presents a TAR system with increased robustness against occlusions and perspective distortions for tangible compositions. Such *composing interfaces*, which rely on several elements that can be assembled in various ways, can for example be used for an interactive movie, where each tangible represents a different part of a story and users decide what will be shown next by attaching further elements, similarly to the interactive storytelling by Gorbet et al. [6]. Alternatively, various board games can easily be realized and extended by an interactive AR view with a composing TAR interface, for example *Carcassonne*, which is based on small cards depicting parts of castles and rivers that need to be combined properly. A last example for a composing interface is configuration, where each tangible corresponds to a different part of an item, which can be assembled in various ways for customization.

To increase the robustness against occlusions and perspective distortions for composing TAR interfaces, the system presented in this paper combines optical marker tracking with embedded computation. This means that microprocessors integrated into tangible elements enable the computation of a composition of such elements independent of the marker tracking. By that, the virtual counterparts of a composition can be arranged beforehand and afterwards attached as soon as a single marker is tracked, hence, the number of required tracked markers to provide proper augmentations decreases from one marker per tangible element to one marker per tangible composition.
In this paper, this concept is presented, implemented and technically evaluated to gather first insights on how the robustness of augmentations increases upon occurring occlusions and perspective distortions for TAR assembling. The use case of a couch configuration tool is explored and tested with users to receive first feedback on the implemented system.

Related Work

The combination of a TUI and AR was firstly presented by Billinghurst et al. [3], who realized an assembly application for furniture in a virtual room. Similarly, Kato et al. [9] present a city-planning system, allowing users to assemble virtual playgrounds with help of a TAR interface. Both of them avoid occlusions by construction, as they rely on indirect interaction by utilizing a marker-tracked paddle, or a cup respectively, as interaction tool to pick up, move and position virtual objects. In contrast, the core concept of composing TAR interfaces is direct and intuitive assembly of several tangibles without an additional interaction tool, wherefore occlusions have to be considered.

One approach to enhance tracking robustness is to augment a single object with several fiducial markers [7, 11]. The position of each marker on the tracked object is specified, such that detecting one of them suffices to compute the position and rotation of the underlying object. Ideally, the markers are attached in such a way that one of them is always visible to a camera and consequently, the underlying object can permanently be tracked.
For composing TAR interfaces, however, this approach is not sufficient. Although each tangible should be augmented with more than a single marker, it nevertheless cannot be guaranteed that a single tangible is not completely occluded during direct assembling interactions.

To avoid occlusions, other devices than the HMD's integrated camera can be used to track the tangible elements, for example external cameras [10, 13], RFID systems [5] or ultrasonic systems [12]. However, all of these approaches restrict the potential flexibility, simplicity and portability of a TAR interaction space, as they require additional external tracking sensors.

For composing TAR interfaces, another approach to increase the robustness against occlusions is to embed microprocessors into the tangible elements and compute the composition solely based on connectivity information, similar to the ActiveCube system by Watanabe et al. [14], Triangles by Gorbet et al. [6] and [1]. While theses systems are capable of computing a tangible composition, they do not overlay them in AR. The system presented in this paper adapts this approach and extends it with a flexible AR view to a composing TAR interface with increased robustness against occlusions and perspective distortions.

Concept and Implementation

The system presented in this paper provides increased tracking robustness by combining embedded computational devices and fiducial marker tracking. The embedded devices are used to maintain a virtual counterpart of a tangible composition. Afterwards, this virtual counterpart is aligned to the tangible composition, relying on the tracking information of a single fiducial marker.

Each tangible of the presented system contains a microprocessor, which supports a wireless network connection. Additionally, the microprocessors are connected to conductive elements, integrated into the tangibles' sides. These conductive elements make contact if two tangibles connect and by that, communication between the microprocessors is enabled, such that connections and disconnections can be detected and communicated wirelessly to AR applications. Based on the received connectivity information, these applications maintain a virtual model consisting of the tangibles' virtual counterparts. Similar to the approach by Gorbet et al [6], magnets are centered into the sides with integrated conductors, such that two tangibles can easily be pushed together and afterwards simply be pulled apart.

In a traditional image tracking application, each object needs to be tracked on its own, thus, the number of markers which need to be tracked to properly overlay a tangible composition equals the number of tangibles the composition consists of. By making use of the connectivity information, a virtual model consisting of the composition's virtual counterparts can be arranged beforehand and attached to a single tracked marker. Thus, the number of markers which need to be tracked decreases to a single one, increasing the robustness against occlusions and perspective distortions. Although a single marker suffices to gather the necessitated spatial information, fiducial markers are attached to all of a

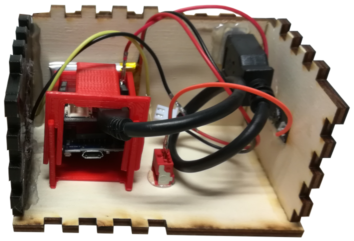

Figure 2: Hardware inside the tangibles. For the microprocessor, the Wemos D1 Mini, which includes a WiFi module, is used. Each block contains a battery shield, a lithium-ion polymer rechargeable battery, and a USB charging cable.

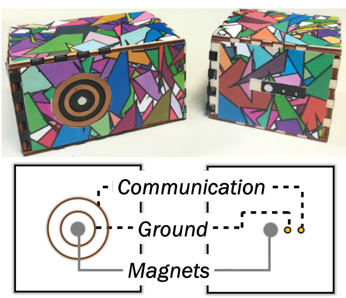

Figure 3: For spatial tracking, the tangibles are augmented with fiducial markers (top). Integrated magnets enable easy connections, resulting in contact between the integrated conductive elements (sketched below).

tangible's sides, such that ideally one marker per composition is always properly visible to the tracking camera.

Five tangible cubic blocks with lengths and depths of 5.6 to 10 cm and heights of 5.6 cm are implemented. Each of them provides two sides with integrated conductive elements, i.e. one side to transmit its own ID and one side to receive incoming IDs. This can, of course, be generalized to bidirectional communication on all of a tangibles sides', however, to evaluate the approach this simple implementation is sufficient. The hardware integrated into each cube can be seen in Figure 2.
The receiving side relies on copper rings as integrated conductors and the transmitting side utilizes pogo pins, which have integrated elastic springs, to ensure a robust connection between the conductive elements of two connected blocks. Consequently, users can identify matching sides visually (rings vs. pins) as well as haptically by the polarity of the integrated magnets. To track the fiducial markers attached to each of a tangible's sides, the system relies on the commonly used library Vuforia. Two implemented cubes, as well as a sketch depicting the connection functionality can be seen in Figure 3.

Instead of directly communicating with the AR application, the tangibles send detected connections and disconnections to a simple node.js web server. This introduces two main advantages. Firstly, the tangibles are capsuled from the AR application, which means that they can be used with several different applications without updating the code on the microprocessors. Secondly, the web server can broadcast received information to several clients, enabling multiple users to interact with the same set of tangibles at the same time, e.g. using multiple HMDs.

The client AR application, which is informed about occurring connections and disconnections by the web server, is implemented using Unity and runs on the Microsoft HoloLens. Each composition is subordinated to a shared parent element within the Unity hierarchy, which means that moving the parent automatically moves the whole composition, while keeping their relative alignment. Thus, only the parent has to be positioned according to a tracked marker. It only considers the spatial information of one marker at the same time, which is randomly chosen out of all currently tracked markers at application start and if the previously tracked marker is lost. The composition itself is maintained in a graph structure, and updated in an event-driven manner on occurring connections and disconnections.

Connecting Elements
A microprocessor continuously writes its own ID and the ID of the corresponding transmitting side to the conductive elements integrated into the transmitting side. Thus, if two cubes connect, the receiving processor detects the connection by receiving the connectivity information of the transmitting tangible. The received information, as well as the own ID and the ID of the own receiving side is then sent to the web server, which broadcasts it to registered clients. The clients now update their virtual model in three steps:
1.) Merge the underlying graphs.
2.) Align the transmitting element's composition to the composition of the receiving element according to the communicated side IDs.
3.) Set the parent of the added composition to the parent of the receiving element's composition.

Disconnecting Elements
During connections, the receiving element continuously receives the connectivity information of the transmitting tangible. Consequently, disconnections can be detected if no information is received anymore. The receiving element then sends its ID and the ID of the side that lost the con-

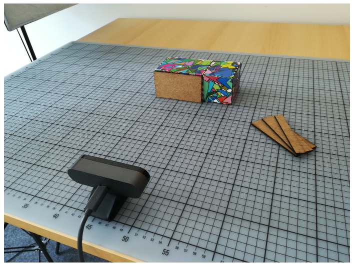

Figure 4: Setup for evaluation of tracking robustness against occlusions. One of two connected blocks is occluded using four different laser-cut wooden rectangles of width 10cm and heights corresponding to 25, 50, 75, and 100% occlusion.

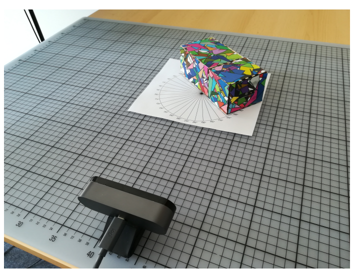

Figure 5: Setup for evaluating tracking robustness against perspective distortions, where the same blocks at the same distance are rotated step-wise by 10 degrees between -70 and +70 degrees.

nection to the web server, which again broadcasts the event to all registered clients. The clients split their virtual model according to the received event as following:
1.) Create a new parent.
2.) Split the underlying graph.
3.) Identify the detached sub-graph (breadth-first-search algorithm) and subordinate it to the newly created parent.

Technical Evaluation

While purely marker-based TAR systems are easy to create with modern AR libraries, these systems suffer in situations where the tangibles are not well visible to the camera. During direct interactions with a composing TAR interface, two frequent scenarios, which can lead to bad visibility, are occlusions and perspective distortions. To showcase the advantages of the presented combination of embedded computation and marker tracking, both of these scenarios are simulated and compared to a purely marker-based system. A Logitech Brio camera is placed with a fixed distance of 25 cm to two tangible blocks. The high-quality webcam is used, because it similarly enhances image tracking for both approaches, preventing tracking errors due to a bad camera resolution. The 25 cm approximate short-range interaction distances in tabletop-scale settings. Two blocks of sizes of 5.6 by 7 cm and 10 by 7 cm are used, both having a height of 5.6 cm. To generalize from the light setting, three different lighting conditions are simulated using spotlights of different intensities and different angles of incidence.

Occlusion: During direct interactions, the most common reason for occlusions are users covering markers with their hands. However, it is unlikely that a user covers a whole composition completely, hence, at least one marker is most likely always properly visible. To simulate this, only the front sides of two attached blocks are shown to the camera and one block is gradually occluded in steps of 25% (see Figure 4). The results are the same for all three lighting conditions: up to 25% occlusion there is no difference between the approaches, as modern tracking libraries like Vuforia can handle small amounts of occlusions. For 50% occlusion the marker-only approach starts flickering, and at 75 and 100% only the hybrid approach can render correctly.

Perspective Distortion: For a single block, distortion is not an issue, since one marker being distorted simultaneously brings another side of the tangible into the camera's view. However, when multiple tangibles are connected, the blocks further away from the camera are only visible in a strongly distorted way, until they are completely occluded. This behavior is analyzed using the same setup as before, but instead of occluding one of the blocks, both are rotated step-wise by 10 degrees between -70 and +70 degrees (see Figure 5). Independent of the lighting condition and angle, our hybrid approach correctly augments both cubes, while the marker-only approach loses track of the distant block after +40/-50 degrees, depending on its size.

Discussion: While TAR systems based only on marker-tracking are a lot easier to implement, the technical evaluation confirms that the hybrid approach improves the robustness both for occlusions and perspective distortions, as soon as a single block in a composition is not properly visible to the tracking device. This is based on the fact that the hybrid approach only requires a single marker per composition to be recognized, while the marker-only approach needs to recognize one marker per block. Even though the technical evaluation is limited by the simplicity of the scenarios, as only one fixed distance is tested and occlusions and perspective distortions are only investigated separately, the clearly observed advantages of the hybrid approach suggest that the results can also be reproduced in more complex situations.

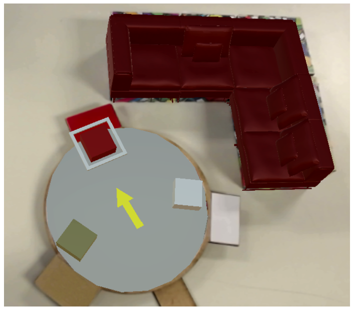

Figure 6: A configured couch as well as the color selection wheel.

Table 1: Likert scale results on 6-point scales.

Question	avg	sd
General		
I like it*	4.67	1.50
it worked well*	4.16	0.98
it was fun to use*	5.34	0.51
Cubes		
cubes too small/big**	3.29	0.49
cubes are handy*	5.14	0.69
Visualization		
I liked the 3D view*	5.28	0.76
AR better than screen*	5.14	1.86

* optimal rate: 6
** optimal rate: 3.5

Use Case: Couch Configuration Tool

The technical evaluation shows that the presented hybrid approach for composing TAR interfaces is more robust towards occlusions and perspective distortions compared to a purely marker-based approach. To test the system in a realistic use case, a couch configuration tool is implemented. For that, the five tangibles are assigned to the most frequent couch elements in online configuration tools: a single-seater, a two-seater, a chaise-longue and two corner-pieces. These elements can be arbitrarily assembled to different couches. For additional customization, marker-tracked TAR configuration wheels are provided. Each wheel is attached to a handle to avoid occlusions, and utilizes a ball bearing to enable the personalization of color, texture and couch type through rotation. Figure 6 shows an AR view of the presented use case.

To gather early user feedback and to investigate whether the implemented use case is suitable to be used in a future comparison study to a purely marker-based approach, seven volunteers (m=3, age 22 to 26) are introduced to the couch configuration tool and configure a couch to their liking. Afterwards, they rate the usability using the SUS [4] and AttrakDiff [8] questionnaires, give personal feedback in an interview and answer Likert scale questions.

A SUS score of 70.5, meaning a 'good' usability [2] and a pragmatic quality, hedonic quality, and attractiveness of 0.82, 1.13, and 1.41, which is positive on all dimensions, were achieved. The results of the Likert scales (see Table 1) show that the volunteers generally liked the system. Even though it suffers from technical limitations, due to the small size of the area, in which markers are recognized and the HoloLens' limited field of view, the system is rated to generally work well and fun to use. The cubes are neither too small nor too big and handy, and furthermore three attendants mention the simplicity and intuitiveness of the configuration system. The attachment of the virtual models to the tangible blocks is stated to be fascinating and preferred over a common computer screen. To conclude, the feedback suggests an overall good usability and user experience.

Conclusion and Future Work

This paper proposes a combination of embedded computation and visual marker tracking for composing TAR interfaces. Conductive elements are integrated into the tangibles' sides, enabling communication between embedded microprocessors. By that, compositions of tangibles can be computed without the need to visually track them and the combined corresponding virtual counterparts can be attached to a single marker. Consequently, the number of tracked markers to properly overlay a tangible composition decreases from one marker per involved tangible element to a single marker per composition of elements. A technical evaluation shows that this approach increases the tracking robustness if a tangible cube in a composition is at least occluded half or perspectively distorted by at least +40/-50 degrees, depending on the block's size. To test this approach in a more realistic setting, a couch configuration tool is built and presented to users, showing promising results regarding usability and user experience.
While currently each tangible supports connections only on two of its sides, it can be generalized to bidirectional communication on all six sides in the future, such that arbitrary three dimensional compositions can be assembled. Additional use cases should be explored, for example board games or interactive movies, and a comparison study between the presented system and an approach solely based on marker tracking can be conducted to investigate differences in the user experience.

REFERENCES

[1] David Anderson, James L Frankel, Joe Marks, Aseem Agarwala, Paul Beardsley, Jessica Hodgins, Darren Leigh, Kathy Ryall, Eddie Sullivan, and Jonathan S Yedidia. 2000. Tangible interaction + graphical interpretation: a new approach to 3D modeling. In *Proceedings of the Conference on Computer Graphics and Interactive Techniques*. 393–402.

[2] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: adding an adjective rating scale. *Journal of Usability Studies* 4, 3 (2009), 114–123.

[3] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2008. Tangible Augmented Reality. *ACM SIGGRAPH ASIA* 7 (2008).

[4] John Brooke and others. 1996. SUS - a quick and dirty usability scale. *Usability Evaluation in Industry* 189, 194 (1996), 4–7.

[5] Audrey Girouard, Erin Treacy Solovey, Leanne M Hirshfield, Stacey Ecott, Orit Shaer, and Robert JK Jacob. 2007. Smart Blocks: a tangible mathematical manipulative. In *Proceedings of the 1st international conference on Tangible and embedded interaction*. ACM, 183–186.

[6] Matthew G Gorbet, Maggie Orth, and Hiroshi Ishii. 1998. Triangles: tangible interface for manipulation and exploration of digital information topography. In *Proceedings of the Conference on Human Factors in Computing Systems*. 49–56.

[7] Taejin Ha and Woontack Woo. 2010. An empirical evaluation of virtual hand techniques for 3D object manipulation in a tangible augmented reality environment. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 91–98.

[8] Marc Hassenzahl, Michael Burmester, and Franz Koller. 2003. AttrakDiff: ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch & Computer 2003*. Springer, 187–196.

[9] Hirokazu Kato, Keihachiro Tachibana, Masaaki Tanabe, Takeaki Nakajima, and Yumiko Fukuda. 2003. A city-planning system based on augmented reality with a tangible interface. In *Proceedings of the Conference on Mixed and Augmented Reality*. 340–341.

[10] Wonwoo Lee, Woontack Woo, and Jongweon Lee. 2005. Tarboard: Tangible augmented reality system for table-top game environment. In *2nd International Workshop on Pervasive Gaming Applications, PerGames*, Vol. 5.

[11] Fotis Liarokapis, Louis Macan, Gary Malone, Genaro Rebolledo-Mendez, and Sara De Freitas. 2009. Multimodal augmented reality tangible gaming. *The Visual Computer* 25, 12 (2009), 1109.

[12] Jannick P Rolland, Larry D Davis, and Yohan Baillot. 2001. A survey of tracking technologies for virtual environments. In *Fundamentals of wearable computers and augmented reality*. CRC Press, 83–128.

[13] Christiane Ulbricht and Dieter Schmalstieg. 2003. *Tangible augmented reality for computer games*. Citeseer.

[14] Ryoichi Watanabe, Yuichi Itoh, Masatsugu Asai, Yoshifumi Kitamura, Fumio Kishino, and Hideo Kikuchi. 2004. The soul of ActiveCube: implementing a flexible, multimodal, three-dimensional spatial tangible interface. *Computers in Entertainment* 2, 4 (2004), 15–15.