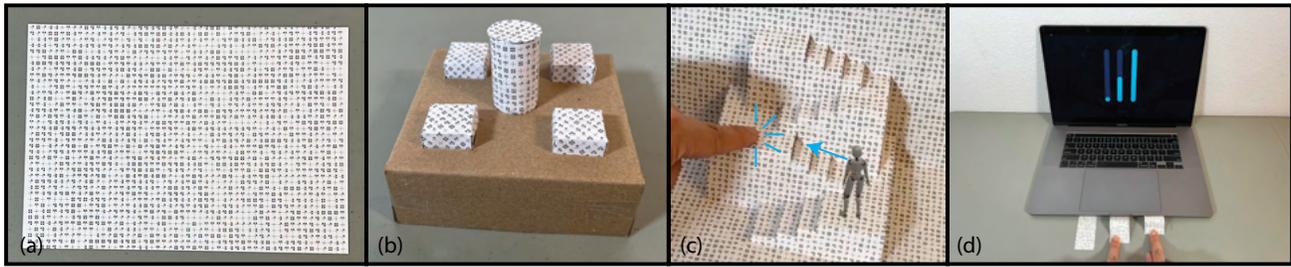


# ReactFold: Towards Camera-based Tangible Interaction on Passive Paper Artifacts

Chang Xiao  
cxiao@adobe.com  
Adobe Research  
San Jose, California, USA



**Figure 1: ReactFold system and its use cases:** Our marker system can be printed on standard paper using a regular printer, enabling interaction with paper-based artifacts and using a single camera for detection. (a) A paper with printed markers. (b) A tangible widget assembled using our system. It contains four push buttons and one rotary dial. (c) A kirigami AR application. An AR avatar can stand on the kirigami stairs, aware of its 3D geometry. A user tapping on the kirigami will trigger the avatar to walk upstairs to that location. (d) A set of deformable strips interface. Our system can track the deformation strength and convert it to an input signal.

## ABSTRACT

In the current era of Extended Reality (XR), tangible interactions play a crucial role in enabling users to achieve more precise and intuitive control. However, most tangible interfaces are built upon electronic devices, offering a fixed interface that is not customizable across different applications. In this work, we explore the potential of utilizing a universally accessible material—paper, to construct tangible interfaces. Our proposed system requires only a single camera for input event detection, a common hardware feature available in almost all XR headsets and mobile devices. We propose a marker-based system that enables camera-based tangible interactions with various paper-based artifacts, including flat sheets, paper attached to everyday objects, and pop-up cards. We examine a set of detectable events, such as occlusion, appearance changes, and deformation. Additionally, we present a range of paper-based interfaces and applications enabled by our system, opening the door to a wealth of new interaction possibilities with paper artifacts. Our system is also validated by extensive experiments, which show robust detection across different paper forms.

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 ACM 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).

TEI '25, 04-07 March 2025, Bordeaux, France

© 2025 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM... \$15.00

<https://doi.org/10.1145/3689050.3704926>

## CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques; Interaction devices.**

## KEYWORDS

Augmented Reality, Mixed Reality, Origami, Deformable Interfaces, Tangible Interfaces

## ACM Reference Format:

Chang Xiao. 2025. ReactFold: Towards Camera-based Tangible Interaction on Passive Paper Artifacts. In *Proceedings of 19th International Conference on Tangible Embedded and Embodied Interaction (TEI '25)*. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3689050.3704926>

## 1 INTRODUCTION

From its remarkable invention approximately 2000 years ago, paper has evolved as an extraordinary cultural milestone, maintaining its importance even in the modern era of information technology. Through innovative yet simple folding techniques, people have transformed this basic material into intricate origami works, dynamic pop-up books, and practical cardboard designs. Its beauty is not solely in its versatility but also in its accessibility: all one needs is a single sheet and the desire to create.

In today's digital era, paper is not merely a medium for flat communication or a folded structure for visual display. Researchers have begun to fuse paper with contemporary sensing techniques to make it interactive. Over recent years, there has been a growing interest in developing tangible interfaces using paper artifacts. Current methods for this fusion mainly center around circuit-based

integration [5, 10, 18, 19, 37, 50, 57]. These methods have significantly expanded the potential of paper as a tangible interface, opening the door for a range of new interactive applications. However, integrating circuits within paper not only requires additional effort to embed the circuitry but also demands expert knowledge for designing circuits tailored to specific interfaces.

In this work, we present ReactFold, a camera-based system that enables interaction detection on various types of passive paper artifacts, including flat sheets, folded paper, pop-up book, and paper attached to other objects in daily life. Our system simply requires our marker pattern to be printed on standard paper. Using our computer vision detection pipeline, we can monitor alterations on the paper artifacts and interpret these changes as interaction events. This method is compatible with any application equipped with a camera, for example, it can easily integrate as a tangible interface within mobile device or Mixed Reality headsets, thereby unlocking the potential to bridge the physical and digital worlds. Using passive paper as a tangible input interface presents several distinct advantages over existing prototyping methods such as 3D printing. First, paper is a widely accessible material, available nearly everywhere and to everyone. Second, paper exhibits an exceptionally high level of flexibility and expressivity, allowing for a vast array of creative designs and real-time manipulations. Third, in most cases, crafting the desired paper artifacts for average users is a more approachable process compared to the intricate steps often associated with 3D modeling or programming.

The contributions of this paper can be summarized as follows:

- We improve the current marker system and detection pipeline to support robust marker detection across various paper forms using a single camera—a capability not fully achievable by existing methods.
- We summarize the design space of interaction primitives for paper-based artifacts, accompanied by new algorithms designed for our system that recognize different interaction events.
- We explore various paper-based interfaces and applications enabled by our system, demonstrating its capabilities and providing a platform for researchers to rapidly prototype new paper-based interfaces.

## 2 RELATED WORK

**Electronic Paper Interface** Utilizing paper as a tangible media and enabling interactivity for it has been extensively studied by the HCI community, as paper is a clean, flexible and recyclable material. One major approach as demonstrated by many works is try to embed electronics into paper itself. For example, Gallant et al. [15] first introduced the idea of a foldable interface to prototype flexible display interactions using only a simple IR webcam and an LCD screen. *IllumiPaper* describe illuminated, digitally controlled papers that enable visual feedback for pen-paper interaction. *Tessella* [7] introduce an interactive origami light that blends traditional craft and soft-circuit techniques to inspire creativity through user interaction. *PrintGami* [10] presents a method combining 3D printing and origami to integrate paper circuits into interactive objects. These work can be categorized in a more generalized area of Origami Electronics [11, 36, 53], which has also been extensively investigated in

the robotics community [1, 32, 40]. Incorporating electronics into paper significantly enhances its interactivity, transforming ordinary paper into a customizable interface. A comprehensive survey on this research line can be found in Signer et al. [43]. However, adding electronics not only makes these devices more complex and challenging for average users to create or utilize, but it also compromises the recyclability and flexibility of paper as a material.

**Camera-based Paper Interface** Consequently, some studies have employed camera-based techniques, enabling interaction with regular, passive papers [2, 6, 8, 29, 39, 41, 56]. Our system also falls into this category. Some early works have identified several useful markers for AR purposes, such as ARTag [13], AprilTag [38], ArUco [16] and Anoto [20]. Building upon these markers, researchers have proposed different camera-based paper interface designs. For instance, *ARcadia* [29] propose a toolkit that combines visual markers and domain-aware programming tools for rapid prototyping of paper-based tangible interfaces. ModelCraft [44] introduces a paper-based method for transferring annotations from physical paper folding to digital CAD models using the Anoto pattern. Zheng et al. [56] and *DynaTags* [41] utilize ArUco marker and leverage paper's flexibility and foldability to facilitate tangible interaction events triggered by marker's visibility. These creatively designed paper interfaces offer passive interaction opportunities across diverse folding designs.

Another line of research [24, 31, 35, 45, 47, 49, 51] explored the concept of projecting digital content onto paper surfaces and simultaneously detecting paper-based events, such as movement or deformation. To support this, several useful dot marker pattern has been proposed like Uchiyama et al. [49] and Narita et al. [35].

While many existing camera-based systems show promise in detecting specific interaction events, they often fail to accommodate other different interaction events across physical variations of paper for various reasons. For instance, the ArUco marker [16] and its variants [13, 38, 55] can only be reliably detected on flat surface. These markers cannot provide accurate geometry information if the paper is largely deformed or if only a portion of the marker is visible, as can frequently occur with folded paper. DDCM [35], on the other hand, can detect deformed or curved surfaces but it faces challenges in robustly processing disjointed pieces, as the marker system's layout is sparse due to the dot cluster design. Similarly, the recent work NeuralMarker [25] provides dense correspondence mapping on user-provided images, but it is unclear how this method can be adapted to folded paper, as it requires the image to appear as a single piece. Zheng et al. [56] demonstrate strong adaption to different forms and input events. However, the design is restricted to pre-defined forms, as it requires the paper to be manipulated in such a way that an entire ArUco marker is visible to trigger an event. In addition, most marker-based methods require a specific layout for each individual interaction purpose. This means that the user needs to carefully design the layout to adapt to the proposed interface. In contrast, our method can utilize the same pattern for different applications. This allows us to use the same set of materials to build the desired applications without needing to reprint.

Here we provide a summary of the detectable events and paper forms that various camera-based methods can support in Table 1. In this comparison, all methods are assumed to be implemented

Methods	Paper Form			Detectable Event				Unified Pattern
	FS	CS	DS	O	AC	OT	DT	
ArUCo Variants [13, 16, 38, 55]	✓			✓		✓		
Anoto [20]	✓			✓				✓
Zhu et al. [58]	✓	✓	✓	✓				✓
DRDM [48]	✓	✓				✓	✓	✓
Origami Guru [52]	✓	✓	✓		✓			
Gustov et al. [21]	✓	✓		✓		✓	✓	✓
DynaTag [41]	✓			✓	✓	✓		
NeuralMarker [25]	✓	✓				✓	✓	✓
DDCM [35]	✓	✓		✓		✓	✓	✓
Zheng et al. [56]	✓		✓	✓	✓	✓		
<b>ReactFold</b>	✓	✓	✓	✓	✓	✓	✓	✓

**Table 1: A capability comparison among existing camera-based methods that enable passive paper-based interactions. FS: Flat Surface. CS: Curved Surface. DS: Disjointed Surface (e.g., in folded paper or kirigami). O: Occlusion. AC: (Discrete) Appearance Change Detection. OT: Object Tracking. DT: (Continuous) Deformation Tracking. Unified Pattern indicates whether a method can employ a single design for all forms and interaction events. A check mark indicates the ability of a method to be integrated with the paper form or input event, or implemented with a unified pattern. For a fair comparison, we assume that all methods utilize only a single RGB camera for detection. Notably, only our method is capable of accommodating all input events across all forms of paper.**

using regular paper printed with patterns or markers, and utilize a standard RGB camera for detection. Notably, our method is the only one capable of supporting all paper forms and input events.

**Deformable Interface** Besides paper interface, a wider scope of research that explore deformable materials, such as rubber or clay, to design input interface has also gain a lot of interests. Some examples includes MetaSense [17], MagnetIO [34] and Flexpad [46]. We refer the reader to Boem et al. [3] for a more comprehensive survey. The materials utilized in these methods are generally less accessible than paper and often require embedded circuits. In contrast, paper stands as a readily available, cost-effective, and versatile medium, thus making it a favorable choice for a wide array of interactive applications.

### 3 SYSTEM AND METHODS

In this section, we explore the design space for tangible interactions with paper artifacts under various forms that paper can take. We then discuss the structure of our proposed markers and the algorithm used to detect interaction events.

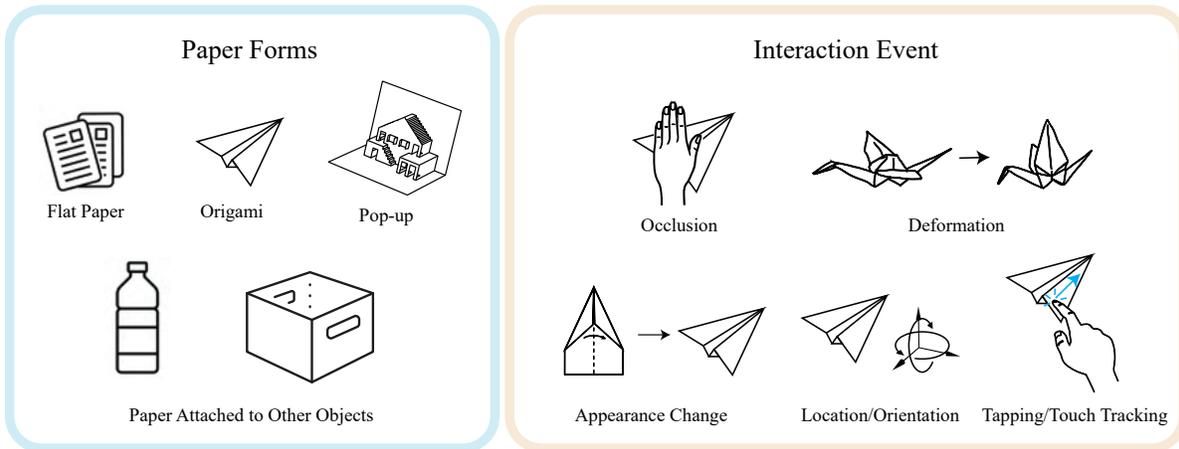
#### 3.1 Design Space of Interactive Paper

Paper comes in various forms, from common flat sheets and cardboard used in printing and packaging, to intricate forms like origami and kirigami, which involve folding and cutting to create specific shapes. These forms offer a plethora of possibilities for tangible interactions. For example, a flat sheet of paper can be used to detect taps or track touch, while origami and kirigami structures enable interactions through deformation and appearance changes. Examining the different forms of paper and the design space of tangible interactions is key to unlocking the full potential of interactive paper artifacts. In Fig. 2, we provide an overview of the various forms of paper along with the respective detectable interaction events,

summarized from previous work on leveraging paper as input interfaces [7, 15, 29, 41, 56]. As discussed in Sec. 2, the majority of existing methods can only accommodate a limited number of forms and events. Our objective is to create a system capable of addressing a wider array of cases in a unified framework, thereby enabling users to construct diverse and functional tangible interfaces using paper artifacts. Note that the design space illustrated here is not an exhaustive representation of all possible paper-based interactions. Instead, we focus on the interactions achievable through passive camera-based sensing. Next, we will discuss how our system accommodates these proposed interaction events across various paper forms.

#### 3.2 Design Considerations

As elaborated in Sec. 2, while numerous marker-based systems show potential in detecting interaction events, they often struggle to adapt to all interaction events given the diverse physical forms of paper. To effectively and robustly support the detection of these conditions, a marker system must meet two primary criteria. Firstly, the “atomic unit” of the marker system should be sufficiently small for a dense layout, accommodating challenges such as occlusion, curved surfaces, and disjointed surfaces. Secondly, each detected marker must possess a distinct identifier, allowing for its remapping to the original pattern for subsequent 3D geometry reconstruction. Previous marker-based systems were usually designed with the purpose of robust detection on planar surfaces. Thus, their “atomic unit” is typically larger to encourage a higher detection rate. Additionally, they were usually designed for locating purposes rather than for interaction purposes and did not account for all the scenarios that might occur with paper artifacts. Fig. 3 illustrates several scenarios and explains why marker systems like ArUco fell short in these aspects.



**Figure 2: Paper Forms and Interaction Events.** In this context, we define “occlusion” as any behavior that obscures a portion of the paper, irrespective of whether physical contact occurs. Both “tapping” and “touch tracking” are recognized only when physical contact is established. “Deformation” involves the continuous monitoring of changes in the paper’s physical geometry, while “appearance change” refers to the identification of switches between two distinct states.

Upon analyzing existing marker systems, we identified two that partially meet our criteria and could potentially be combined to create a new system satisfying both conditions. The first, DDCM [35], utilizes dot clusters as the “atomic unit”, which is a simple pattern and can be detected robustly. This system can generate many unique identifiers by varying the dot cluster patterns. However, printing these clusters densely on paper causes overlaps, which could lead to confusion in the detection algorithm. The second system, Anoto [20], is a commercial product primarily used for digital pen tracking. It features a grid pattern with offset dots, allowing for robust detection even when printed in small sizes. Yet, it provides only a limited number of unique identifiers due to the constrained variability in grid and dot layouts. Additionally, since it is a commercialized system, details about the specific detection algorithm and how the system adapts to curved or disjointed paper surfaces are also unclear.

Our system is mainly inspired by these two marker systems; we combine their designs, use both cross shapes and dot clusters, to create a new marker system that can be printed in small form, enables robust detection, and provides a rich number of distinct identifiers.

### 3.3 ReactFold Markers

To meet the outlined requirements, we introduce ReactFold Marker (RM), a marker system engineered with dense marker arrangement to support robust detection and correspondence mapping. As illustrated in Fig. 4-a, an RM incorporates a cross shape and dots arranged in its surroundings; the simplicity of RMs’ shapes allows for detection through basic image filters. To guarantee rotational invariance and ensure the marker remains unchanged under any transformation, each RM adopts one of six distinct types of layouts (Fig. 4-b), wherein the dots can occupy different quadrants of the cross. An RM pattern is formed by a matrix of RMs, and each RM is

selected from one of the six types. In practice, we randomly choose the RM type to construct an RM pattern of any desired size.

Similar to prevalent image feature detection methods such as SIFT [33], our system adopts a two-stage pipeline: first to detect candidate RMs, then compute an identifier for each candidate RM to match with the original design.

**Detecting Individual RM** The precise location and type of the RM can be inferred using simple image filtering techniques. Upon receiving an image of paper artifacts with the RM pattern, we initiate the detection by adaptively binarizing the captured image. In our implementation, the local threshold value is determined from a mean of the pixel values in the neighborhood. Following this, we apply the Harris Corner Detector [22] to extract all center of the crosses. Next, we crop the region around each cross in the identified RM and use OpenCV’s SimpleBlobDetector to locate the dots, determining the RM type based on the arrangement of these dots.

**Compute Identifier** To this end, both the number and positions of the RMs have been recognized, setting the stage for the next step: extracting unique identifiers for each RM. This phase is vital for mapping the detected RMs to their respective positions in the original pattern. Although our system draws inspiration from DDCM and Anoto, it still requires a completely new algorithm to compute the identifier and correspondence.

Given that a single RM manifests in only six variations, solely relying on the type information to determine its exact original position is not reliable, owing to the possibility of detection errors and ambiguous matches. Hence, we incorporate information from the focal RM and its eight neighboring RMs to craft a distinctive identifier. This identifier is represented as a 9-dimensional vector, with the initial entry denoting the focal RM’s type ID (0 to 5) and the subsequent entries encoding the type IDs of the eight neighboring RMs, arranged in a circular order starting with any neighbor. As we



**Figure 3:** Here we illustrate various scenarios where the popular ArUco marker system falls short. A red rectangle highlights successfully identified markers. (a) Within a folded cube setup, detection is possible only when the entire marker is visible and does not span the fold or crease in the structure. (b) Although using smaller markers promotes a denser layout, this substantially reduces the detection rate due to inadequate resolution. (c) The system struggles to identify markers on a curved surface accurately. Even if a marker is recognized, the extracted borders are incorrect, as the detection algorithm presumes a flat surface. (d) Similarly, using smaller markers leads to inconsistent detection outcomes, rendering them unsuitable for interaction purposes.

will consider all cyclic groups of such vector during matching, this allows us to accommodate any marker orientations that may be present in an image. In instances where neighbors are absent due to geometric borders or detection failures, a value of  $-1$  is assigned to the respective entry. The boarder of the paper folding is detected through the Hough Line transform [26]. Examples of computing identifiers can be observed in Fig. 4.

**Matching Detected RM** To match the detected RM with its counterpart in the original pattern, we propose a distance metric to measure the similarity between them. The distance metric  $D$  is defined as:

$$D(\mathbf{x}, \mathbf{y}) = \min_i H[(x_1, x_2, \dots, x_9), (y_1, C_i y_{\setminus 1})], \quad (1)$$

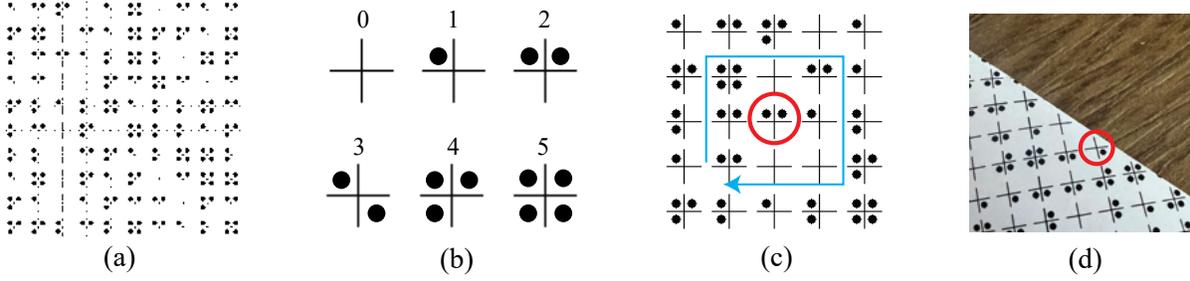
where  $(x_1, \dots, x_9)$  and  $(y_1, \dots, y_9)$  represent the vector representations of the detected RM ( $\mathbf{x}$ ) and a RM in original pattern ( $\mathbf{y}$ ) respectively. Here,  $H$  denotes the weighted Hamming distance metric, where the first entry has a weight of 4 and others are weighted equally at 1. This enables us to place greater emphasis on the focal RM.  $C_i y_{\setminus 1}$  is the  $i$ -th cyclic permutation of  $(y_2, \dots, y_9)$ , accounting

for the indeterminate neighbor orientation encountered in practical scenarios when an RM is captured by the camera.

To identify the correct match for a detected RM, we denote this RM as  $\mathbf{v} = (v_1, v_2, \dots, v_9)$  and its most similar counterpart in the original RM pattern as  $\mathbf{w}^i$ . Ideally, The index  $i$  can be determined simply through  $\arg \min_i D(\mathbf{v}, \mathbf{w}^i)$ . Nevertheless, there exist scenarios where multiple  $\mathbf{w}^i$  yield the smallest value of  $D$ . This might be due to detection errors or complications at the edges of the paper (note that we assign a value of  $-1$  to indicate detection failures and border issues). Another potential cause is that multiple RMs in the original design share the same identifier following random generation. The latter case is quite uncommon; our experiment indicates that only about 0.5% of RMs might possess the same identifier within a randomly generated RM pattern on an A4 paper. To maximize the number of successful matches and ensure their accuracy, we employ the following algorithm: Initially, if a detected RM  $\mathbf{v}$  has a unique counterpart in the original pattern  $\mathbf{w}$  and  $D(\mathbf{v}, \mathbf{w}) \leq 4$ , we accept  $\mathbf{v}$  and  $\mathbf{w}$  as a valid match. Subsequently, for any remaining  $\mathbf{v}$  associated with multiple  $\mathbf{w}^i$  with identical  $D$  values, we inspect each  $\mathbf{v}$  and  $\mathbf{w}$  pairs to check if at least one valid match exists between their respective neighboring RMs. If this condition and  $D(\mathbf{v}, \mathbf{w}^i) \leq 4$  are met, we accept them as a valid match. Any remaining RMs with unresolved matches are omitted. As demonstrated in our evaluation in Sec. 5, this matching protocol can achieve a significantly higher match rate across various paper forms compared to other existing methods.

**Size of RMs** We have outlined the detection and matching processes for RMs. Yet, selecting the optimal dimensions of the RMs remain undiscussed. The dimensions are primarily influenced by two factors: the size of the cross and the size of the dots. To determine the most suitable dimensions for RMs, we conducted a synthetic experiment wherein we created a series of virtual RM-patterned A4 papers with different RM dimensions, and rendered them at random camera angles from a distance of 80cm using Blender [14]. This experiment also included random cropping to simulate the effects of folding or cutting that are commonly observed in paper artifacts.

We examined cross lengths ranging from 1mm to 7mm, and the ratio of dot diameter to cross length between 0.1 and 0.4. In all cases, the dots were centrally located in the cross quadrants. Our findings revealed that cross lengths larger than 3mm maintain high overall detection rates, exceeding 97%. Conversely, lengths below 3mm led to a declining trend in detection rates, mainly because the details become less recognizable from the camera’s perspective. Therefore, we opted for an RM size of 3mm × 3mm to facilitate a dense layout. Furthermore, we determined that a dot diameter ratio of 0.3 (equivalent to 0.45mm) offered the best performance across various settings. We incorporated a 1mm margin between adjacent RMs to also encourage a denser layout without interfering the detection. Consequently, a standard A4 paper (210mm × 297mm) can accommodate a layout of 52 × 74 RMs. Additionally, our system imposes no restrictions on the number of pieces of paper that can be used in a single interactive scene. Indeed, we permit the simultaneous presentation of multiple paper artifacts, provided they do not share the same RM patterns.



**Figure 4: An overview of ReactFold Marker (RM).** (a) Pattern consists of randomly generated RM. (b) The 6 types of RM. These RM is unique under different orientation. The number above each RM indicates its type ID. (c) The identifier of the RM is represented as a 9-dimensional vector, which the first entry is the focal RM type and the following entries are the neighbor RMs arranged in any circular order. In this example, the identifier of red circle RM can be calculated as  $(2, 2, 2, 5, 0, 2, 1, 0, 0)$  (d) When compute identifier in real photo, the neighbor RM may missing due to folding design. We assign  $-1$  to the respective entry. In the red circle region, the RM’s identifier is  $(1, 2, 2, 0, -1, -1, -1, -1, 5, 4)$ .

### 3.4 Interaction Event Detection

There are two primary categories of interaction event in Sec. 3.1: visibility-based and geometry-based. The first one involves changes only of the markers’ visibility against the camera, whereas the second involves changes of the paper artifact’s geometry and location. Different techniques are required for detecting interaction events in these categories. In the following, we describe our proposed techniques for detecting these different interaction events.

**3.4.1 Visibility-based Interaction.** Visibility-based interaction occurs when a user causes changes to the artifacts, which solely affect the detectability of the RMs. To support this, we introduce the ReactFold Visibility Map (RVM), representing the identified part of the original RM pattern (Fig. 5-a). Formally, it is defined as

$$M_{ij} = \begin{cases} 1, & \text{if } M_{ij} \text{ has at least 1 detected RM,} \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

where  $M_{ij}$  is a overlapping region with  $3 \times 3$  RMs in the original RM pattern. The RVM identifies whether a specific RM region is successfully recognized. Ideally, a camera view containing the entire RM paper should have  $M_{ij} = 1$  for all  $i, j$ . As each  $M$  cell contains 9 RMs, this could provide redundancy to ensure a region is identified as visible even if some RMs are not detected. Recall that the physical size of our RM is  $3mm$ . This could result in a smallest detectable region of approximately  $1cm \times 1cm$ , which is suitable for most user interactions. Fig. 5-a demonstrate the process of computing the RVM of a paper cootie catcher.

**Occlusion** The simplest visibility-based interaction event to identify is occlusion. By monitoring the RVM, it is possible to detect if a particular region transitions from 1 to 0. If this occurs, the transition could be identified as an occlusion event. This can be useful in recognizing events such as a user’s hand hovering over a specific area on the paper artifacts.

**Touch Detection** We could also extend this framework to detect touch event. To determine if a touch event has occurred at location  $x$ , we first identify two circular regions centered at  $x$ ,  $M_x^a$  and  $M_x^b$ , with diameters  $a$  and  $b$  respectively. A touch event at  $x$  is triggered if i) no RMs are detected in  $M_x^a$ , and ii) at least 70% of the RVM in  $M_x^b$

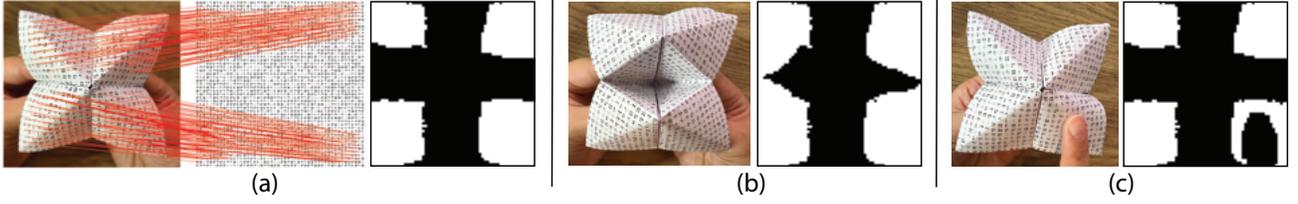
(excluding the area of  $M_x^a$ ) is visible. The logic behind this algorithm is that since a finger is long but only touches at the fingertip, the algorithm anticipates that the fingertip will occlude all RMs within the area of the fingertip ( $M_x^a$ ). However, due to the finger’s length, the body of the finger will also occlude some of the surrounding area ( $M_x^b$ ). This observation is the basis for setting a 70% visibility threshold in the outer region. In practice, we set  $a = 1.3cm$  and  $b = 3cm$  to accommodate the average human finger size, but these values can be adjusted if using other objects (e.g., a pen) to trigger touch.

**Appearance Change** Detecting appearance changes can also be achieved by simply monitoring changes in the RVM, as altering the appearance typically involves activating new parts of the RVM and deactivating previous ones, as shown in Fig. 5-b. Therefore, the most straightforward method to identify changes in appearance is to designate a specific region as a trigger. If the RVM value for that region changes, it indicates a change in appearance, provided that no touch event is detected in that region.

**3.4.2 Geometry-based Interaction.** In geometry-based interactions, we adopt a distinct approach for detection as they require an understanding of the artifact’s 3D geometry before proceeding with any interaction. Thus, the initial step involves reconstructing the 3D geometry of the given artifact. To facilitate this, we instruct the user to capture around 10 different angled photographs of the artifact intended for 3D interaction. Subsequently, we extract the pixel coordinates of all matched RMs from each image. We then follow the standard Structure-from-Motion (SfM) procedure to acquire the artifact’s 3D point cloud [4], that is, by solving the minimization problem

$$\min_{X_1, X_2, X_3, \dots} \sum_{i=1}^m \sum_{j=1}^n \|x_{ij} - Proj(X_j, M_i)\|^2. \quad (3)$$

Here,  $m$  represents the number of photographs used for reconstruction,  $n$  is the total count of RMs in the original pattern,  $x_{ij}$  denotes the observed pixel location of the  $j$ -th RM in the  $i$ -th image,  $X_j$  is the  $j$ -th RM’s 3D world coordinate, and  $M_i$  is the  $i$ -th image’s projection matrix. In our implementation, we use the OpenCV SfM



**Figure 5: Illustration of RVM. (a) Left: the input image of an RM-patterned cootie catcher. Middle: the matched RM in the original design. Right: the computed RVM, where the black region indicates 0 and the white regions indicate 1. (b) When the cootie catcher is "opened," its corresponding RVM changes because there are more matched RMs. (c) When a finger occludes a certain part of the cootie catcher, its corresponding part in the RVM turns to 0.**

module, substituting the feature-matching step with our matched RMs.

**6-DoF Tracking** 6-DoF tracking involves computing the artifact’s location and rotation in 3D world coordinates from real-time captured frames. This process forms the foundation for numerous AR applications that require the precise geometry of paper artifacts. When the point cloud of all RMs is known, calculating the world coordinates is equivalent to optimizing the translation  $t$  and rotation matrices  $R$  given the camera’s projection matrix  $P$  [23]. Thus, the 6-DoF problem for a specific frame can be reformulated as another minimization problem:

$$\min_{R,t} \sum_{i=1}^n \|x_i - P[R|t]X_i\|^2, \quad (4)$$

where  $P$  denotes the camera projection matrix which only depends on the camera type and was previously computed by solving Eq. (3). Additionally,  $n$  represents the count of matched RM in the current frame,  $x_i$  denotes the observed pixel coordinates of the matched RM, and  $X_i$  is the 3D world coordinates of the RM retrieved from the reconstructed point cloud.

**Deformation** Another unique interactive event that paper enables is its ability to deform. Similar to other marker-based systems [35, 49], our method could also detect the deformation of the artifacts at any time. In order to use the deformation process as the input signal, we need a metric to quantify the degree of deformation. Measuring the degree of deformation is still an open problem [42] and it highly depends on the application domain and the deformable object (e.g., face). In our case, as a proof-of-concept, we simply use the Euclidean distance between the deformed artifact and its original shape as the signal of such input interface. More specifically, we use the reprojection error from optimizing Eq. (4) as the quantity to measure the deformation strength  $d = \sum_{i=1}^n \|x_i - P[R|t]X_i\|^2$ . This simple calculation could result in a 1D input signal controlled by the degree of deformation of a given artifact.

## 4 PROPOSED INTERFACES AND APPLICATIONS

In this section, we showcase a variety of interfaces and applications that leverage the previously described detectable events to develop innovative interactive experience that come to life. All of these applications can be realized in AR/VR environments where the user is wearing AR/VR glasses or using a handheld mobile device.

Some of the applications are also practical for desktop usage when a webcam is in operation.

**Tangible Input Widgets** The great versatility of paper allows it to mimic physical input widgets commonly encountered in everyday life. For example, we can craft various widgets such as trackpads, push buttons, switches, and dials using paper. A plain flat sheet can be transformed into a trackpad with the aid of our touch tracking method. Likewise, push buttons, switches, and dials can be effectively realized through our 6-DoF tracking algorithm. These widgets can be built in less than 10 minutes by following the online tutorials step by step even for user without crafting experience.

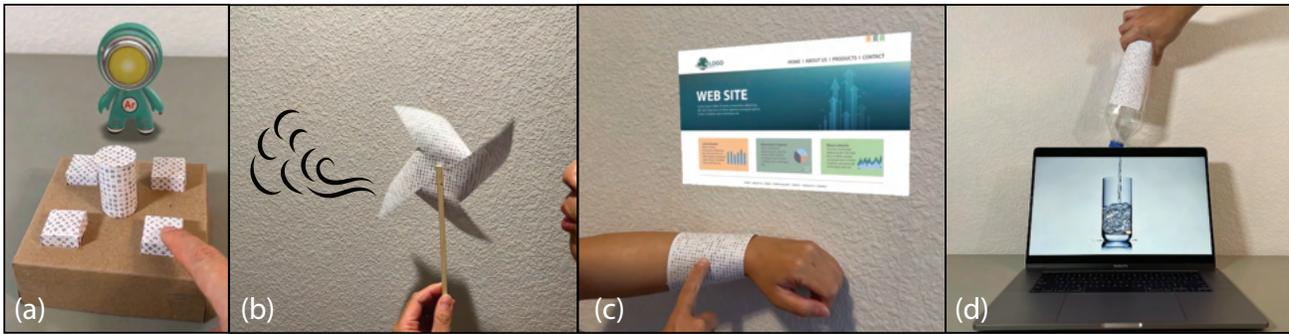
Furthermore, we introduce a new deformable slider that is simpler to create and utilizes deformation tracking to capture a continuous signal. As depicted in Fig. 1-d, this deformable slider is composed of a series of paper strips. When a user slides their finger across a strip, the paper deforms, and the magnitude of this deformation can be detected. This feature can be particularly useful for simulating audio interfaces, such as faders, to produce real-time audio feedback.

These widgets can also be applied in other scenarios. For instance, they can support the control of AR objects, enabling actions such as scaling objects, adjusting volume, or other controls like triggering animation, adding new objects, or initiating actions in an AR game (Fig. 6-a). Moreover, considering the growing trend of AR glasses being launched or planned for future release, researchers anticipate that AR glasses may soon become a daily wear, serving as personal computing hubs [9]. In such scenarios, our paper-based tangible widgets could function as smart home controllers, assisting users in developing customized interfaces for smart home management.

**Enhancing Interactivity of Daily Objects** In addition to creating artifacts purely from paper, a unique feature of our method is the capability to attach patterned paper to a variety of everyday objects, thereby making them interactive. For instance, one can attach RM-patterned paper to the grip of a cup, thereby transforming it into a control interface capable of scrolling through a website on a virtual AR screen.

Furthermore, everyday objects often exhibit unique interactive modalities, such as the ability to deform. We can apply our paper to deformable objects, like a plastic water bottle, to craft functional AR experiences that are activated through deformation triggers, as demonstrated in Fig. 6-d.

Paper can also be transformed into wearable interfaces, such as a wristband. Fig. 6-c shows an example in an AR scenario, assuming



**Figure 6: Gallery of Applications.** (a) Tangible console made of paper. It consists of four push buttons and one rotary dial. Pressing a button can select a specific AR avatar, and rotating the dial can change its scale. (b) An RM pinwheel. Blowing on it to spin the pinwheel can trigger an AR wind effect, and the strength of the wind is determined by the spin speed of the pinwheel. (c) A wristband made of paper. Swiping on this wristband allows the user to scroll through a website rendered in AR. (d) Paper attached to a water bottle. Squeezing the water bottle will cause deformation and trigger an animation of pouring water on the display.

AR glasses are worn by the user. The user can tap a specific point on the wristband to open a browser, and swipe on the wristband to scroll through the website.

**Interactive Pop-up Card** Pop-up cards or kirigami are the fascinating works created through folding and cutting of paper. These paper artifacts are widely used by people to share creativity or as unique gifts. Moreover, they display a rich variety of mechanical movements, making them excellent candidates for interactive experiences. In Fig. 1-c, we constructed a staircase kirigami model using RM-patterned paper. When our system recognizes that the kirigami has been opened, an AR avatar appears at the bottom of the staircase. Using the algorithm described in Sec. 3.4.2, our system is aware of the detailed 3D structure of the entire kirigami. Therefore, users can further interact with the scene by tapping on the desired staircase location to prompt the avatar to move to that spot through the stairs in 3D. This experience seamlessly connects digital assets with customizable physical content, offering a more engaging and immersive AR experience.

Similarly, by following the numerous crafting tutorials online, users can create a diverse set of paper pop-ups for various interactive experiences.

**Paper-based Sensor** Our system can also be expanded to create paper-based sensors. For example, a pinwheel, which consists of a paper wheel affixed to a stick by a pin at its axle, is designed to spin when exposed to breath or wind. By utilizing our system to monitor the pinwheel’s rotational speed, we can transform the design into an anemometer for gauging wind speed or a flow meter for evaluating water flow velocity.

In an AR context, this mechanism can be leveraged to craft enriched interactive experiences. As shown in Fig. 6-b, users can blow on the pinwheel to make it spin, triggering a visual display of AR wind originating from the pinwheel. The intensity the wind correspond to the spin speed of the pinwheel.

## 5 EVALUATION

The objective of our system evaluation is to address two questions: 1) What is the overall detection rate of RM markers in various paper forms, and how does this performance compare to existing marker-based methods? 2) How accurate is the detection of the individual interaction events defined in Sec. 3.4?

### 5.1 Detection Accuracy of Markers under Different Paper Forms

To address the first question, we conducted experiments using three distinct types of paper artifacts, each exhibiting different levels of complexity: a flat sheet, paper wrapped around a cylinder to represent curved surface, and a staircase kirigami structure to represent disjointed surface. All artifacts were created using A4 paper. For a comparative analysis, we also used two other methods to create the same paper artifacts: ArUco [16] and DDCM [35]. With the ArUco marker, we implemented three different configurations: a regular version ( $3\text{cm} \times 3\text{cm}$ ), a small version ( $1\text{cm} \times 1\text{cm}$ ), and a tiny version ( $3\text{mm} \times 3\text{mm}$ ) which has the same size as an RM. For DDCM, we adhered to the configurations and detection methods described in the original paper, since a denser layout would lead to a reduced detection rate [35]. This setup allowed for a comprehensive comparison across five different configurations, including our method.

For each specific configuration, the artifact was positioned on a motorized rotating stand situated at a predetermined location. The camera was positioned 60 cm away from the artifact, a distance chosen to reflect a typical scenario where a user holds a mobile device and places an AR scene on an office desk. We then captured 10 images of each artifact at 720p from varying angles as the artifact rotated on the stand. Subsequently, each photo underwent marker detection processing using the method respective to its type of marker.<sup>1</sup>

<sup>1</sup>For ArUco, we utilized OpenCV’s detector (`cv::aruco::ArucoDetector`), and generated ArUco markers using the `DICT_4X4_50` library, which is the most robust one for detection.

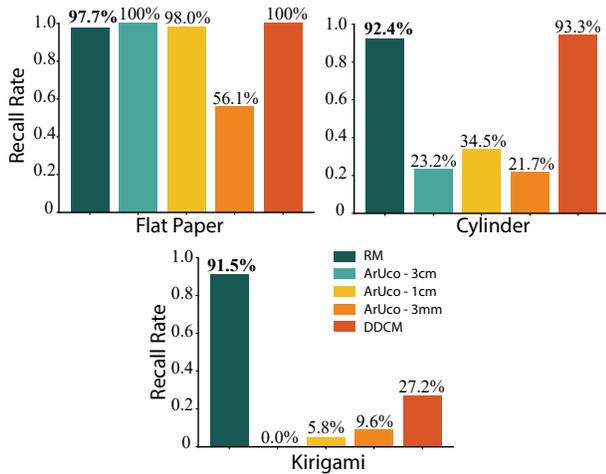


Figure 7: Recall rate of marker detection.

The performance of the detection process was assessed based on the recall rate, defined as the ratio of successfully detected markers to the total markers present in a frame. Precision was not used in this comparison, as we did not observe a noticeable number of false positives across all the methods. A marker is categorized as “detected” if it either matches accurately with its original design (in the case of our method) or is identified with the correct ID (for ArUco and DDCM). We determined the total number of markers present in each frame through manual inspection, counting only the markers whose complete shape is clearly visible.

The results are depicted in Fig. 7. Our method consistently achieves an overall recall rate exceeding 90% across all test artifacts. Both the regular and small versions of ArUco exhibit satisfactory performance on flat paper; however, their accuracy significantly declines when the paper is either curved or has a disjointed surface. Due to its limited marker size, tiny ArUco does not produce adequate performance on all artifacts. While DDCM manages to deliver reasonably good performance on both flat paper and cylindrical shapes, it struggles to robustly detect features on kirigami surfaces. A high recall rate is foundational to reliable camera pose estimation, as noted in Kalaitzakis et al. [28], and directly impacts the performance of subsequent interaction events. This implies that our method can naturally adapt to various paper forms, making them ready for easy interaction.

## 5.2 Accuracy Evaluation For Different Interaction Event

The previous evaluation demonstrates that our system pipeline enables robust detection of individual markers. Next, we aim to analyze how the detection of these markers assists in recognizing distinct input events.

To carry out this evaluation, we implemented the RM detection algorithm from Sec. 3.3 and the event detection algorithm described in Sec. 3.4 on an iPhone 14 Pro. Six different types of paper artifacts were used for this evaluation. These include the three artifacts utilized in the previous experiment, along with a folded push button, an origami cootie catcher, and the deformable

strips. However, considering the varied nature of these paper artifacts, it is impractical to assess all artifacts across every interaction event. Consequently, our primary focus is to evaluate each artifact within specific evaluations where it suitably accommodates the corresponding input event. The corresponding images/shapes for each artifact can be found as follows: Flat Paper: Fig. 1-a, Cylinder: Fig. 3-c, Kirigami: Fig. 1-c, Cootie Catcher: Fig. 5, Push Button: Fig. 6-a, Deformable Strips: Fig. 1-d.

The evaluation protocol for each detectable event is detailed below:

- **Tapping:** This is evaluated on the kirigami and the cylinder. We establish four touch locations on each artifact. For the kirigami, reflecting its stair-like structure, the touch points corresponded to the flat surfaces of the four levels, as shown in Fig. 1-c. For the cylinder, the four touch points were uniformly spaced along the curved surface, with the first point near the top edge and the last point near the bottom. Participants are instructed to touch the designated locations on each artifact in a random sequence. Our system then records whether the touch event was activated at the correct location.
- **Touch Tracking:** This is evaluated on the flat paper and the cylinder. We define two non-intersecting paths on each artifact, with one straight path and one circular path, to prevent any interference with each other. Each path is marked with six checkpoints. A touch tracking event is considered successful if all the checkpoints on the path are activated in the correct sequence.
- **Appearance Change:** This is evaluated on the push button and the cootie catcher. For the push button, we recognize two states: the pressed state and the released state. For the cootie catcher, we recognize three states: closed, opened vertically, and opened horizontally. The system evaluates whether the intended appearance change has been accurately identified.
- **Deformation:** This is evaluated on the deformable slips and the cootie catcher. The evaluation of deformation differs from other tests as we only calculate a scalar value, denoted as “deformation strength”, instead of tracking the 3D geometry in every frame. To assess this, we instruct the user to replicate the same action everytime on the evaluated artifacts, and measure the standard deviation of the deformation strength obtained from each attempt. This value potentially indicates the reliability of this signal when the user executes a consistent action at a human level.

We engaged 8 participants to take part in this study. During the evaluation, participants were asked to secure the phone to their heads using a camera headband, simulating the experience of wearing an AR/VR headset. In each individual trial, the artifacts were evaluated separately. Participants were asked to head at the artifact, ensuring that the camera was also focused on it, and then interact with the artifact using the designated actions 30 times by following our protocol. The precision and recall for each event and artifact are calculated to understand the detection performance.

Setup	Tapping		Touch Tracking		Appearance Change	
	Kirigami	Cylinder	Flat Paper	Cylinder	Push Button	Cootie Catcher
Precision	92%	95%	100% / 100%	100% / 100%	99%	94%
Recall	90%	91%	89% / 62%	91% / 47%	95%	92%

**Table 2: Results of event detection. For touch tracking, the first number indicates the straight path, and the second number indicates circular path.**

The results are displayed in Table ?? . Our system demonstrates promising accuracy in both tapping and appearance change detection, suggesting that these events could be utilized more frequently when constructing paper tangible interfaces.

For touch tracking, we observe varied results depending on the gesture. The straight path, which can be interpreted as a swipe gesture, exhibits very high accuracy. However, the circular path demonstrates a low recall rate, indicating a significant number of missed detections on both artifacts. This encourages us to design simple gestures, such as swipes, when utilizing the touch tracking functionality, instead of aiming to treat it as a high-precision 2D trackpad or design complex gestures for it.

The deformation results also hint at promising usability, particularly with the deformable slips. They have a standard deviation of 9% of the average signal strength, indicating a consistent level of signal. Conversely, the cootie catcher exhibits more fluctuating results, with a notably high standard deviation of 35%. One plausible explanation is that users can perform actions more consistently when the shape of the deformed artifacts is simple. More complex objects may also elevate the detection error rate. This leads us to recommend the use of simpler shapes when designing deformation-based interactions.

## 6 DISCUSSION AND CONCLUSION

Like all existing vision-based approaches to tangible interaction, our system also has several limitations stemming from fundamental principles. One primary constraint is the necessity for a line of sight with the object in interaction. In practical terms, this implies that users wearing AR glasses can only interact with the artifacts if they are within their direct view. Furthermore, although we have demonstrated a wide range of adaptability regarding the form of the paper and interaction events, certain events with specific artifacts can pose difficulties in detection. For instance, touch tracking has shown to be less robust compared to other interaction events. Therefore, we recommend using swipe gestures with simple paths instead of relying on full 2D touch tracking when developing paper interfaces.

Another limitation is the durability of paper-based interface. Paper, especially the one used for printing, is considered less durable than other solid materials. As time passes, the markers printed on the paper may fade and become less visible. Tears or other forms of wear could also compromise detection accuracy. Thus, we recommend using materials such as cardboard or plastic paper for those seeking to construct enduring interactive artifacts.

At present, we gauge deformation by measuring the shape difference between the original and the current states, represented by a scalar value. However, as noted by Boem et al. [3], deformations can manifest in a variety of ways, including twists, bends, and shears. A promising future development is the creation of algorithms that can

recognize different deformation events, leveraging our method’s ability to detect geometry changes in patterned artifacts. This is especially significant as deformations present unique features enabled by paper, as opposed to other tangible interfaces. Moving forward, it would be interesting to apply this technique to other deformable objects using methods such as hydrographic printing [54], thereby enabling interactivity with a variety of deformable objects.

A final area of limitation in our current system is that interaction events are manually programmed into our detection system. In the future, we aim to provide an authoring tool that will allow average users to construct and personalize tangible interaction experiences using paper.

As another potential future extension, since our design consists only of black-and-white shapes, the system can naturally detect patterns through an infrared camera when printed with IR ink. This integration fosters dual-channel data transmission, effectively presenting human-readable information on paper while simultaneously enabling detectable interaction [12, 27, 30].

In conclusion, we have introduced the novel marker-based system, ReactFold, which supports various fundamental interaction types with paper artifacts. These detectable events include occlusion, tapping, touch tracking, appearance change, tracking of RM-patterned objects, and deformation. All interaction events can be monitored just using a standard RGB camera, which is a standard hardware available on almost all AR/VR headset of mobile device. We validated our approach by analyzing the drawbacks of previous marker-based systems and conducting extensive experiments to showcase our system’s advantages and robustness. Opening the doors to a range of novel interfaces and applications, we believe our system serves as a functional prototype for experimenting with paper-based interactions and holds the potential to inspire future developments.

## REFERENCES

- [1] Christoph H Belke and Jamie Paik. 2017. Mori: a modular origami robot. *IEEE/ASME Transactions on Mechatronics* 22, 5 (2017), 2153–2164.
- [2] Steve Benford, Boriana Koleva, Anthony Quinn, Emily-Clare Thorn, Kevin Glover, William Preston, Adrian Hazzard, Stefan Rennick-Egglestone, Chris Greenhalgh, and Richard Mortier. 2017. Crafting interactive decoration. *ACM Transactions on Computer-Human Interaction (TOCHI)* 24, 4 (2017), 1–39.
- [3] Alberto Boem and Giovanni Maria Troiano. 2019. Non-rigid HCI: A review of deformable interfaces and input. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 885–906.
- [4] Matthew Brown and David G Lowe. 2005. Unsupervised 3D object recognition and reconstruction in unordered datasets. In *Fifth International Conference on 3-D Digital Imaging and Modeling (3DIM'05)*. IEEE, 56–63.
- [5] Zekun Chang, Heeju Kim, Kunihiro Kato, Kazuya Saito, Tung D Ta, Weiwei Jiang, Koya Narumi, Yoshinobu Miyamoto, and Yoshihiro Kawahara. 2019. Kirigami keyboard: Inkjet printable paper interface with kirigami structure presenting kinesthetic feedback. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–5.
- [6] Max Chen, Shano Liang, and Gillian Smith. 2023. Stackable Music: A Marker-Based Augmented Reality Music Synthesis Game. In *Companion Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 22–28.

- [7] Billy Cheng, Maxine Kim, Henry Lin, Sarah Fung, Zac Bush, and Jinsil Hwaryoung Seo. 2012. Tessella: interactive origami light. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*. 317–318.
- [8] Enrico Costanza and Jeffrey Huang. 2009. Designable visual markers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1879–1888.
- [9] Oscar Danielsson, Magnus Holm, and Anna Syberfeldt. 2020. Augmented reality smart glasses in industrial assembly: Current status and future challenges. *Journal of Industrial Information Integration* 20 (2020), 100175.
- [10] Claudia Daudén Roquet, Jeeun Kim, and Tom Yeh. 2016. 3D Folded PrintGami: transforming passive 3D printed objects to interactive by inserted paper origami circuits. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 187–191.
- [11] Himani Deshpande, Clement Zheng, Courtney Starrett, Jinsil Hwaryoung Seo, and Jeeun Kim. 2022. Fab4D: An Accessible Hybrid Approach for Programmable Shaping and Shape Changing Artifacts. In *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 1–7.
- [12] Mustafa Doga Dogan, Alexa F Siu, Jennifer Healey, Curtis Wigington, Chang Xiao, and Tong Sun. 2023. StandARone: Infrared-Watermarked Documents as Portable Containers of AR Interaction and Personalization. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7.
- [13] Mark Fiala. 2005. ARTag, a fiducial marker system using digital techniques. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 2. IEEE, 590–596.
- [14] Blender Foundation. 2021. *Blender* (3.0 ed.). Blender Foundation, Amsterdam. Available from: <https://www.blender.org/>.
- [15] David T Gallant, Andrew G Seniuk, and Roel Vertegaal. 2008. Towards more paper-like input: flexible input devices for foldable interaction styles. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*. 283–286.
- [16] Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco José Madrid-Cuevas, and Manuel Jesús Marín-Jiménez. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47, 6 (2014), 2280–2292.
- [17] Jun Gong, Olivia Seow, Cedric Honnet, Jack Forman, and Stefanie Mueller. 2021. MetaSense: Integrating sensing capabilities into mechanical metamaterial. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 1063–1073.
- [18] Nan-Wei Gong, Jürgen Steimle, Simon Olberding, Steve Hodges, Nicholas Edward Gillian, Yoshihiro Kawahara, and Joseph A Paradiso. 2014. PrintSense: a versatile sensing technique to support multimodal flexible surface interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1407–1410.
- [19] Daniel Groeger and Jürgen Steimle. 2019. Lasec: Instant fabrication of stretchable circuits using a laser cutter. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [20] Anoto Group. 2017. Anoto Digital Pen. <https://www.anoto.com/>.
- [21] Igor Guskov, Sergey Klibanov, and Benjamin Bryant. 2003. Trackable surfaces. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*. Citeseer, 251–257.
- [22] Chris Harris, Mike Stephens, et al. 1988. A combined corner and edge detector. In *Alvey vision conference*, Vol. 15. Citeseer, 10–5244.
- [23] Richard Hartley and Andrew Zisserman. 2003. *Multiple view geometry in computer vision*. Cambridge university press.
- [24] David Holman, Roel Vertegaal, Mark Altosaar, Nikolaus Troje, and Derek Johns. 2005. Paper windows: interaction techniques for digital paper. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 591–599.
- [25] Zhaoyang Huang, Xiaokun Pan, Weihong Pan, Weikang Bian, Yan Xu, Ka Chun Cheung, Guofeng Zhang, and Hongsheng Li. 2022. Neuralmarker: A framework for learning general marker correspondence. *ACM Transactions on Graphics (TOG)* 41, 6 (2022), 1–10.
- [26] John Illingworth and Josef Kittler. 1988. A survey of the Hough transform. *Computer vision, graphics, and image processing* 44, 1 (1988), 87–116.
- [27] Weiwei Jiang, Jorge Goncalves, and Vassilis Kostakos. 2024. Mobile Near-Infrared Sensing-A Systematic Review on Devices, Data, Modeling and Applications. *Comput. Surveys* (2024).
- [28] Michail Kalaitzakis, Brennan Cain, Sabrina Carroll, Anand Ambrosi, Camden Whitehead, and Nikolaos Vitzilaios. 2021. Fiducial markers for pose estimation: Overview, applications and experimental comparison of the artag, apriltag, aruco and stag markers. *Journal of Intelligent & Robotic Systems* 101 (2021), 1–26.
- [29] Annie Kelly, R Benjamin Shapiro, Jonathan de Halleux, and Thomas Ball. 2018. ARcadia: A rapid prototyping platform for real-time tangible interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–8.
- [30] Han-Jong Kim, Ju-Whan Kim, and Tek-Jin Nam. 2016. Ministudio: Designers' tool for prototyping ubicomp space with interactive miniature. In *Proceedings of the 2016 CHI Conference on human factors in computing systems*. 213–224.
- [31] Byron Lahey, Audrey Girouard, Winslow Burleson, and Roel Vertegaal. 2011. PaperPhone: understanding the use of bend gestures in mobile devices with flexible electronic paper displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1303–1312.
- [32] Dongchi Lee, Kazuya Saito, Takuya Umedachi, Tung D Ta, and Yoshihiro Kawahara. 2018. Origami robots with flexible printed circuit sheets. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. 392–395.
- [33] David G Lowe. 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60 (2004), 91–110.
- [34] Alex Mazursky, Shan-Yuan Teng, Romain Nith, and Pedro Lopes. 2021. MagnetIO: Passive yet interactive soft haptic patches anywhere. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [35] Gaku Narita, Yoshihiro Watanabe, and Masatoshi Ishikawa. 2016. Dynamic projection mapping onto deforming non-rigid surface using deformable dot cluster marker. *IEEE transactions on visualization and computer graphics* 23, 3 (2016), 1235–1248.
- [36] Masaya Nogi, Natsuki Komoda, Kanji Otsuka, and Katsuki Suganuma. 2013. Foldable nanopaper antennas for origami electronics. *Nanoscale* 5, 10 (2013), 4395–4399.
- [37] Simon Olberding, Sergio Soto Ortega, Klaus Hildebrandt, and Jürgen Steimle. 2015. Foldio: Digital fabrication of interactive and shape-changing objects with foldable printed electronics. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. 223–232.
- [38] Edwin Olson. 2011. AprilTag: A robust and flexible visual fiducial system. In *2011 IEEE international conference on robotics and automation*. IEEE, 3400–3407.
- [39] William Preston, Steve Benford, Emily-Clare Thorn, Boriana Koleva, Stefan Rennick-Egglestone, Richard Mortier, Anthony Quinn, John Stell, and Michael Worboys. 2017. Enabling hand-crafted visual markers at scale. In *Proceedings of the 2017 Conference on Designing Interactive Systems*. 1227–1237.
- [40] Daniela Rus and Michael T Tolley. 2018. Design, fabrication and control of origami robots. *Nature Reviews Materials* 3, 6 (2018), 101–112.
- [41] Cassandra Scheirer and Chris Harrison. 2022. DynaTags: Low-Cost Fiducial Marker Mechanisms. In *Proceedings of the 2022 International Conference on Multimodal Interaction*. 432–443.
- [42] Yanjun Shen. 2023. A survey of global conformal parameterization. In *International Conference on Pure, Applied, and Computational Mathematics (PACM 2023)*, Vol. 12725. SPIE, 320–325.
- [43] Beat Signer. 2017. *Fundamental concepts for interactive paper and cross-media information spaces*. BoD—Books on Demand.
- [44] Hyunyoung Song, François Guimbretière, Chang Hu, and Hod Lipson. 2006. ModelCraft: capturing freehand annotations and edits on physical 3D models. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*. 13–22.
- [45] Martin Spindler, Sophie Stellmach, and Raimund Dachselt. 2009. PaperLens: advanced magic lens interaction above the tabletop. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. 69–76.
- [46] Jürgen Steimle, Andreas Jordt, and Pattie Maes. 2013. Flexpad: highly flexible bending interactions for projected handheld displays. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 237–246.
- [47] Dominique Tan, Maciej Kumorek, Andres A Garcia, Adam Mooney, and Derek Bekoe. 2015. Projectagami: A foldable mobile device with shape interactive applications. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. 1555–1560.
- [48] Hideaki Uchiyama and Eric Marchand. 2011. Deformable random dot markers. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 237–238.
- [49] Hideaki Uchiyama and Hideo Saito. 2011. Random dot markers. In *2011 IEEE Virtual Reality Conference*. IEEE, 35–38.
- [50] Nirzaree Vadgama and Jürgen Steimle. 2017. Flexy: Shape-customizable, single-layer, inkjet printable patterns for 1d and 2d flex sensing. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*. 153–162.
- [51] Pierre Wellner. 1993. Interacting with paper on the DigitalDesk. *Commun. ACM* 36, 7 (1993), 87–96.
- [52] Nuwee Wiwatwattana, Chayangkul Laphom, Sarocha Aggaitchaya, and Sudarat Chattanon. 2016. Origami guru: An augmented reality application to assist paper folding. In *Information Technology: New Generations: 13th International Conference on Information Technology*. Springer, 1101–1111.
- [53] Po-Kang Yang, Zong-Hong Lin, Ken C Pradel, Long Lin, Xiuhuan Li, Xiaonan Wen, Jr-Hau He, and Zhong Lin Wang. 2015. Paper based origami triboelectric nanogenerators and self-powered pressure sensors. *ACS nano* 9, 1 (2015), 901–907.
- [54] Yizhong Zhang, Chunji Yin, Changxi Zheng, and Kun Zhou. 2015. Computational hydrographic printing. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 1–11.
- [55] Zhuming Zhang, Yongtao Hu, Guoxing Yu, and Jingwen Dai. 2022. Deeptag: A general framework for fiducial marker design and detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 3 (2022), 2931–2944.

- [56] Clement Zheng, Peter Gyory, and Ellen Yi-Luen Do. 2020. Tangible interfaces with printed paper markers. In *Proceedings of the 2020 ACM designing interactive systems conference*. 909–923.
- [57] Clement Zheng, HyunJoo Oh, Laura Devendorf, and Ellen Yi-Luen Do. 2019. Sensing kirigami. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 921–934.
- [58] Kening Zhu, Owen Noel Newton Fernando, Adrian David Cheok, Mark Fiala, Theam Wei Yang, and Hooman Aghaebrahimi Samani. 2010. A SURF-based Natural Feature Tracking System for Origami Recognition. In *Proceedings of 20th International Conference on Artificial Reality and Telexistence, Adelaide, Australia*.