Journal of the Brazilian Computer Society, 2026, 32:1, doi: 10.5753/jbcs.2026.7542

A survey on XR visuo-haptic illusions using real-virtual proxy pairing

David Ohara [**Universidade Federal de Pernambuco** | ***dosc@cin.ufpe.br***]
Marina Holanda Kunst [**Universidade Federal de Pernambuco** | ***marina.kunst@ufpe.br***]
Carlos Gabriel Bezerra Pereira [**Universidade Federal de Pernambuco** | ***cgbp@cin.ufpe.br***]
Fátima de Lourdes dos Santos Nunes Marques [**Universidade de São Paulo**| ***fatima.nunes@usp.br***]
João Marcelo Teixeira [**Universidade de São Paulo, Universidade Federal de Pernambuco**| ***jmxnt@cin.ufpe.br***]

Center of Informatics, Universidade Federal de Pernambuco, Av. Jornalista Aníbal Fernandes, s/n, Cidade Universitária, Recife, PE, 50740-560, Brazil.

Received: 01 February 2026 • **Accepted:** 28 May 2026 • **Published:** 31 July 2026

Abstract. Real-virtual object pairing is a central strategy for enabling convincing visuo-haptic experiences in Extended Reality (XR), allowing users to touch and manipulate physical proxies while perceiving aligned virtual counterparts. This paper presents a systematic review of XR systems that pair real and virtual objects to support tangible interaction, passive and encountered haptics, and illusion-based remapping. We organize the literature around three guiding questions: which pairing solutions and archetypes are used to achieve real-virtual alignment, how technological advances in head-mounted displays, sensing, and software pipelines shape these solutions over time, and what open challenges and future directions remain for scalable, robust deployments. Across the included studies, we synthesize recurring solution patterns, such as one-to-one tracked proxies, proxy reuse via retargeting and reset strategies, dynamic redirection for safe contact, and approaches that reduce or exploit mismatch through perceptual manipulations. We further consolidate reported challenges into cross-cutting categories (e.g., misalignment, tracking and latency issues, hardware and safety limitations, ergonomics, cybersickness, and experimental setup constraints), highlighting the persistent engineering and methodological barriers that limit generalization beyond controlled settings. Finally, we summarize the experimental landscape using a participant-tier matrix, covering 74 experiments and emphasizing the dominance of small-to-medium participant evaluations. Based on these findings, we outline actionable research directions toward perception-aware co-registration, standardized reporting and benchmarking, safer dynamic/robot-mediated pairing, and authoring tools that scale pairing to complex, multi-object XR scenarios.

Keywords: Visuo-Haptic Illusion, Real-Virtual Proxy Pairing, Extended Reality, Immersion

1 Introduction

Extended Reality (XR) has evolved from largely visual, screen-based simulations into increasingly embodied experiences in which users see, move, and act in spatially registered environments. Among XR modalities, Mixed Reality (MR) occupies a particularly relevant position in the Reality-Virtuality continuum by enabling virtual content to remain anchored to the physical world while supporting bidirectional interactions between real and virtual elements [Milgram and Kishino, 1994]. As MR hardware and tracking pipelines mature, a central opportunity emerges: calibrating everyday physical objects as passive haptic proxies that users can naturally touch, grasp, lean on, or manipulate, while perceiving them as virtual artifacts. This paradigm reduces dependence on costly active haptic devices and shifts the design focus to perceptual coherence, namely, the tight coupling of spatial alignment, visual rendering, and interaction timing.

The feasibility of such experiences is grounded in multisensory integration. Human perception combines cues from multiple modalities in a reliability-weighted manner [Ernst and Banks, 2002], and in many everyday conflicts between vision and touch, visual information can dominate or bias tactile judgments, for instance, in classic visuo-tactile illusions such as the rubber hand illusion [Botvinick and Cohen, 1998]. In XR, this has motivated a broad family of visuo-haptic techniques that deliberately exploit cross-modal effects to evoke sensations that are not physically present. Pseudo-haptics is a notable example, in which purely visual manipulations of motion, deformation, or interaction dynamics induce the impression of forces, stiffness, friction, or weight without dedicated force-feedback hardware [Lécuyer, 2009]. More recently, MR systems have used physical proxies and remapping strategies to let a limited set of tangible objects stand in for a richer set of virtual objects, including dynamic remapping of passive haptics to expand interaction bandwidth while preserving plausibility [Azmandian *et al.*, 2016].

A recurring technical and methodological challenge in this space is that the effectiveness of visuo-haptic illusions depends not only on perception but also on engineering details: tracking accuracy, anchor stability, calibration, and the ability to maintain persistent, low-drift registration between virtual assets and their real counterparts. This is particularly critical for real-virtual proxy pairing, where even small spatial errors can break the illusion and reduce the sense of presence. In practical MR development, spatial anchors and room-scanning pipelines have therefore become enabling technologies, allowing virtual content to be re-localized in a consistent physical frame across sessions and supporting repeatable experimental protocols. Our own tutorial work, Visuo-Haptic Immersion Made Easy: A Fast-Track Course on Spatial Anchors, systematizes this development pipeline by guiding practitioners to understand how real-virtual paired room can be created through important steps, such as environment scanning to get real object scales, 3D object manipulations to match real props scanned mesh, anchor creation and matched placement at their physical counterpart, and Unity-based deployment for Meta Quest 3, resulting in a functional visuo-haptic prototype and a reproducible workflow for co-located interaction [Ohara *et al.*, 2025a].

In parallel, empirical evidence has continued to accumulate showing that carefully constructed visual cues can substantially modulate tactile perception when a real surface is paired with a virtual rendering. In Tactile Perception in Matched Virtual Reality: Evaluating Virtual Textures on a Real Surface, we quantified how far texture based visuo-haptic illusions can go when multiple virtual materials are mapped onto a single, unchanging surface, physical table proxy. In a controlled study with four virtual textures, disposed in six left-right different texture pairs, our 35 participants had to touch both sides at same time, and then rate each one in a 1-7 (smoothest to roughest) Semantic Differential Scale. We observed robust differences in perceived roughness ratings and introduced a Deception Index to characterize user susceptibility, and notably, 71% of participants' ratings varied between texture pairs, despite the unchanged physical surface, in at least 5 out of 6 comparison cycles, indicating that visual appearance can override constant tactile input under tight spatial alignment [Ohara *et al.*, 2025b]. These findings reinforce a broader trend: a substantial portion of visuo-haptic immersion in XR is increasingly achieved through combinations of MR-grade registration, real-virtual proxy pairing and illusion-oriented interaction design.

Despite this momentum, the literature remains fragmented across venues, domains, such as training, prototyping, accessibility, entertainment, and technical choices, which can be proxy types, tracking stacks, anchoring strategies and rendering manipulations. As a result, it is difficult for researchers and practitioners to: Compare solution archetypes on a shared set of dimensions, understand which technical bottlenecks recur across studies, and identify research gaps that persist even as hardware improves. This paper addresses that need by presenting a systematic survey of XR visuo-haptic immersion approaches that explicitly rely on real-virtual object and environment pairing. Concretely, our survey is organized around three main axes: MR, real-virtual proxy pairing, and visuo-haptic illusions, and focuses on works that substitute or reinterpret real objects/environments with virtual counterparts while their spatial alignments stay paired, and other properties, such as material appearance, texture, or haptic response, can be changed. Following a structured search and screening process, we review 75 studies spanning the period from 2021 to 2025 and compare them in terms of experimental objectives, methodologies, technical challenges, alignment strategies, and participant sampling.

Our study is guided by the following objectives:

- Evaluate the state of the art (2021-2025) in XR research that pairs physical and virtual objects/environments in real-world coordinates to enable visuo-haptic interaction.
- Identify which problems are being addressed by lightweight and deployable approaches for visuo-haptic immersion based on comparison and substitution between real and virtual elements.
- Report limitations (both overcome and persistent) relative to the technologies available within the analyzed time frame.
- Identify trends and promising future directions for visuo-haptic immersion in MR.

In this paper, we use the terms "lightweight" and "deployable" descriptively, not as a binary classification criterion. They refer to approaches that seek to reduce the cost, complexity, or setup burden of visuo-haptic XR systems, for example by using passive physical proxies instead of specialized force-feedback devices, commodity HMDs instead of laboratory-only displays, standard game engines or SDKs instead of entirely custom software stacks, and reproducible alignment procedures instead of highly controlled one-off setups. Thus, our goal is not to classify each study as basic or non-basic, but to analyze how the literature attempts to make real-virtual proxy pairing more practical, scalable, and reproducible.

To operationalize these objectives, we formulate three research questions:

1. What archetypes of visuo-haptic immersion solutions have been reported using real-virtual pairing of objects and environments?
2. How visuo-haptic approach evolved (in terms of software and hardware) in the last decade?
3. Which challenges still persist and what is expected from future XR platforms regarding visuo-haptic illusion supported by real-virtual pairing?

Beyond consolidating prior work, we also justify this survey through the lens of our two complementary contributions: an empirical demonstration that visual texture remapping over a single proxy can reliably bias tactile judgments under rigorous alignment [Ohara *et al.*, 2025b], and a hands-on, reproducible development pathway for building anchored visuo-haptic MR applications with commodity hardware and widely available toolchains [Ohara *et al.*, 2025a]. Together, these works expose both the promise (low-cost immersion via perception-driven design) and the complexity (registration stability, proxy constraints, evaluation variability) of the area, motivating the need for a systematic synthesis that characterizes solution families, recurring challenges, and open research directions at scale.

The remainder of this paper is organized as follows. Section 2 introduces the core concepts underlying MR, real-virtual proxy pairing, and visuo-haptic illusions, and positions our previous contributions as reference points. Section 3 describes the review protocol and screening process (aligned with PRISMA 2020 [Page *et al.*, 2021]). Section 4 reports the main findings, including timelines, archetypes, and methodological patterns across studies. Section 5 discusses implications, gaps, and trends in relation to our research questions. Finally, Section 6 summarizes contributions and outlines opportunities for future work on scalable visuo-haptic immersion in MR.

2 Background

XR is commonly used as an umbrella term covering Virtual Reality, Augmented Reality, and Mixed Reality. A classical perspective is the Reality-Virtuality continuum, which describes a spectrum ranging from purely real environments to fully virtual ones, with AR and Augmented Virtuality occupying intermediate regions [Milgram and Kishino, 1994]. While VR replaces the user's sensory input with a synthetic environment, AR overlays virtual content onto the physical world; MR emphasizes consistent spatial registration and interaction between real and virtual entities, enabling users to act on both with coherent perceptual consequences.

From a visuo-haptic standpoint, MR is especially relevant because it supports co-located interaction, where the user can see a virtual counterpart at the same location where their hand contacts a physical proxy (Figure 1.A). This alignment is critical: if the system preserves geometry, position, and timing within tolerable bounds, the user's perceptual system can fuse cues across modalities and maintain plausibility during manipulation [Cosco *et al*., 2012]. Consequently, MR has become a practical substrate for building illusion-oriented experiences in which limited physical resources (such as a single proxy) can represent a larger set of virtual objects through perceptual substitution and remapping.

Real-virtual pairing refers to the deliberate mapping of a physical object or surface (the proxy) to a virtual representation so that interaction occurs in shared, real-world coordinates. The simplest case is passive haptics, where the proxy provides contact geometry, friction, and compliance without active actuation, while rendering supplies the visual appearance and interaction logic. Pairing can also be applied to environments (e.g., tables, walls, floors, cockpit elements), enabling co-located constraints and touchpoints that enhance presence and task performance.

A central motivation for pairing is scalability: realistic force-feedback hardware remains expensive and application-specific, while passive proxies can deliver rich tactile cues at low cost if alignment is maintained. However, pairing introduces engineering and methodological constraints. Co-located interaction requires accurate tracking of the headset and user input, calibration between physical and virtual frames, robust persistence across sessions when longitudinal evaluation is desired, and predictable error characteristics so that researchers can interpret perceptual outcomes.

To expand interaction bandwidth, the literature has proposed proxy repurposing, in which a single physical object stands in for multiple virtual objects. A prominent example is haptic retargeting, which dynamically realigns the mapping between the user's motions, the virtual scene, and the proxy (Figure 1.B) so that different virtual objects can be touched using the same physical prop while maintaining plausibility [Azmandian *et al*., 2016]. Such approaches explicitly rely on perceptual tolerance to small spatial or kinematic discrepancies, making their effectiveness contingent on both human factors and system accuracy.

Finally, persistent MR often depends on anchoring and relocalization mechanisms (for instance, spatial anchors and scene understanding pipelines (Figure 1.D)) that allow content to be placed and recovered in stable physical locations. Recent work on shared anchors highlights that persistence and cross-user consistency can incur non-trivial overheads, such as multi-step relocalization and synchronization costs, particularly when anchors must be exchanged across devices or sessions [Vidal-Balea *et al*., 2025]. These constraints directly impact experimental repeatability and the feasibility of basic (deployable) visuo-haptic systems outside controlled labs.

Visuo-haptic illusions exploit the fact that perception is multisensory and inferential: the brain combines cues from different modalities according to their relative reliability and the assumed common cause. In foundational evidence for statistically efficient integration, humans combine visual and haptic information in a near-optimal manner under many conditions [Ernst and Banks, 2002]. When cues conflict yet are interpreted as originating from the same object or event, one modality may bias the other, producing systematic distortions in perceived touch. The rubber hand illusion (Figure 1.E) is a canonical demonstration that congruent visual and tactile evidence can reshape body ownership and tactile experience [Botvinick and Cohen, 1998].

In XR, these principles motivate software-driven approaches to haptic enhancement. Pseudo-haptics, in particular, uses visual manipulations of motion, deformation, or interaction dynamics to induce impressions of force, stiffness, friction, or weight without dedicated force-feedback devices [Lécuyer, 2009]. Contemporary surveys emphasize that sensory-illusion-based haptics remains an active and diversifying research area in XR, spanning pseudo-haptics, remapping techniques, and hybrid visual-tactile strategies [Kurzweg *et al*., 2024]. Recent empirical work continues to refine illusion design for mid-air interaction and commodity hand tracking (Figure 1.C); for example, pseudo-haptic rendering can improve perceived button-press realism and user experience even when no physical contact occurs [Kim and Xiong, 2022].

In MR specifically, visuo-haptic illusions increasingly handle pairing to anchor tactile expectations to real contact, and to selectively modulate only the properties that the proxy cannot change (the surface appearance, micro-texture cues, or material semantics). A related trend is the use of video see-through manipulation to reshape how the user visually interprets the physical world while preserving contact: MaskWarp, for instance, demonstrates visuo-haptic illusions in MR by altering passthrough video using real-time inpainting, enabling masked physical interactions that support substitution and plausibility [Matthews *et al*., 2023a]. Across application domains, passive haptic strategies have also been adopted in training and education as pragmatic alternatives to complex actuation, with studies reporting that tangible proxies can improve engagement and task grounding when integrated into immersive workflows (Figure 1.F) [Calandra *et al*., 2023; Johnson-Glenberg *et al*., 2023].

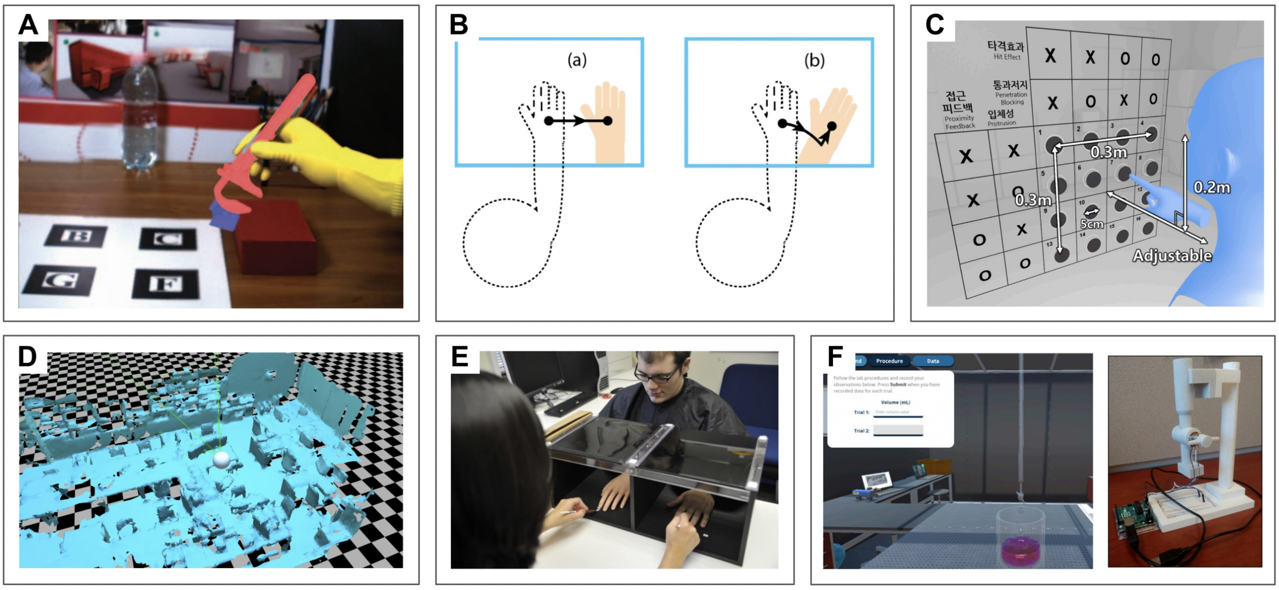

Figure 1. Examples from haptic-related works: A) virtual counterpart at the same location [Cosco *et al.*, 2012]; B) haptic retargeting [Azmandian *et al.*, 2016]; C) pseudo-haptic button for mid-air interaction [Kim and Xiong, 2022]; D) 3D mesh of the spatial mapping (room-sized anchor) [Vidal-Balea *et al.*, 2025]; E) Rubber hand illusion [Botvinick and Cohen, 1998]; F) tangible proxies integrated into immersive workflows [Johnson-Glenberg *et al.*, 2023].

3 Methodology

3.1 PRISMA 2020

To ensure transparency and reproducibility, this survey follows the PRISMA 2020 guideline for reporting systematic reviews [Page *et al.*, 2021]. PRISMA 2020 provides a 27-item checklist and a flow-based structure to document how records are identified, screened, assessed for eligibility, and ultimately included in the synthesis. Although originally designed to standardize reporting across evidence-based disciplines, PRISMA has also become a widely adopted framework for technology-oriented surveys because it makes the review pipeline explicit (presenting the consulted sources, screening decisions, and data handling), reducing ambiguity and selection bias in the reporting of literature searches [Page *et al.*, 2021].

In this work, PRISMA 2020 is used as a reporting backbone for the end-to-end review protocol, from query formulation to final synthesis. Specifically, we report the information sources and search strategy used to retrieve candidate records, the screening stages and decision rules applied to exclude out-of-scope studies, the eligibility assessment based on full-text reading, and the resulting included corpus that supports the comparative analyses presented in Sections 4 and 5 [Page *et al.*, 2021]. The PRISMA logic is reflected in our screening workflow, according to Figure 2, which separates initial retrieval from progressively stricter screening and full-text eligibility assessment.

We adopt PRISMA 2020 without changing its core phases (identification, screening, eligibility, inclusion), however, we extend the reporting and extraction to account for characteristics that are central in XR visuo-haptic research yet typically absent from PRISMA-focused health/behavioral reviews. XR studies are often defined as much by engineering choices (hardware, tracking, anchoring, rendering, interaction modality) as by experimental outcomes, and these choices can be decisive for whether a visuo-haptic illusion is feasible, reproducible, and comparable across labs. Therefore, in addition to standard PRISMA elements, we explicitly track XR-specific descriptors during data extraction, including:

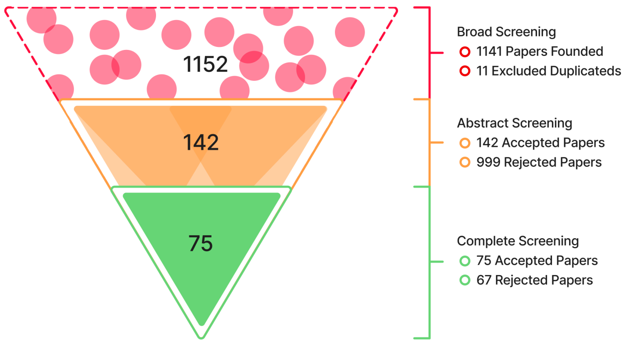

Figure 2. Screening Method

- XR modality and display pipeline: whether the system is VR/AR/MR and, when relevant, whether interaction relies on optical see-through or video see-through rendering.
- Real-virtual pairing strategy: proxy type (object/surface/environment), mapping strategy (one-to-one vs. repurposing), and co-location assumptions.
- Alignment and persistence mechanisms: calibration approach, tracking stack, and anchoring/relocalization mechanisms when cross-session stability is required.
- Illusion mechanism: which visual manipulations or pseudo-haptic cues are used to bias or complement tactile perception.
- Evaluation and participants: task type, outcome variables, and participant sampling (including sample size reporting and recruitment context when available).

Table 1 summarizes how PRISMA 2020 stages map to our workflow, and highlights the XR-oriented additions that support the taxonomy-driven synthesis presented later in the paper.

By combining PRISMA 2020 with XR-specific descriptors,

Table 1. Operationalization of PRISMA 2020 for this survey, including XR-specific reporting extensions used in data extraction.

PRISMA phase	Purpose in PRISMA 2020	Operationalization in this survey (XR-oriented)
Identification	Retrieve candidate records from selected information sources using a defined search strategy	Query formulation and database searches; record management and de-duplication; capture of source metadata (venue type, year)
Screening	Rapid exclusion of clearly irrelevant records based on minimal fields	Title/keyword and abstract-based filtering to remove out-of-scope topics; early enforcement of XR + visuo-haptic + pairing relevance
Eligibility	Full-text assessment against explicit inclusion/exclusion criteria	Complete reading to confirm real-virtual pairing, visuo-haptic mechanism, and empirical/technical contribution; extraction of hardware, alignment, and proxy details
Included	Final corpus used for synthesis and reporting	Structured dataset supporting taxonomy and comparative analysis (modalities, proxy pairing archetypes, illusion mechanisms, challenges, and evaluation characteristics)

the review remains faithful to a standard, auditable reporting format while also enabling a systems-level synthesis that is actionable for both researchers and practitioners working on deployable visuo-haptic MR experiences.

3.2 Screening Methods

Our screening procedure followed a staged selection pipeline consistent with PRISMA 2020 recommendations for transparent identification, screening, eligibility assessment, and inclusion reporting [Page *et al*., 2021]. In line with the scope of our survey, the screening targeted peerreviewed studies in XR (with emphasis on Mixed Reality) that explicitly pair a physical proxy (object, surface, or environment) with a spatially aligned virtual counterpart and employ visuo-haptic phenomena (such as pseudo-haptics, visuo-tactile bias, remapping/retargeting, or illusion-oriented interaction design) to expand or reshape tactile experience beyond the intrinsic properties of the proxy.

Information sources. We queried three major publication platforms that consistently index XR/HCI venues relevant to visuo-haptic interaction: the ACM Digital Library (ACM), IEEE Xplore (IEEE), and Elsevier ScienceDirect (SD). Searches were restricted to the survey time window (2021-2025), and results were exported for subsequent consolidation and screening.

Search strategy. Because terminology varies across subcommunities (for instance, visuo-haptic vs. visuotactile, passive haptics vs. physical proxy, retargeting vs. remapping), we adopted a concept-driven Boolean strategy that combines three facets: XR/MR context, real-virtual proxy pairing, and visuo-haptic illusion or perception-driven haptics. The core query was instantiated with platform-specific syntax but preserved the same semantic structure:

```
( "virtual reality" OR "mixed reality" OR
"augmented reality" )
AND ( "haptic feedback" OR "force
feedback" OR "visuo-haptic" )
AND ( "real environment" OR "spatial
alignment" OR "physical proxy" )
```

On ScienceDirect (SD), in addition to the Boolean query, we applied the platform's built-in filters (year range and publication type) to reduce clearly out-of-scope records before manual screening.

It is important to acknowledge that the haptics-related facet of the query was intentionally concise and therefore does not exhaustively cover all terminology used in the XR haptics literature. Although the terms "haptic feedback", "force feedback", and "visuo-haptic" capture a broad portion of the target literature, related studies may describe similar mechanisms using alternative expressions such as "passive haptics", "encountered haptics", "pseudo-haptics", "visuotactile", "tangible interaction", "haptic retargeting", "remapping", "physical props", or "physical proxies". Consequently, papers that exclusively use these alternative terms in their titles, abstracts, or keywords may not have been retrieved by the initial search.

This choice reflects a methodological trade-off between recall, precision, and cross-platform reproducibility. The three consulted platforms (ACM Digital Library, IEEE Xplore, and ScienceDirect) differ in their support for long Boolean expressions, nested operators, wildcards, and field-specific searches. We therefore adopted a compact query that combined three central facets of our review scope: XR modality, haptic/visuo-haptic interaction, and real-virtual spatial pairing. To reduce the risk of excluding conceptually relevant studies, the subsequent screening stages did not rely solely on the exact terminology used in the query. During abstract and full-text screening, papers were retained when they implemented real-virtual proxy pairing and relied on relevant visuo-haptic mechanisms, including passive haptics, pseudo-haptics, retargeting, remapping, visuo-tactile bias, or illusion-oriented interaction design.

This restriction was also motivated by differences among the search interfaces of the consulted platforms, since not all of them handled long Boolean expressions, nested synonyms, and wildcard terms consistently. Therefore, we prioritized a compact and reproducible query centered on XR modality, haptic/visuo-haptic interaction, and real-virtual spatial pairing.

The screening and extraction workflow was conducted by two coauthors. These coauthors defined the extraction questions, applied the assisted extraction procedure described in Section 3.3, and manually verified the extracted information against the original papers. We note that this procedure did not constitute a fully independent double-review process in which two reviewers separately extracted all fields for every paper before reconciliation. Instead, after the candidate corpus was defined, the papers were processed using a standardized extraction prompt, and the resulting entries were checked by the two coauthors through consultation of the original PDFs. When the extracted information was ambiguous, incomplete, or inconsistent with the source paper, the coauthors discussed the case and reached a consensus decision based on the original text.

Eligibility criteria. The following inclusion criteria (IC) and exclusion criteria (EC) were applied throughout the stages below.

- IC1: XR/MR setting. The study is situated in XR, with mixed reality, augmented virtuality, or closely related spatially registered AR/MR configurations.
- IC2: Real-virtual proxy pairing. The study includes an explicit pairing between a physical element (object, surface, or environment) and a virtual representation intended to be interacted with in (approximately) the same real-world coordinates.
- IC3: Visuo-haptic mechanism. The study relies on visuo-haptic effects or illusion-oriented design (such as pseudo-haptics, visuo-tactile bias, remapping/retargeting of passive haptics, visual manipulation of perceived material/texture/forces) to modulate touch or haptic perception.
- IC4: Peer-reviewed and accessible. The work is peer-reviewed (journal, conference, or workshop proceedings) and provides sufficient methodological/technical detail to support extraction (system description and/or evaluation protocol).
- IC5: Time window. The publication date falls within 2021-2025 (inclusive).
- EC1: Short papers. Studies with ≤ 4 pages of substantive content (which does not count references).
- EC2: Reviews. Papers focused on systematic reviews, literature surveys, and meta-analyses.
- EC3: English papers. Papers whose full text is not available in English.
- EC4: Duplicate papers. When the same study appears on multiple platforms (preprint, repository, conference, journal).

3.2.1 Broad Screening

The broad screening stage aimed to maximize recall while maintaining topical relevance. We executed the platform-specific queries, exported all retrieved records, and consolidated them into a single candidate set. Duplicate records across platforms were removed by matching title, authors, venue, and year (with manual verification for ambiguous cases, such as extended versions). At this stage, no study was excluded based on experimental design quality; instead, records were retained whenever their metadata suggested potential alignment with IC1-IC3 (XR/MR context, proxy pairing, and visuo-haptic mechanism). Records clearly violating EC3 (non-scholarly outputs) were removed during consolidation.

3.2.2 Abstract Screening

In the abstract screening stage, we evaluated titles, abstracts and articles metadata to filter out false positives produced by broad terminology (for example, haptics used in unrelated robotics contexts, or anchors used only for tracking without interaction). Each record was assessed against the inclusion/exclusion criteria with particular focus on IC2-IC3, since proxy pairing and visuo-haptic mechanisms are often stated at a high level in abstracts, and EC1-EC2, given that most studies present relevant metadata to determine their number of content pages, and whether they are review-focused. When the abstract was insufficient to determine eligibility (common in short venue formats), the record was provisionally retained for full-text screening rather than discarded, prioritizing recall over premature exclusion.

3.2.3 Complete Screening

The complete screening stage assessed full texts to confirm that included studies truly implemented real-virtual proxy pairing and that the reported interaction relied on visuo-haptic effects beyond generic haptic feedback claims. We required that the paper provide enough technical detail to characterize: the physical proxy (type, role, and constraints), the pairing/alignment strategy (tracking, calibration, anchoring, or retargeting/remapping when applicable), and the visuo-haptic mechanism (visual manipulations, illusion framing, or perceptual rationale) in a way that supports consistent data extraction. Works that mentioned proxies only as incidental props, or that described haptics exclusively through actuation without proxy pairing, were excluded at this stage (EC1-EC2). The final set of included studies comprises 75 papers within 2021-2025, which were then passed to data extraction.

3.3 Data Collection

After the complete screening stage, the remaining 75 papers (2021-2025) were transformed into a structured dataset to support the comparative analyses reported in Section 4. The goal of data collection was to capture, in a standardized form, what each study attempted to achieve, characterize the technical choices that enable, or constrain, real-virtual proxy pairing, and summarize empirical outcomes and reported limitations in a way that enables aggregation across heterogeneous XR systems.

To improve reproducibility, the extraction prompt, the final verified spreadsheet, and the per-paper extraction notes can be made available as supplementary material upon request.

Operational meaning of lightweight and deployable approaches. Because the term "basic" may suggest a formal quality or complexity category, we avoid using it as an analytical label. Instead, we discuss "lightweight" and "deployable" approaches as a descriptive dimension of the reviewed systems. During extraction, this dimension was interpreted through observable implementation characteristics, including:

- (i) whether the system relied on passive physical proxies rather than specialized force-feedback or robotic haptic devices;
- (ii) whether it used commodity or broadly available XR hardware;
- (iii) whether the implementation was based on common development platforms, such as Unity and public SDKs;
- (iv) whether the alignment procedure could plausibly be reproduced without a highly specialized laboratory setup; and
- (v) whether the paper explicitly framed its solution as scalable, low-cost, portable, easy to implement, or suit-

able for deployment outside tightly controlled conditions.

This dimension was not used as an inclusion or exclusion criterion, nor did we assign a binary "basic" label to each paper. Instead, it supported the interpretation of recurring solution archetypes and challenges, especially the tension between practical passive-proxy solutions and the continued need for external tracking, calibration, custom software, and controlled environments.

Record management and workspace. All candidate records and screening decisions were registered in Parsif.al to centralize metadata and to keep a consistent audit trail across the PRISMA stages. For the final included set, each paper was assigned a unique identifier and stored in a shared workspace used by the authors for extraction, verification, and synthesis. The extraction workflow relied on Google Docs for per-paper notes and Google Sheets for consolidating the extracted fields into a single analyzable table. Figure production and graphic organization were supported by design tools (Figma and Canva) after the dataset was stabilized.

Extraction schema. For each of the 75 included papers, we extracted the following fields:

- Objective: the stated goals of the study, which included, among others, proposing an interaction technique, evaluating an illusion mechanism, enabling a workflow, and validating a system in an application domain.
- Method: the primary research method used to achieve the objective, which may be a controlled experiment, framework proposal, tutorial, tool evaluation, system demonstration, or comparative study.
- Hardware: the Head-Mounted Display (HMD) and relevant supporting devices explicitly used, such as hand tracking, controllers, depth sensors, trackers, tangible props, or custom-built components.
- Software: the application and/or development stack used in the study, which may include game engine, SDK or toolkits. We explicitly noted whether the application was authored by the paper's authors or if it was a third-party/free application adopted as-is.
- Conclusion: the main outcomes reported by the authors, encompassing measured effects on perception or performance, qualitative feedback, feasibility claims, or deployment insights.
- Challenges: limitations and difficulties acknowledged in the paper, which may involve tracking/registration drift, calibration overhead, proxy constraints, cybersickness, sampling limitations, or evaluation threats to validity.

ChatGPT-assisted extraction. Given the scale of the corpus and the heterogeneity of reporting styles across venues, we employed ChatGPT as an assistance tool to support the initial organization of information extracted from the included papers. The tool was not used to decide whether a paper should be included or excluded, nor were its outputs accepted as final evidence without author verification. Its role was limited to producing a first-pass structured extraction for six predefined fields (Objective, Method, Hardware, Software, Conclusion, and Challenges) and to indicate the part of the paper where the corresponding information could be found.

The extraction questions were prepared by two coauthors before running the assisted extraction. For each paper, the full text was provided to ChatGPT and the following prompt was used:

> You are assisting with data extraction for a systematic survey on XR visuo-haptic immersion using real-virtual object or environment pairing. Based only on the provided paper, extract the following fields:
>
> 1. Objective: What is the main goal of the paper?
> 2. Method: What method, study design, system proposal, prototype, experiment, or evaluation was used?
> 3. Hardware: Which HMDs, sensors, trackers, controllers, haptic devices, props, or other physical devices were used?
> 4. Software: Which software, game engine, SDKs, toolkits, or custom applications were used? Indicate whether the software appears to be author-developed or an existing third-party application.
> 5. Conclusion: What are the main findings or conclusions reported by the authors?
> 6. Challenges: What limitations, technical issues, methodological constraints, or threats to validity are reported?
>
> For each field, provide:
>
> - a concise extracted answer;
> - the section, paragraph, table, or page where the information appears;
> - a short supporting excerpt from the paper.
>
> Do not infer information that is not explicitly stated. If the paper does not report a field, write "Not reported". If the information is ambiguous, write "Ambiguous" and explain why.

The use of ChatGPT accelerated the extraction process mainly by helping locate relevant sections and by producing an initial standardized summary for each predefined field. This reduced the time spent manually searching across heterogeneous paper structures for apparatus descriptions, implementation details, evaluation protocols, conclusions, and limitations. However, ChatGPT was treated only as a navigation and transcription aid. The final dataset was not generated automatically from the model outputs.

Author verification and correction. All ChatGPT-assisted outputs were manually verified by the two coauthors against the original PDFs. The verification focused on the sections or passages indicated by ChatGPT, but the authors also consulted surrounding sections whenever the extracted information was incomplete, ambiguous, or inconsistent with the paper. During this step, the coauthors confirmed whether each extracted field was explicitly supported by the original text. If a field was not clearly reported in the paper, the entry was marked as "Not reported" rather than inferred.

When discrepancies were identified, such as missing devices, incorrect software descriptions, overgeneralized conclusions, or challenges not actually stated by the authors, the entries were corrected based on the original paper. If the two coauthors disagreed with the ChatGPT output, or if the interpretation of a field was unclear, the final entry was decided by consensus after re-reading the relevant sections of the paper. Therefore, ChatGPT outputs served only as preliminary extraction candidates; the verified dataset used in the results was produced through human checking and correction.

Normalization for synthesis. To support aggregate analyses and visualizations, the verified data were normalized in the consolidated spreadsheet. This included harmonizing device names to maintain consistency in HMD model nomenclature; standardizing method labels by grouping closely related formats such as user study and controlled experiment under a common category when appropriate; and consolidating recurring challenge themes into higher-level descriptors used later to form challenge archetypes. The resulting dataset serves as the basis for the tables and figures in Section 4 and the cross-paper comparisons in Section 5.

4 Results

This section reports the main findings of our survey, synthesizing the evidence extracted from the final corpus of 75 peer-reviewed studies published between 2021 and 2025. The results are presented to support both an overall characterization of the recent state of the art in XR-centered visuo-haptic immersion based on real-virtual proxy pairing and the comparative analyses required to address our research questions, including timelines, archetypes, and methodological patterns across studies.

The corpus also shows that lightweight or deployable does not mean technically simple (Table 2). Even when studies use passive haptic proxies or commodity XR devices, many still depend on external trackers, depth/RGB-D sensors, calibration procedures, instrumented objects, or custom software pipelines to maintain real-virtual alignment. Therefore, deployability should be understood as a direction of design effort rather than as an achieved property shared by all included studies. This distinction is important because the reviewed literature often reduces dependence on specialized active haptic devices, but it does not eliminate engineering complexity.

4.1 Highlighted papers (largest participant samples studies)

From the 75 papers included in the final corpus, we selected a subset of ten studies (Table 3) for detailed highlighting based on a single, objective criterion: the papers reporting the largest participant samples (defined in N values). This choice is motivated by two practical considerations in visuo-haptic XR research. First, large participants N value studies tend to provide more stable estimates of perceptual and behavioral effects under the variability typical of immersive setups (tracking quality, embodiment, novelty, and individual differences), and a overall usability problems perception, reaching a minimum rating of 55% accuracy problem finding at 5 users samples, increased to 90% at 15 users, and 98% at 50 users [Faulkner, 2003]. Second, high-participation evaluations are often associated with applied contexts, such as training and safety, where real-virtual pairing must remain robust beyond carefully controlled demonstrations. The highlighted set therefore complements our taxonomy-driven synthesis by emphasizing works with strong empirical support and deployment-oriented constraints.

Table 2. Descriptive dimensions used to discuss lightweight and deployable visuo-haptic XR approaches.

Dimension	More lightweight/deployable indicators	Less lightweight/deployable indicators
Haptic mechanism	Passive proxies, everyday objects, physical surfaces, pseudo-haptic visual manipulation	Robotic platforms, specialized force-feedback devices, complex actuation systems
XR hardware	Commodity HMDs, standalone MR devices, standard controllers or hand tracking	Laboratory-only displays, high-end tracking rooms, specialized optical systems
Tracking and alignment	Inside-out tracking, spatial anchors, markerlight or markerless procedures	External motion capture, dense calibration, controlled-room assumptions
Software stack	Common engines and SDKs, reusable toolkits, reproducible pipelines	Fully bespoke software, undocumented pipelines, application-specific integration
Deployment assumptions	Portable, low-cost, scalable, or explicitly designed for practical use	Fixed installations, expert-only setup, high calibration burden

Large-scale comparisons of immersive setups and multisensory reconstruction. The largest sample in our selection is the study Effect of perceptive enrichment on the efficiency of simulated contexts: Comparing virtual reality and immersive room settings ($N = 279$), which compares performance and related outcomes between an HMD-based VR configuration and an immersive room/CAVE-like setup while examining whether perceptive enrichment improves efficiency in simulated contexts. Its main contribution, from the standpoint of this survey, is large-sample evidence on efficiency and experience trade-offs across immersive display modalities, a recurring concern when designing comparable conditions for proxy-based interaction and multisensory grounding [Gouton *et al*., 2023].

A second large-sample highlight is Towards Temporal Telepresence & Multisensory Spatial Intelligence: Reconstructing the Human Experience through TeleAbsence ($N = 115$), which frames TeleAbsence as a direction for reconstructing richer human experiences via telepresence, emphasizing multisensory and spatial intelligence aspects, and reporting evaluation signals with a sizable participant group. In the context of our review, this work stands out for connecting multisensory reconstruction to telepresence design, providing conceptual and technical framing supported by large-N feedback [Pillis, 2025].

Perceptual studies probing visuo-haptic conflicts and embodiment cues. Putting Vision and Touch into Conflict: Results From a Multimodal Mixed Reality Setup ($N = 90$) directly targets a core issue underlying real-virtual proxy pairing: what happens when visual and tactile cues conflict. By using a multimodal MR setup to probe dominance, integration, and perceptual consequences under controlled mismatch, the paper contributes empirical results and methodological guidance for studying cross-modal conflict, both as an intentional illusion mechanism and as an inadvertent artifact of imperfect alignment [Kang *et al*., 2023].

Complementing this, An Empirical Evaluation of How Virtual Hand Visibility Affects Near-Field Size Perception and Reporting of Tangible Objects in Virtual Reality ($N = 80$) examines how the visibility of virtual hands influences near-field size perception and the reporting of tangible objects in VR. Within our survey scope, it highlights that embodiment cues (hands) are not merely cosmetic: they can serve as perceptual references and reshape how users interpret and report co-located/tangible targets, yielding actionable design implications for proxy-based interaction [Murmu *et al*., 2025].

Applied training and safety contexts with substantial participant pools. Practical skills training in enclosure fires: An experimental study with cadets and firefighters using CAVE and HMD-based virtual training simulators ($N = 74$; 67 cadets + 7 firefighters) compares two virtual training systems for enclosure-fire nozzle person training: a CAVE and an HMD VR simulator. The study evaluates acceptance/usability, presence/realism, sickness, and stress considerations, offering practical evidence and design discussion that are especially relevant when real-virtual pairing must remain credible and safe in high-stakes training scenarios [Grabowski, 2021].

Similarly, Simulating Social Pressure: Evaluating Risk Behaviors in Construction Using Augmented Virtuality ($N = 64$) investigates whether simulated social pressure can influence risk behaviors in a construction context using augmented virtuality. In our survey, this work exemplifies how immersive systems can be used not only to train procedural skills, but also to probe behavioral outcomes under social influence, reinforcing the importance of ecological validity when physical surrogates and virtual overlays are combined [Pooladvand *et al*., 2025].

In healthcare training, ViTAWiN - Interprofessional Medical Mixed Reality Training for Paramedics and Emergency Nurses ($N = 44$; 30 paramedic students + 14 nursing trainees) describes an MR training setup in which different professional groups jointly train a shared burn-patient scenario integrating a haptic mannequin. The reported evaluation addresses usability, presence, sickness, and practical constraints, emphasizing challenges that emerge in collaborative, real-world training deployments (such as comfort, stability, and fidelity requirements) [Schild *et al*., 2022].

Tangible interaction primitives via real-virtual objects and props. Real-Virtual Objects: Exploring Bidirectional Embodied Tangible Interaction with a Virtual Human in World-Fixed Virtual Reality ($N = 40$) explores real-virtual objects that enable bidirectional tangible interaction between a user and a virtual human in world-fixed VR. For our survey, it offers an interaction concept and implementation framing that treats paired objects as shared interaction primitives, accompanied by evaluation-driven insights into embodied tangible interaction [Bozgeyikli, 2024].

Tangiball: Foot-Enabled Embodied Tangible Interaction with a Ball in Virtual Reality ($N = 40$) proposes a foot-enabled tangible interaction by pairing a tracked physical ball with a virtual ball and evaluating performance/experience in a room-scale task. Its contribution to our synthesis is the demonstration of a tangible prop as an interaction primitive for lower-limb/foot-based embodied interaction, alongside empirical evidence on when such pairing improves outcomes and practical guidance for similar applications [Bozgeyikli and Bozgeyikli, 2022].

Finally, Mixed Reality and Haptic-Based Dental Simulator for Tooth Preparation: Research, Development, and Preliminary Evaluation ($N = 34$) presents a MR haptics dental simulator and reports a preliminary face-validity evaluation with novices and skilled participants. The paper contributes an end-to-end simulator description and initial evidence regarding usability/acceptance and user experience, illustrating how proxy-based interaction and haptic instrumentation can be combined for task-specific training [Li *et al*., 2022].

These ten papers provide a high-participation cross-section of the recent literature, spanning large-scale comparisons of immersive display modalities and multisensory reconstruction, controlled studies of visuo-haptic conflict and embodiment, applied training and safety scenarios with realistic constraints, and tangible interaction primitives grounded in real-virtual pairing. This subset helps contextualize the broader corpus-level trends and challenge archetypes reported in Sections 4.2-4.6.

4.2 Timeline (2021-2025)

To better understand the overall behavior of immersive visuo-haptic illusions with real-virtual proxy pairing's technological usage, Figure 3 summarizes the temporal distribution of our screened corpus and highlights how hardware and software choices evolved across the 2021-2025 window. In total, our search retrieved 142 records, in which 75 papers (2021: 10; 2022: 16; 2023: 14; 2024: 10; 2025: 25) were retained after the Complete Screening process. This yields an approximate overall inclusion rate of 52.8% (75/142), with a clear increase in both retrieved and included papers in 2025, consistent with the recent growth of MR-capable, consumer-grade XR platforms and tooling.

Hardware trends. Across the 75 included papers, we identified 78 HMD instances because three studies reported using two different headsets in their evaluations. Overall, headset reporting is high but not universal: six papers, despite using a HMD, do not explicitly mention which model has been used in their studies, and two papers fall into a No HMD category, one in 2022 (uses a Samsung Series S6 Smartphone to show and pair virtual objects with physical props) and other one in 2024 (uses a world-fixed projector, mirroring the real environment in a real one). Among the headsets that are reported, HTC Vive devices are the most common family (28 instances), spanning HTC Vive (10), Vive Pro (11), Vive Eye (4), and Vive Pro 2 (3). Meta Quest devices form the second largest group (17 instances), with Meta Quest (1), Quest 2 (7), Quest Pro (2), and Quest 3 (7), notably concentrated in 2025 where

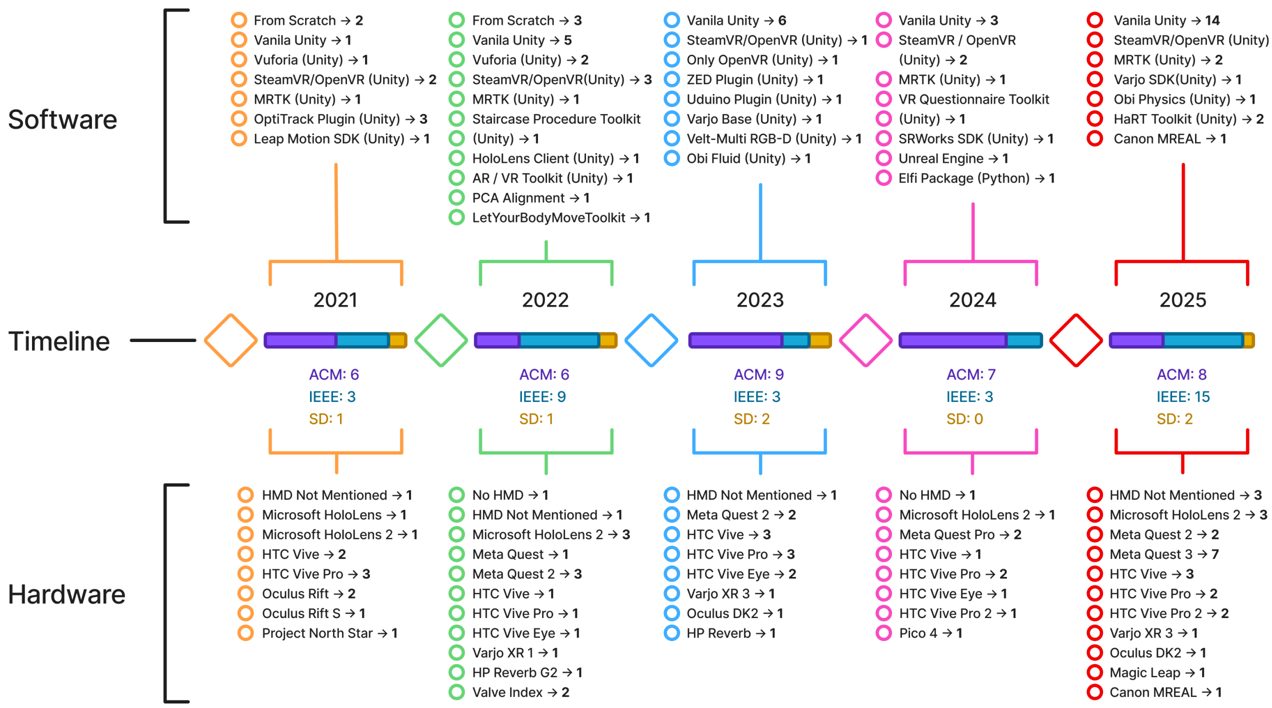

Figure 3. Timeline distributing the accepted papers by ears and research platform. It also states used software and hardware across the time span.

Quest 3 accounts for seven instances. Microsoft HoloLens appears in nine instances overall (HoloLens: 1; HoloLens 2: 8), distributed across the window with increased presence again in 2025 (three instances of HoloLens 2). The remaining headsets (16 instances) include Varjo XR-series, Oculus Rift/DK2 variants, HP Reverb (and Reverb G2), Magic Leap 2, Pico 4, Project North Star, Canon MREAL, and Valve Index.

When examining the year-by-year profile, 2021 is dominated by earlier PC-tethered headsets (in which are Vive/Vive Pro and Rift-class devices), while 2025 exhibits the strongest shift toward newer standalone MR headsets and platforms, most prominently Meta Quest 3 and renewed HoloLens 2 usage. This pattern aligns with the growing availability of commodity MR headsets and improved hand tracking, passthrough, and anchoring pipelines that facilitate real-virtual proxy pairing in deployable settings.

External device support. Most included studies rely on additional hardware beyond the HMD/controller stack. Specifically, 68 out of 75 papers use at least one external supporting device (2021: 9; 2022: 13; 2023: 14; 2024: 9; 2025: 23), while only seven papers stated that they did not depend on other auxiliary hardware than their own HMDs (2021: 1; 2022: 2; 2024: 2; 2025: 2). This underscores a key practical feature of visuo-haptic proxy pairing: even when the haptic channel is passive (physical props and surfaces), systems often require trackers, depth/RGB-D sensors, calibration fixtures, or instrumented objects to maintain robust alignment and capture interaction variables.

Software stacks and development choices. Software development is overwhelmingly author-driven in this corpus. We found no cases in which studies relied solely on an existing off-the-shelf application without author modification. Instead, most papers report custom applications built by the authors (67), based in pre-existent content (such as game engines, toolkits and SDKs) and a smaller subset report fully author-developed stacks (8). This strongly suggests that, during 2021-2025, visuo-haptic proxy pairing remains largely a systems-and-prototyping endeavor requiring bespoke pipelines rather than turnkey applications.

In terms of platforms, Unity dominates. Vanilla Unity (i.e., Unity without the named higher-level platform/toolkit categories in the dataset) is the most frequent baseline (29 total; rising to 14 in 2025 alone). Common Unity-based integrations include SteamVR/OpenVR (11) and Vuforia (3), while MR-specific tooling appears via MRTK (5). A long tail of specialized plugins and SDKs reflects the heterogeneity of tracking and sensing setups: OptiTrack plugin: 3; ZED plugin: 1; Leap Motion SDK: 1; SRWorks SDK: 1; Varjo Base/SDK: 2 total across years; HaRT toolkit: 2. Unreal Engine appears in one study (2024), and Python tooling appears once (2024). The presence of a Canon MREAL platform instance in 2025 reflects that a minority of studies still rely on specialized industrial MR stacks.

The timeline indicates a rise in publication volume in 2022, followed by a slight decrease in 2023 and 2024, however, presented a strong increase at 2025; Continued reliance on external devices for alignment and sensing; And strong convergence around Unity-based custom development. Hardware diversity remains high, but there is a visible shift from predominantly PC-tethered VR systems early in the window toward newer standalone MR headsets and mixed-reality toolchains in 2025, which is consistent with the maturation of practical pipelines for real-virtual proxy pairing.

4.3 Thematic Distribution

To make sense of the diverse solutions and study goals observed in the final corpus, we grouped papers into four recurrent thematic categories: Frameworks, Focused Experiments, Prototypes, and System Designs, as demonstrated by Figure 4. This categorization is driven by the primary contribution type reported in each paper, which may be a reusable technique/toolkit, a controlled perceptual study, a novel artifact evaluated as a proof-of-concept, or a domain-oriented system integrating multiple components. Across categories, a consistent pattern emerges: most works rely on custom author-developed applications (predominantly in Unity) and frequently combine HMD-based immersion with external sensing/actuation to maintain real-virtual alignment and to deliver visuo-haptic cues.

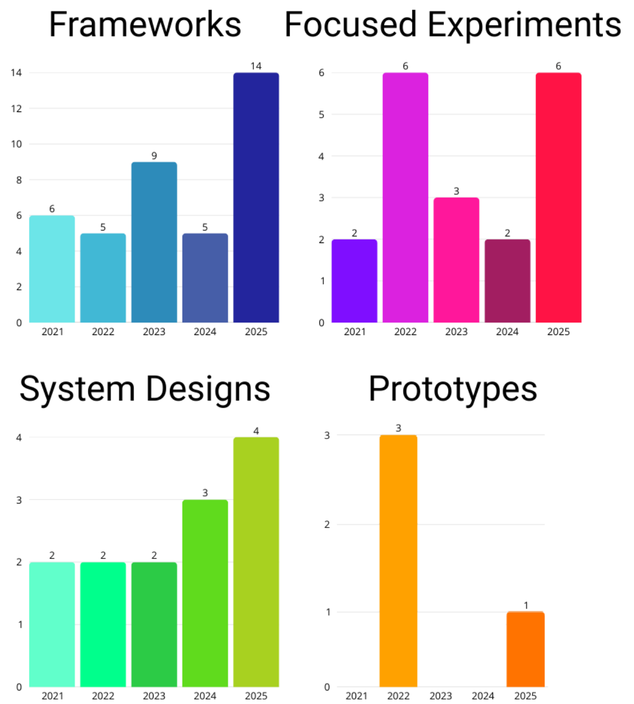

Figure 4. Thematic distribution of papers per year.

Frameworks: reusable techniques and toolkits for visuo-haptic XR. Framework papers provide generalizable methods, pipelines, and toolkits intended to be reused beyond a single scenario. In this category, several works focus on proxy construction and mapping between physical artifacts and virtual objects to improve realism and interaction quality. For instance, Feick *et al.* [2023a] formalize a modular approach for building handheld proxies that combine tactile and kinesthetic cues and evaluate them against standard controllers. Similarly, Auda *et al.* [2021b] propose a toolkit that physically delivers UI elements via quadcopters, highlighting a general approach to bringing haptics to the hand even when the virtual interface is spatially distributed. Other framework papers emphasize perceptual modulation through visuo-haptic illusions and retargeting. Examples include hand-redirection generalization [Geslain *et al.*, 2022] and control-to-display manipulations for perceived stiffness or resistance [Weiss *et al.*, 2023; Feick *et al.*, 2023b]. Another strong cluster targets multisensory enrichment and domain deployment readiness, such as in-car multisensory VR using built-in vehicle actuators [Yeo *et al.*, 2025] and tangible or wearable MR augmentation [Teng *et al.*, 2021; Cai *et al.*, 2025]. Finally, multiple framework works address interaction scalability (multiple objects, multiple users, and diverse environments), such as ownership transfer with passive props in remote collaboration [Auda *et al.*, 2021a], distraction-aware VR mappings [Tao and Lopes, 2022], or markerless tangible interaction with everyday objects [Fan *et al.*, 2025]. Overall, this category contributes the design vocabulary and reusable mechanisms that many other papers build upon: proxy pairing, retargeting, illusion shaping, and multisensory integration.

Focused Experiments: controlled evidence on perception, thresholds, and transfer. Focused experiments are characterized by hypothesis-driven, controlled studies aimed at quantifying perceptual effects, detection thresholds, or behavioral outcomes in visuo-haptic XR. A central theme is the characterization of tolerance to mismatch between physical and virtual cues. Psychophysical approaches appear repeatedly, such as interleaved staircase procedures and detection-threshold measurements for proxy-based discrepancies [Feick *et al.*, 2021, 2022]. A second theme is cross-modal dominance and expectation, where studies show that visual rendering, embodiment cues, and task context can systematically bias haptic judgments. Examples include the role of visual material cues in roughness perception under constrained physical textures [Günther *et al.*, 2022], size congruency under visuo-haptic discrepancy [Zheng *et al.*, 2024], and the influence of embodiment cues on perceptual reporting [Murmu *et al.*, 2025]. Finally, several focused experiments connect visuo-haptic manipulations to transfer, stress, or collaborative behavior. Examples include proactive transfer of motor skills enabled by hand-redirection illusions [Geslain *et al.*, 2021], physiological stress modulation using physical elevation plus virtual height exposure [Zhu *et al.*, 2023], and the impact of haptic/visual inconsistencies on collaborative performance [Weiss *et al.*, 2025a]. Collectively, this category provides the quantitative backbone for understanding when and why proxy pairing and illusion techniques work.

Prototypes: proof-of-concept artifacts and feasibility-driven evaluations. Prototype papers typically introduce a new device, sensing approach, or integrated artifact and validate feasibility with smaller-scale evaluations. These works often address practical bottlenecks in proxy pairing, such as robust touch sensing on ad hoc surfaces [Mollyn *et al.*, 2025], portable shape/softness rendering [Zhu *et al.*, 2022], or mixed environments that combine real/virtual objects with haptic devices [Iwamoto, 2022]. In our corpus, prototypes commonly serve as technology probes: they explore what becomes possible if a key sensing/actuation limitation is relaxed, and they provide early evidence (accuracy, usability, or perception) that motivates later, more complete systems.

System Designs: integrated solutions targeting applied tasks and deployment constraints. System design papers report end-to-end systems that integrate hardware, software, and interaction design to support an applied scenario (training, assessment, manipulation, or cultural experiences), typically including evaluation with representative users or experts. Several works focus on training and assessment in high-stakes contexts: fire drill training with tangible props and hybrid setups [Grabowski, 2021; Kang *et al*., 2024], medical skill training using MR manikins and tools [Uhl *et al*., 2023; Li *et al*., 2022], and operational assessment with passive haptics [Dufresne *et al*., 2024]. Other system designs emphasize interaction paradigms that depend on reliable proxy mapping and instrumentation, such as opportunistic tangible proxy selection and hand-object mapping in AR [Feick *et al*., 2023b], mobile devices as haptic interfaces [Stellmacher *et al*., 2024], and ungrounded impact-force rendering [Chen *et al*., 2023]. These systems foreground the engineering constraints that determine whether proxy pairing is viable outside controlled demonstrations: calibration effort, robustness to drift, multiuser coordination, and the trade-off between fidelity and complexity.

Synthesis across categories. Across the four categories, the literature reveals a coherent progression: Frameworks define reusable building blocks (proxy pairing, retargeting, multisensory pipelines), Focused Experiments quantify perceptual tolerances and behavioral effects, Prototypes test feasibility of new sensing/actuation primitives, and System Designs integrate these elements to meet domain constraints (training, assessment, telepresence, and applied interaction). This thematic structure helps contextualize the remaining analyses in our survey by clarifying what the field is building (systems and devices) and how it validates them (controlled perception studies versus deployment-driven evaluations).

4.4 Representative Archetypes

From the observed trends, we derived ten recurring representative archetypes, presented in Figure 5, that captures how passive/visuo-haptic systems are typically framed, implemented, and validated in the literature: Textures, Weight, Shape, Depth, General Touch, Environmental, Matched Reality Construction, Simulators, Tangible UI, and Sensory/Emotional Experiences. Rather than describing individual systems in isolation, these archetypes summarize the dominant design intents (what is being manipulated or supported), the common technical mechanisms (how the effect is produced), and the evaluation styles (how success is measured), while pointing to representative works.

Textures. Texture-oriented works manipulate perceived surface qualities such as roughness, slipperiness, or compliance using combinations of limited physical stimuli and strong visual priors. A representative example is the work of Günther *et al*. [2022], where a wearable stroking device and a small set of sandpapers are paired with multiple virtual material renderings; results suggest that only coarse physical roughness levels can produce convincing texture variation when visuals are well-matched, and that physical strokes outperform vibrotactile baselines for realism and enjoyment. Texture is also expanded beyond classical roughness into mechanical properties: Weiss *et al*. [2023] show that control-to-display (C/D) gain manipulations during pressing can shift perceived stiffness within undetected ranges, yielding practical mapping functions for designers. Wearable MR texture devices further emphasize dexterity preservation and perceptual tuning, as in the works from Teng *et al*. [2021] and Cai *et al*. [2025], the latter explicitly demonstrating that roughness can be both increased and decreased (via vibrotactile stimulation vs. reduced fingerpad contact). Finally, texture archetypes also include locomotion-relevant cues: Degraen *et al*. [2024] demonstrate that sole materials with different friction coefficients map consistently to perceived slipperiness and that visuo-haptic correspondence increases realism. Overall, texture studies show a repeated strategy: use a small, controllable set of physical stimuli and let visual context, and congruency, do the heavy lifting.

Weight. Weight archetypes target perceived heaviness, inertia, or internal center-of-mass (CoM) through actuation, retargeting, or perceptual priors. Faltaous *et al*. [2022] combine passive props with EMS to bias muscular effort, reporting the strongest weight increases with biceps/triceps stimulation and suggesting promising multi-phase actuation strategies. Complementary to muscle-based approaches, Sagheb *et al*. [2023] relocate an internal active mass to follow a simulated fluid's CoM across vessel shapes, yielding congruent, low-latency weight-shift sensations for fluid handling. Retargeting and passive-dynamic methods extend this archetype to virtual weight distribution in hand-held objects; Zenner *et al*. [2021] report stronger perceivable CoM shifts and tolerance to larger spatial dislocations than either technique alone. Finally, weight archetypes include the role of learned priors and transfer: Nam *et al*. [2025] highlight internal-model robustness as the main driver of virtual-to-real transfer, with presence acting as a secondary amplifier in some conditions. Together, these works show that weight in XR is rarely a single cue; it emerges from effort, mass distribution dynamics, and expectation/learning.

Shape. Shape archetypes address geometry consistency between physical proxies and virtual objects, including tolerance to mismatch, modality dominance, and synthesis/mapping techniques. Kang *et al*. [2023] provide a strong representative result: for global shape discrimination under conflict, touch dominates vision, and users rely on haptics in incongruent trials. In contrast, other works highlight situations where visual dominance biases judgments: Zheng *et al*. [2024] report substantial tolerance to tactile size distortions when cues are synchronized, often with visual information driving overestimation. A recurring technical approach is to reshape the virtual object rather than forcing strict physical matching; Fukumoto *et al*. [2022] propose chimera-object synthesis that merges proxy and target shapes to reduce discomfort and improve visuo-haptic consistency. Shape effects are also mediated by environment and display conditions: Kahl *et al*. [2022] show that brighter indoor lighting widens acceptable size deviations but can reduce presence and usability. Finally, shape archetypes extend to collaborative contexts, where inconsistencies become interactional rather than purely percep-

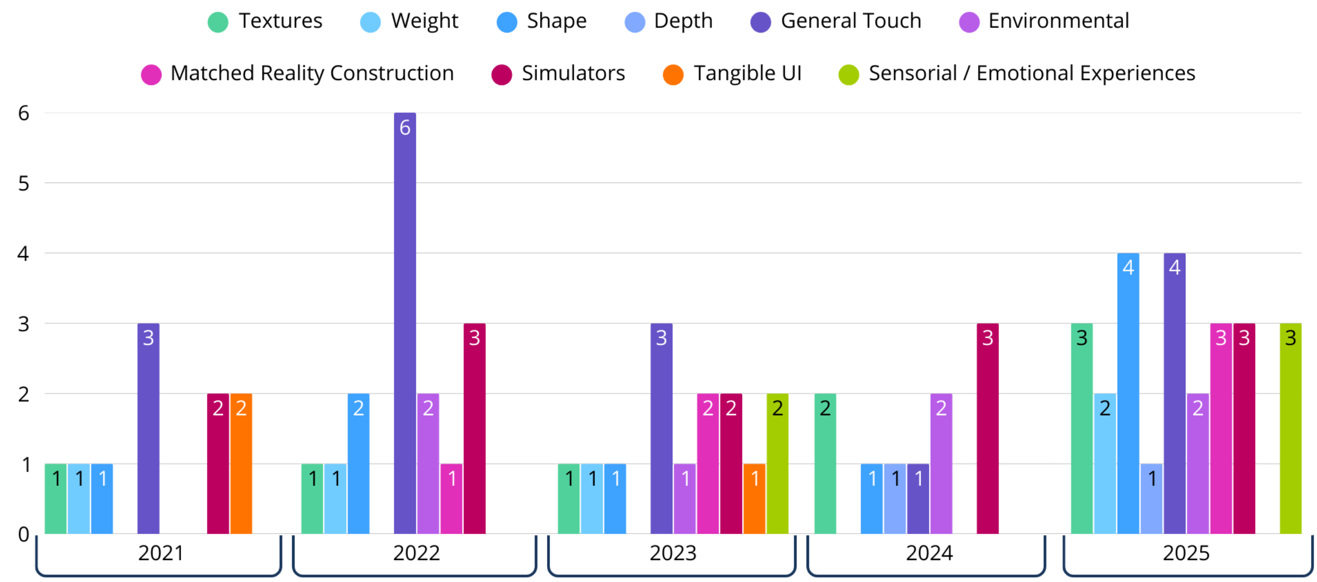

Figure 5. Representative archetypes per year.

tual: Weiss *et al*. [2025a] report that shape/size mismatches disrupt collaborative performance and experience, prompting partial adaptation strategies but not full compensation. Overall, shape archetypes emphasize consistency management, via synthesis, constraints, or context and task dependence (discrimination vs. manipulation vs. collaboration).

Depth. Depth archetypes focus on near-surface interaction states such as hover distance, touch onset, and proprioceptive/visual sensitivity-often because depth ambiguity is a primary cause of misalignment and reduced realism in MR. Mollyn *et al*. [2025]'s work is representative of a sensing-driven depth archetype: by combining IR illumination geometry with an egocentric camera and lightweight ML, it achieves high touch segmentation accuracy and mm-scale hover estimation on everyday surfaces, enabling ad-hoc surface input without instrumentation. Complementarily, Lebrun *et al*. [2024] formalize how individual visual and proprioceptive precision can be measured in VR and used to categorize users, motivating adaptive models for depth-related illusion detectability. Together, these depth works show how accurate touch/hover estimation and user-specific sensitivity models reduce mismatch and support robust interaction design.

General Touch. General-touch archetypes capture a broad class of systems that do not target a single physical attribute (like texture or weight) but instead study or enable consistent contact, manipulation, and interaction through redirection, passive proxies, or ungrounded actuation. Psychophysical threshold studies define the limits of acceptable discrepancy [Feick *et al*., 2021, 2022], which provide conservative bounds and highlight the effects of movement constraints and individual differences. General touch also includes interaction techniques for specific control primitives: Feick *et al*. [2023b] demonstrate rotational C/D manipulation with a physical knob to simulate resistance, enabling more realistic room-scale interaction. For impulsive events, Chen *et al*. [2023] show a hardware module that amplifies air-jet forces to produce stronger, more immediate impact sensations preferred over baselines. A parallel line targets embodiment cues and calibration: Murmu *et al*. [2025] show that hand-avatar visibility can improve perceptual calibration and reduce overestimation, especially when present during reporting. In summary, general-touch operationalize touch as contact consistency plus interaction reliability across diverse tasks and different tactile stimuli.

Environmental. Environmental archetypes integrate real-world stimuli or context cues (sound, airflow, heat, vibration, space geometry, co-located artifacts) to reduce breaks in presence and strengthen ecological validity. Tao and Lopes [2022] map real stimuli (detected or sensed) into coherent virtual counterparts, reporting increased presence and reduced perceived distraction relative to baseline strategies such as noise-canceling. Environmental can also mean co-registration fidelity: Mubarrat *et al*. [2024] characterize spatial and temporal alignment errors in a physics-based prop interaction system, supporting its suitability for applications requiring high-fidelity synchronization. At the evaluation level, Effect of perceptive enrichment on the efficiency of simulated contexts: Gouton *et al*. [2023] show that enrichment improves scenario credibility (VR approaching real-world credibility) while leaving product-evaluation performance largely unchanged, illustrating that environmental realism and task outcomes may decouple. Finally, environmental archetypes can include social and autobiographical place reconstructions: Pillis [2025] report heightened emotional engagement and transport to the past, positioning environment reconstruction as a primary experiential mechanism.

Matched Reality Construction. Matched-reality construction archetypes explicitly address how to build and maintain mappings between physical proxies and virtual objects at scale. These works contribute toolkits, selection/recommendation strategies, and retargeting pipelines that reduce authoring cost and interaction overhead. Feick *et al*. [2023a]

exemplify modular proxy construction with multiple plates (shape/texture/weight/tracking), reporting higher realism and enjoyment than controllers while revealing trade-offs (such as increased size/weight). Jain *et al*. [2023] advance beyond geometry matching by recommending everyday proxies based on desired hand-object interactions, and demonstrate competitive pose accuracy and positive usability across tutoring, games, and smart-home scenarios. In VR retargeting pipelines, Sun *et al*. [2025] and Liu *et al*. [2025] show how optimizing reset points and selecting a small set of important proxies reduces resets and travel while improving comfort and workload. Beever and John [2022] developed a construction archetype for entertainment, using consumer devices to scan spaces and import objects as passive haptics for room-scale SR play. Overall, matched-reality construction turns isolated demos into authorable, repeatable, scalable experiences.

Simulators. Simulator archetypes are end-to-end training, rehearsal, or assessment systems where passive/visuo-haptic alignment is a means to improve skill transfer, fidelity, or decision making. Several simulators integrate tangible props or manikins: Uhl *et al*. [2023]'s work overlays a virtual patient on a physical manikin and reports high physical/self-presence and technology acceptance with physiological arousal during scenarios. Yigitbas *et al*. [2022] show that AR/VR can increase engagement and reduce workload, while traditional training still yields higher exam scores-illustrating that simulator presence does not automatically translate to superior learning outcomes. Domain simulators also target operational assessment: Dufresne *et al*. [2024] report that passive proxies increase embodiment and perceived fidelity for ConOps assessment, but also raise workload. Task-specific rehearsal systems include Cao *et al*. [2025]'s work, where real instruments and patient-specific models support tangible rehearsal with strong performance and surgeon-reported value. Beyond training, simulator archetypes include creative practice and cultural sport: Fender *et al*. [2023] combine real brush haptics with persistent VR rendering and is preferred for hedonic/haptic qualities, and He *et al*. [2024] show higher exertion and convincing rowing with IMU/exertion controllers in a digital twin campus. Finally, simulators can be used as research instruments: Geslain *et al*. [2021] demonstrate positive motor-skill transfer from VR training using hand redirection, supporting simulators that trade hardware complexity for perceptual techniques.

Tangible UI. Tangible-UI archetypes emphasize physical controls and tangible interaction metaphors as the primary interface, often prioritizing engagement, learnability, and embodied manipulation. Auda *et al*. [2021b] provide a canonical example: quadcopters deliver tangible UI widgets (button/knob/slider/etc.) into the user's reachable space, increasing engagement and perceived haptics while acknowledging limitations versus standard controllers. Kawazoe *et al*. [2021] extend tangible UI toward wearable augmentation, capturing touch vibrations and returning tactile/audio echoes that increase responsiveness and agency in free-hand interactions. At the interaction-technique level, Matthews *et al*. [2023b] study how reset methods reduce required resets and improve efficiency when aligning sequences of UI interactions to a physical interface. As a result, tangible UIs often combine physical affordances with tracking/retargeting strategies to preserve coherence during extended interaction sequences.

Sensory/Emotional Experiences. This archetype foregrounds affective, therapeutic, or social outcomes where touch and multisensory cues are used to shape emotional state, bonding, or trust. Sensing Nature represents a relaxation-focused installation combining a fabric-based haptic tree, spatial audio, and vibrations, reporting increased presence and positive emotional effects (calm, restored energy) alongside stronger connectedness toward nature. Clavelin *et al*. [2023] demonstrate that synchronized visual manipulations can increase perceived discomfort (virtual pain) without changing the underlying nociceptive stimulus, positioning perceptual alignment as a lever for affective intensity. Social affect and bonding are central in Fouad *et al*. [2025]'s work, where passive haptics (a plush toy) primarily drive social presence and bonding while active vibrotactile gloves increase behavioral engagement and can compensate when passive embodiment is absent. Finally, trust and acceptance in clinical contexts appear in Song *et al*. [2025]'s work, where immersive visualization choices modulate comfort and acceptance. Across this archetype, success is measured less by perceptual thresholds and more by experienced meaning: comfort, trust, bonding, and emotional transport.

Cross-archetype observations. Across archetypes, three broad patterns recur in the provided corpus: authors overwhelmingly rely on custom-built software (mostly Unity-based) to tightly synchronize rendering, sensing, and actuation; many contributions combine passive physical artifacts with perceptual techniques (C/D manipulations, mapping, or synthesis) to expand what a limited set of props can represent; and evaluations range from psychophysical threshold studies (Textures/Shape/General Touch/Depth) to domain validation (Simulators/Environmental/Sensory), reflecting a field that alternates between establishing perceptual operating envelopes and integrating them into applied systems.

4.5 Challenges

According to our review processes, accepted studies reported limitations and practical hurdles that concentrates around eleven recurring challenge categories: Misalignments, Tracking Issues, Latency Issues, Bug Issues, Limited Generalisability, Hardware Limitations, Safety Limitations, Limited Participant Samples, Cybersickness/Presence, Ergonomics Issues, and Experimental Setup Challenges, as shown in Figure 6, which distributes those topics across the past five years. While these challenges often co-occur, the literature reveals stable patterns in how they constrain fidelity, scalability, and deployment beyond lab settings.

Misalignments. Physical-virtual misalignment is among the most pervasive constraints, spanning passive-haptics co-registration, proxy placement, avatar alignment, and real-to-virtual mapping across VR/MR setups. Even when small,

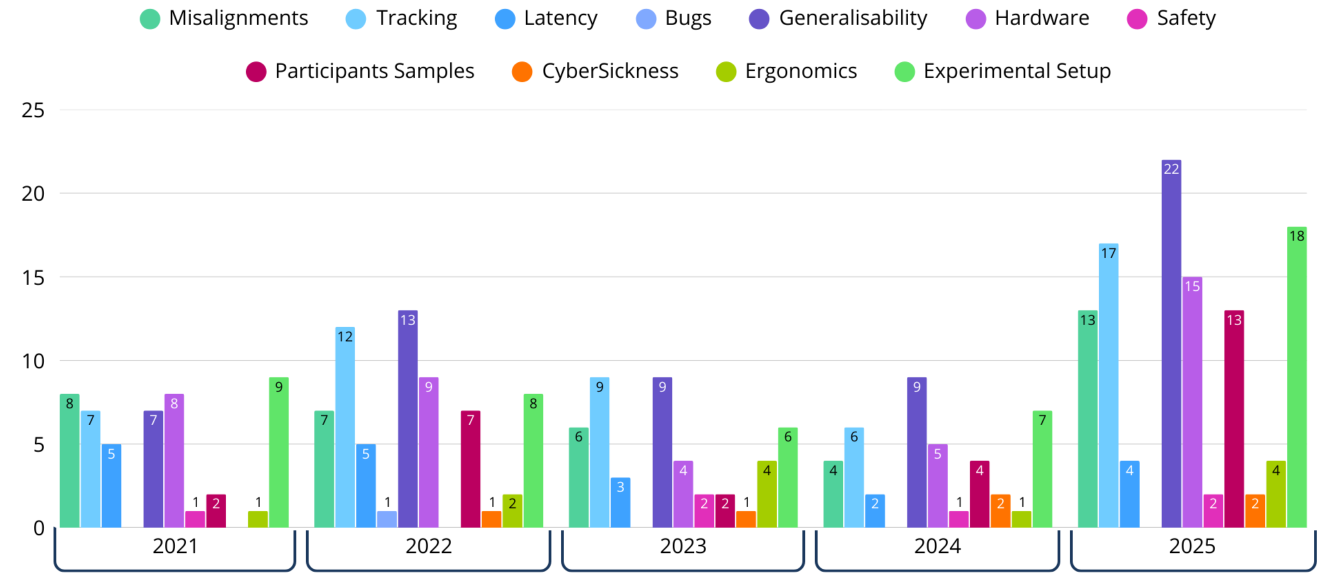

Figure 6. Challenges Graphic presenting the reported limitations within different years timeframes.

misalignments can reduce perceived realism, degrade performance, and introduce measurement artifacts. Studies grounded in strict physical co-location frequently highlight calibration sensitivity, drift, and representation mismatches (for instance, a case where props do not match their virtual counterparts under motion) [Mubarrat *et al.*, 2024; Grießel *et al.*, 2025]. Misalignments also appear as a limiting factor in interactive proxy systems and retargeting pipelines where repeated resets or proxy switching can accumulate error [Sun *et al.*, 2025; Liu *et al.*, 2025]. In AR/MR, misalignments are exacerbated by optical see-through registration constraints and environmental lighting [Kahl *et al.*, 2022] and teleoperation-style mock-up overlays [Xu *et al.*, 2025]. Finally, even highly instrumented experiences note that alignment demands can shape interpretation of outcomes, as can be seen with performance penalties when alignment is imperfect [Cherix *et al.*, 2025].

Tracking Issues. Tracking robustness remains a core bottleneck across systems, regardless of modality (inside-out, outside-in, optical markers, depth cameras, or hybrid). Many prototypes depend on multi-sensor stacks (Vive trackers, OptiTrack, depth cameras, hand tracking), which increases occlusion risk, calibration complexity, and sensitivity to lighting/material conditions. Examples include systems reliant on external tracking for precise proxy alignment [Günther *et al.*, 2022; Geslain *et al.*, 2021] and toolkits that require stable tracker attachment and consistent geometry [Feick *et al.*, 2023a]. Markerless or ad-hoc interaction further stresses tracking [Mollyn *et al.*, 2025; Fan *et al.*, 2025]. Overall, the reported trend is that tracking limitations are not merely engineering concerns, they directly bound illusion thresholds, interaction precision, and validity of behavioral/physiological measures.

Latency Issues. Latency is repeatedly noted as a fidelity limiter, especially in closed-loop multisensory mappings (audio/wind/heat integration, active haptics, or remote/proxy-mediated interaction). Pipeline latency can arise from sensing (external microphones/cameras), transmission (OSC/network), rendering/physics, and actuator control. This is particularly visible in real-world stimulus mapping [Tao and Lopes, 2022], mobile/robotic or mechanically mediated interaction [Auda *et al.*, 2021b; Chen *et al.*, 2023], and AR setups where visual-haptic asynchrony influences perception [Normand *et al.*, 2024]. Latency also constrains retargeting and reset techniques that must maintain tight spatial-temporal coupling during sequences of interactions [Matthews *et al.*, 2023b]. In sum, latency is commonly treated as a first-order determinant of perceptual plausibility (and not only a QoS metric), because small delays can be reinterpreted as inconsistency between modalities.

Bug Issues. Bug-related limitations are comparatively rarely foregrounded, but when present they are consequential because these systems integrate heterogeneous components: embedded devices, signal processing, networking, game engines. For example, Tao and Lopes [2022] explicitly report bug issues alongside latency and tracking concerns, illustrating how software instability can confound both user experience and experimental reliability when multiple subsystems must remain synchronized.

Limited Generalisability. A dominant theme is that many contributions are validated in narrowly scoped tasks, controlled environments, and bespoke hardware configurations, limiting direct transfer to broader applications. This appears in illusion work where thresholds depend on task geometry and constraints [Feick *et al.*, 2021; Geslain *et al.*, 2022], in tangible proxy toolkits tuned to specific object sets [Feick *et al.*, 2023a], and in domain-focused training systems where realism requirements are context-dependent [Yigitbas *et al.*, 2022; Kang *et al.*, 2024]. Even when systems show strong within-study benefits, authors frequently note that scaling to different spaces, users, tasks, and device ecosystems remains open. The pattern suggests that generalisability is currently constrained by hardware specialization, calibration/authoring overhead, and the lack of standardized benchmarks that span multiple XR modalities.

Hardware Limitations. Hardware constraints are repeatedly cited, including actuator force and bandwidth, device bulk, power/battery limitations, sensing coverage (FOV), and HMD display limitations (resolution, passthrough quality). Wearables and handheld props highlight trade-offs between realism and practicality, such as device weight/size tradeoffs [Feick *et al.*, 2023a] and constraints in tactile/texture devices [Teng *et al.*, 2021; Cai *et al.*, 2025]. Systems that depend on auxiliary cameras or depth sensors also inherit their limitations (mounting, bandwidth, lighting sensitivity), as in EclipseTouch: Touch Segmentation on Ad Hoc Surfaces using Worn Infrared Shadow Casting and TangiAR: Markerless Tangible Input for Immersive Augmented Reality with Everyday Objects. In applied domains, limitations include insufficient realism of force feedback or missing modalities (for example, a soft-tissue forces), noted in Mixed Reality and Haptic-Based Dental Simulator for Tooth Preparation: Research, Development, and Preliminary Evaluation. Overall, hardware limitations frequently define the ceiling of achievable fidelity more than the interaction concept itself.

Safety Limitations. Safety emerges as a distinct constraint when systems introduce moving machinery, vehicles, elevated platforms, or high-energy actuation. Quadcopter-based tangible interaction explicitly raises safety limitations and operational constraints [Auda *et al.*, 2021b]. In-car XR introduces safety-critical concerns in both deployment and evaluation [Yeo *et al.*, 2025]. Physical elevation paradigms for stress induction require careful risk management [Zhu *et al.*, 2023]. Robotics-augmented XR similarly demands safety boundaries and conservative interaction design, as implied by frameworks that physically align robot-mounted props or encountered-type devices [Senk *et al.*, 2025]. These works collectively show that safety is not peripheral: it can dictate interaction vocabulary, experiment duration, and participant eligibility.

Limited Participant Samples. Participant limitations are widespread, particularly in hardware-heavy or early-stage system papers where recruiting and running controlled sessions is costly. Many studies rely on small-to-moderate samples (often in the 5-30 range) for lab experiments and prototype evaluations, which is frequently acknowledged as limiting external validity, e.g., Uhl *et al.* [2023]; Mollyn *et al.* [2025]; Fukumoto *et al.* [2022]. Some works do include large-scale evaluations, see Gouton *et al.* [2023], but the broader picture is that sample size is often bounded by experimental setup complexity and hardware availability, and can present narrow participant profile variability, despite the large sample, as shown in Pooladvand *et al.* [2025] study, in which all users were students and none worker was included. This reinforces the need for reusable protocols and shared platforms to reduce evaluation costs.

Cybersickness and Presence. Cybersickness is periodically reported as an explicit concern, especially in experiences involving locomotion, exertion, or non-standard display conditions. Exertion-based rowing highlights simulator sickness as a design risk even when physiological engagement is the goal [He *et al.*, 2024]. Whole-body interaction and world-fixed or projection-based setups also note cybersickness/presence considerations [Bozgeyikli and Bozgeyikli, 2022; Bozgeyikli, 2024]. Several systems frame cybersickness as intertwined with presence: interventions that improve embodiment or realism can also raise workload or discomfort if alignment is imperfect, or if interaction induces atypical motion cues, as suggested by Murmu *et al.* [2025], and context-enrichment comparisons in Gouton *et al.* [2023]. Overall, cybersickness is not uniformly dominant across papers, but it repeatedly surfaces as a gating factor for sustained, deployable XR scenarios.

Ergonomics Issues. Ergonomic constraints appear in the form of fatigue, device weight distribution, restrictive form factors, and discomfort from wearables/attachments or complex props. Toolkit-style and modular devices report tradeoffs between realism and bulk, where plates/accessories increase perceived realism but also weight/size in Feick *et al.* [2023a]. Systems using body-worn stimulation or larger carried setups highlight comfort and encumbrance, such as EMS components in Faltaous *et al.* [2022]. Hand- and finger-worn devices also emphasize dexterity and comfort constraints [Teng *et al.*, 2021; Cai *et al.*, 2025]. For high-energy handheld devices, ergonomics includes handling and sustained use feasibility [Chen *et al.*, 2023]. Collectively, the literature suggests ergonomic feasibility is a central prerequisite for moving from compelling demonstrations to repeated-use applications (training, therapy, industrial work).

Experimental Setup Challenges. Finally, many papers emphasize the complexity and fragility of experimental setups: multi-device calibration, marker placement, environmental constraints (space, lighting), safety protocols, and long condition matrices. For example, highly parameterized studies require extensive condition control and synchronization [Günther *et al.*, 2022; Feick *et al.*, 2022]. Systems that integrate external sensors and stimulus actuators must manage networking, timing, and reproducibility [Tao and Lopes, 2022]. Installation-style experiences also require careful staging and instrumentation [Zhang *et al.*, 2025]. Training environments frequently depend on elaborate props and staged spaces (for example, manikins, tracked tools, or hybrid AR/VR flows), as shown in Yigitbas *et al.* [2022] and Kang *et al.* [2024]. These repeated reports reinforce that evaluation difficulty is itself a research barrier: setup overhead limits sample size, replication, and cross-study comparability.

The reviewed challenges suggest that advancing tactile and passive-haptics XR requires progress on robust tracking and co-registration under realistic conditions, low-latency multimodal pipelines, ergonomically viable devices, safety-aware interaction design for mobile/robotic/vehicular contexts, and evaluation protocols that reduce setup overhead and improve generalisability. These constraints shape not only engineering feasibility but also how confidently results can transfer to real-world applications and longitudinal use.

4.6 Experiments

Only the [Senk *et al.*, 2025] study performs no experiments, which means that 74 experiments were reported across the selected papers. To summarize the empirical evidence base, we categorize experiments into participant-sample tiers and report their distribution, as Figure 7 presents. The tiering follows the rationale of diminishing returns in human-subject evaluations discussed by Faulkner [2003]: in usability-style studies, smaller samples can already reveal a substantial portion of issues, while additional participants tend to yield progressively fewer new findings. Consequently, tiers provide a compact way to communicate whether an experiment is closer to an exploratory/pilot scale (few participants), a typical HCI/XR lab evaluation, or a larger-scale validation.

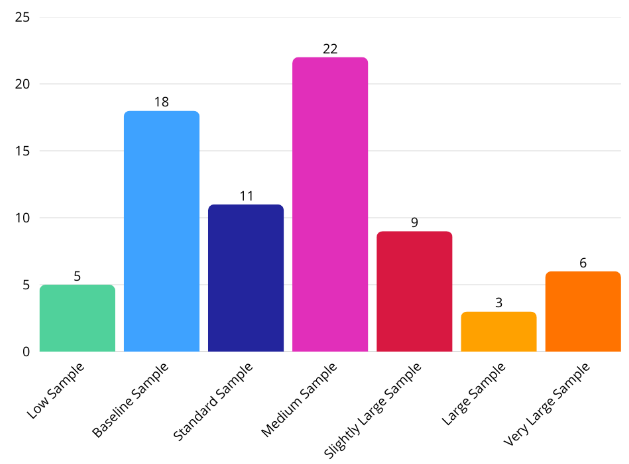

Figure 7. Experiments chart.

Based on Faulkner's discussion of sample-size trade-offs [Faulkner, 2003], we interpret tiers as follows:

- **Very Low Sample** ($N \leq 4$): typically pilot/feasibility, insufficient for stable quantitative inference;
- **Low Sample** ($5 \leq N \leq 9$): early validation where qualitative insights are expected but statistical power is limited;
- **Baseline Sample** ($10 \leq N \leq 14$): common minimum for controlled XR/HCI studies, enabling more reliable central tendencies and basic comparisons;
- **Standard Sample** ($15 \leq N \leq 19$): improved robustness for within-/between-subject designs;
- **Medium Sample** ($20 \leq N \leq 29$): stronger inferential stability and better coverage of individual differences;
- **Slightly Large Sample** ($30 \leq N \leq 39$);
- **Large Sample** ($40 \leq N \leq 49$);
- **Very Large Sample** ($N \geq 50$): broader validation, typically enabling finer subgroup analyses and more confidence in general patterns.

Table 4 reports the experiment counts per tier (also visualized in Figure 7). Overall, most experiments fall in the Baseline-Medium range ($10 \leq N \leq 29$): 1/74 experiments (68.9%). A smaller but meaningful portion uses $N \geq 30$ (18/74 experiments, 24.3%), while only 5/74 experiments (6.8%) are in the $5 \leq N \leq 9$ tier; no experiment in this corpus used $N \leq 4$.

Table 4. Experiment matrix by participant-sample tier (total = 74). Tiers defined following the sample-size rationale discussed [Faulkner, 2003]

Tier (participants N)	**# Experiments**	**Share**
Very Low Sample ($N \leq 4$)	0	0.0%
Low Sample ($5 \leq N \leq 9$)	5	6.8%
Baseline Sample ($10 \leq N \leq 14$)	18	24.3%
Standard Sample ($15 \leq N \leq 19$)	11	14.9%
Medium Sample ($20 \leq N \leq 29$)	22	29.7%
Slightly Large Sample ($30 \leq N \leq 39$)	9	12.2%
Large Sample ($40 \leq N \leq 49$)	3	4.1%
Very Large Sample ($N \geq 50$)	6	8.1%

The dominance of 10-29 participant experiments suggests that the literature largely emphasizes controlled lab validation where both quantitative comparisons and qualitative feedback are feasible, but where generalization to broader populations can still be constrained by sample size. Conversely, the presence of $N \geq 30$ experiments indicates that a non-trivial subset of studies pursue larger validations, supporting more stable aggregate conclusions and (when applicable) subgroup analyses. Reporting results with these tiers helps the reader quickly contextualize whether a finding stems from exploratory evidence or from broader empirical coverage, consistent with the sample-size trade-offs emphasized by Faulkner [2003].

5 Discussion

This section synthesizes the evidence presented in the previous results to answer the study's guiding questions introduced in Section 1. Rather than reiterating individual paper outcomes, we discuss the broader patterns that emerge from the reviewed corpus, connecting

- *what* kinds of real-virtual object pairing solutions are being built and studied,
- *how* enabling technologies (hardware/software) have evolved over time, and
- *where* the field still faces persistent barriers and promising opportunities.

A limitation of this review concerns the search terminology. The query was designed to identify XR studies involving haptic or visuo-haptic interaction grounded in real-virtual spatial pairing; however, the field uses heterogeneous and partially overlapping terminology. Terms such as "passive haptics", "encountered haptics", "pseudo-haptics", "visuotactile interaction", "tangible interaction", "retargeting", and "remapping" are commonly used to describe mechanisms that are relevant to our scope, but they were not all explicitly included in the initial Boolean query. As a result, some relevant studies may have been missed, especially if they did not use the terms "haptic feedback", "force feedback", or "visuo-haptic" in searchable metadata. This limitation means that the corpus should be interpreted as a focused synthesis of XR real-virtual proxy pairing studies retrieved through a reproducible query and screening protocol, rather than as a complete census of all XR haptics terminology. Future updates of this survey should adopt a broader query vocabulary and may combine database searches with backward and forward snowballing to improve recall across terminology clusters.

Another limitation concerns the use of ChatGPT as an auxiliary tool during data extraction. Although the tool helped standardize the initial extraction format and locate candidate sections for each field, large language models are not fully reliable instruments for systematic literature reviews. They may omit relevant information, misinterpret methodological details, hallucinate unsupported claims, or provide incomplete references to the source text. To mitigate these risks, ChatGPT was not used to make inclusion/exclusion decisions, and its outputs were not accepted as final data. All extracted entries were manually checked against the original PDFs by two coauthors, and disagreements with the tool were resolved by consensus based on the source paper. Nevertheless, the process was not equivalent to a fully independent double-extraction protocol, and future updates of this survey should consider independent extraction by at least two reviewers per paper, inter-rater agreement reporting, and publication of the extraction forms and prompts as supplementary material.

Although we discuss lightweight and deployable approaches, we did not apply a formal quantitative complexity score to rank the included systems. The interpretation is based on reported implementation characteristics such as hardware dependence, proxy type, tracking setup, software stack, and calibration burden. Future work could extend this survey by defining a dedicated deployability index for visuo-haptic XR systems.

Throughout, we contrast these observed patterns with the assumptions and recommendations embedded in the proposed tutorial workflow, highlighting where current practice aligns with the tutorial and where the literature suggests additional guidance, constraints, or alternative pathways are needed.

5.1 Research Question 1 (Real-Virtual Pairing Solutions)

Across the reviewed studies, real-virtual pairing is consistently operationalized as the co-existence of physical referent (object, surface, tool, body segment, or environment feature) and virtual counterpart (rendered object/effect) whose perceived realism, usability, or meaning depends on maintaining a coherent cross-domain relationship. In practice, the corpus reveals a set of recurring solution niches with application goals and contexts, and pairing archetypes with technical patterns for achieving the pairing. Below, we synthesize what appears frequently, what appears occasionally, and what is largely absent from the accepted studies, and how these patterns vary over time.

Niches observed in the corpus. A substantial portion of the literature uses pairing as a means to increase interaction realism and embodiment through tangible proxies and passive haptics. This niche spans creative tasks (such as painting with a real wet brush and proxy canvas [Fender *et al.*, 2023]), everyday-object interaction and play [Huard *et al.*, 2022]; [Bozgeyikli and Bozgeyikli, 2022], and social/affective interaction [Fouad *et al.*, 2025]. A second major niche is training, rehearsal, and assessment, where pairing is used to preserve tool-use fidelity or ecological validity: medical and first-responder contexts [Uhl *et al.*, 2023; Yigitbas *et al.*, 2022; Schild *et al.*, 2022; Kang *et al.*, 2024; Cao *et al.*, 2025] and workplace/ergonomic assessment [Grießel *et al.*, 2025]. A third niche focuses on perception and psychophysics, explicitly quantifying detection thresholds and perceptual biases when physical and virtual cues diverge, as happend in hand-redirection and C/D-ratio manipulations [Geslain *et al.*, 2021; Feick *et al.*, 2021, 2022; Geslain *et al.*, 2022]; and in stiffness/texture-related manipulations [Weiss *et al.*, 2023; Normand *et al.*, 2024]. Finally, a smaller but notable niche explores multisensory and contextual realism beyond touch, pairing real stimuli (wind/heat/sound/vibration) or real environments (vehicles, installations, reconstructed places) to coherent virtual events [Tao and Lopes, 2022; Yeo *et al.*, 2025; Zhang *et al.*, 2025; Pillis, 2025].

Independent of niche, we observe four pairing archetypes that repeatedly appear:

1. **Co-located tangible proxies (tracked passive haptics).** Here, a physical object (prop, tool, surface, or wearable) is tracked and registered to a virtual counterpart, enabling contact where the user expects it. This is the dominant archetype in social/play and training contexts [Huard *et al.*, 2022; Bozgeyikli and Bozgeyikli, 2022; Auda *et al.*, 2021a; Harley *et al.*, 2021; Uhl *et al.*, 2023; Kang *et al.*, 2024] and multiple medical rehearsal/training systems (2022-2025). In these works, the pairing success is primarily governed by registration fidelity (pose accuracy) and interaction design (what tasks the proxy can credibly support).
2. **Retargeting and multiplexing (many virtual objects per few physical proxies).** This archetype deliberately relaxes strict one-to-one mapping by using redirection or importance-based placement so that a limited set of physical props can represent a larger virtual set. Representative examples include haptic retargeting pipelines and reset optimization [Sun *et al.*, 2025; Liu *et al.*, 2025; Matthews *et al.*, 2023b] and chimera/synthesis approaches that merge proxy and target shapes to improve consistency [Fukumoto *et al.*, 2022]. This archetype is especially aligned with scalable deployments where prop count and setup time must be minimized.
3. **Visuo-haptic illusions and controlled discrepancy (psychophysical pairing).** In these studies, the pairing is intentionally imperfect but constrained to remain undetected or acceptable. The core mechanism is often C/D-ratio manipulation, redirection, or remapping between sensed motion/contact and rendered motion/contact. This archetype underpins much of the threshold-focused literature [Geslain *et al.*, 2021; Feick *et al.*, 2021, 2022; Geslain *et al.*, 2022; Feick *et al.*, 2023b; Weiss *et al.*, 2023, 2025b]. It also appears in size/weight perception work where pairing quality is evaluated through perceptual outcomes rather than task performance alone [Zheng *et al.*, 2024; Nam *et al.*, 2025].
4. **Opportunistic and markerless pairing with everyday objects/surfaces.** A growing subset of studies targets pairing without fiducials or heavy instrumentation, lever-

aging sensing and ML to identify/track ad hoc surfaces and objects. Examples include touch sensing on uninstrumented surfaces [Mollyn *et al*., 2025] and markerless tangible tracking for AR [Fan *et al*., 2025], as well as systems that recommend or utilize everyday objects as proxies [Jain *et al*., 2023]. This archetype signals a shift from lab-centric, instrumented pairing toward more practical and deployable pairing strategies.

Niche-archetype relations (what combinations dominate). The reviewed corpus suggests that training/rehearsal work tends to favor co-located tangible proxies often with domain-specific instrumentation such as manikins or tools, because fidelity and safety requirements motivate direct physical grounding [Schild *et al*., 2022; Uhl *et al*., 2023; Kang *et al*., 2024; Cao *et al*., 2025]. By contrast, psychophysical work tends to favor controlled discrepancy archetypes, because the research questions explicitly target detection thresholds, dominance effects, or transfer (for example, hand redirection, stiffness/texture illusions across 2021-2025). Social/play studies predominantly use simple tangible co-location (plush toys, balls, cards), where interaction naturalness is prioritized and the physical proxy is semantically straightforward [Bozgeyikli and Bozgeyikli, 2022; Huard *et al*., 2022; Fouad *et al*., 2025]. Finally, scalability-oriented systems (retargeting, reset optimization, proxy-importance) appear as a bridge between controlled lab setups and realistic deployment constraints (2023-2025), aiming to reduce prop counts, resets, and setup overhead while retaining pairing benefits.

Temporal trends (year-to-niche and year-to-archetype relations). From 2021 onward, pairing solutions frequently rely on instrumented tracking pipelines and carefully constrained environments, especially for illusions and collaborative interaction, in 2021 with hand-redirection and proxy-based illusions, and remote multi-user haptic props [Auda *et al*., 2021a]; early wearable/tangible actuators [Teng *et al*., 2021] In 2022-2023, the corpus shows diversification: richer application niches emerge, in 2022 with cultural heritage tangibles in 2022; and in 2023 with creative painting proxies, and pairing begins to be formalized as reusable methods and toolkits [Feick *et al*., 2023a]; adaptive reset methods (2023). By 2024-2025, there is stronger representation of applied, context-rich deployments: in-car VR [Yeo *et al*., 2025]; multisensory installations [Zhang *et al*., 2025]; broader telepresence/heritage framing [Pillis, 2025], alongside more practical sensing for markerless or ad hoc pairing [Mollyn *et al*., 2025; Fan *et al*., 2025]. In parallel, optimization-oriented retargeting work (2023-2025) reflects a maturing focus on efficiency, comfort, and continuous interaction rather than one-off demonstrations.

What is under-represented or absent. Despite the breadth above, several pairings remain comparatively scarce in the accepted studies. First, longitudinal and in-the-wild evaluations of pairing solutions are rare: most systems are assessed in lab studies or bounded deployments, even when the target use case implies extended real-world use, such as training curricula, workplace assessment. Second, cross-system comparability is limited: while multiple papers study similar archetypes, for instance, redirection, passive proxies, they often use different tasks, metrics, and calibration procedures, which complicates synthesis into unified design rules beyond conservative thresholds. Third, multi-user, remote, and networked pairing appears but is not dominant [Auda *et al*., 2021a; Weiss *et al*., 2025a], yet systematic exploration of synchronization strategies and failure modes is limited relative to single-user pairing. Fourth, although multisensory pairing beyond touch is present [Tao and Lopes, 2022; Yeo *et al*., 2025; Zhang *et al*., 2025], standardized multisensory authoring and validation pipelines are still uncommon, with most solutions remaining bespoke and domain-specific.

Implications for the tutorial framing. Overall, the accepted studies support the tutorial's framing that real-virtual pairing is not a monolithic technique but a family of solutions whose feasibility depends on the intended niche (training, social interaction, creative work, psychophysics), and the pairing archetype (direct co-location, retargeting/multiplexing, controlled discrepancy/illusions, or opportunistic markerless pairing). The corpus also suggests that guidance should explicitly help practitioners choose between fidelity-first pairing (strict co-location), scalability-first pairing (retargeting and opportunistic tracking), and perception-first pairing (illusion-based discrepancy), since each archetype leads to different design constraints, evaluation strategies, and expected failure modes.

5.2 Research Question 2 (Technological Advances)

The accepted studies collectively show that progress in real-virtual pairing is tightly coupled to tracking and sensing, the computational/interaction envelope of XR headsets or non-HMD installations, and authoring/integration software used to connect physical signals to coherent virtual effects. Rather than a single linear trajectory, the timeline suggests a co-evolution: improvements in consumer HMDs and engines reduce barriers for deployment, while pairing-centric research continues to introduce custom peripherals, external tracking, and bespoke middleware to reach accuracy, latency, and realism requirements.

Across 2021-2025, consumer VR HMDs remain the most common platform for pairing systems, including Oculus/Meta (Rift/Rift S, Quest 2, Quest 3, Quest Pro), HTC Vive variants (Vive, Vive Pro, Vive Pro Eye, Vive Pro 2), and HP Reverb-class devices (Reverb G2). These devices are repeatedly adopted in both psychophysics studies (e.g., hand redirection, stiffness and texture perception) and applied systems for training, social play or creative authoring, indicating that commodity VR hardware provides a sufficiently stable baseline for controlled experiments and rapid prototyping. In parallel, AR/MR HMDs appear frequently when pairing requires direct interaction with the physical world (for example, HoloLens 2 in tangible AR, markerless tangible tracking, mixed mock-up teleoperation, and medical/dental training). High-end mixed-reality headsets, such as Varjo XR series, are used when optical fidelity, pass-through depth handling, or instrumentation (LiDAR/chroma-key) is critical for training and simulation scenarios.

Notably, the corpus also includes non-HMD or hybrid display settings that broaden the notion of pairing: projection or world-fixed setups [Bozgeyikli, 2024] and projector/CAVE-like training comparisons [Grabowski, 2021] show that pairing solutions are not exclusive to head-worn displays. However, these configurations remain comparatively rare, and most methodological innovations target head-mounted pipelines.

The technological throughline across years is that pairing quality is tracking-limited: systems continually trade off between external, high-precision tracking and more deployable inside-out or markerless approaches.

Many systems still rely on OptiTrack-class tracking (or similar) to achieve stable co-registration for props, hands, or tools, especially in studies where subtle discrepancies matter (for instance, hand redirection and visuo-haptic illusion thresholds), or where tool-use fidelity is central in cases of medical training/manikin systems, collaborative haptic props. This reflects that the demands of precise real-virtual alignment remain higher than what pure consumer inside-out tracking can consistently guarantee across objects, occlusions, and constrained contact tasks.

Fiducial markers (Vive trackers mounted to objects, ArUco/Vuforia-like markers, or bespoke marker rigs) are frequently used to maintain robust object pose estimation for passive haptic props and instruments, such as extinguishers, manikin tools, proxy sliders/knobs. This is especially visible in training and rehearsal systems where reliability must dominate.

Recent work increasingly attempts to reduce instrumentation by using depth cameras, that can be Azure Kinect, RealSense, ZED mini, egocentric cameras, and lightweight ML pipelines to recognize touch and track everyday objects without fiducials [Mollyn *et al*., 2025; Fan *et al*., 2025; Jain *et al*., 2023]. These approaches expand deployability but typically introduce new forms of noise, latency, and failure modes that must be managed through interaction design and filtering.

Overall, the timeline indicates that while headsets improved, tracking for tangible pairing did not disappear as a research problem; instead, it diversified into two paths: high-fidelity lab tracking, which is still dominant in controlled evaluation, and pragmatic markerless sensing, which is growing but not yet universal.

Another clear advance is the increased sophistication of supporting hardware attached to, carried by, or surrounding the user. Early works often pair one prop with one illusion or interaction (for example, physical proxies for redirection, tracked objects for shared manipulation), whereas later systems integrate richer stacks: wearable finger actuators and foldable mechanisms [Teng *et al*., 2021], vibrotactile+pneumatic roughness modulation [Cai *et al*., 2025], EMS-based weight manipulation [Faltaous *et al*., 2022], quadcopter-delivered tangible UI elements [Auda *et al*., 2021b], and multi-actuator, multisensory environments (thermal/airflow/motion in vehicles [Yeo *et al*., 2025], fabric+haptics+audio installations [Zhang *et al*., 2025], real-world stimulus mapping [Tao and Lopes, 2022]). These systems suggest that pairing is increasingly treated as a sensing-actuation pipeline, not only a geometric registration problem.

At the same time, many applied systems purposely use passive hardware (cards, balls, brushes, manikins, mock-ups) because passive proxies often yield the best realism-per-cost ratio, particularly when paired with strong visual dominance effects and conservative alignment strategies, such as multiple passive-haptics evaluations and training systems. Thus, advances are not purely a move toward more actuation; rather, they reflect better matching of modality to task.

On the software side, the timeline shows consolidation around mainstream real-time engines: Unity is the dominant authoring platform across most systems (VR, AR, and MR), frequently paired with headset SDKs and plugins (SteamVR/OpenVR, MRTK, Vuforia, Varjo SDK, ZED/RealSense/Azure Kinect integrations). Unreal Engine appears less often but supports high-fidelity simulation contexts [Dufresne *et al*., 2024]. This dominance matters because it standardizes rendering, interaction, and deployment workflows; consequently, technological differentiation shifts to custom middleware that binds physical devices to the engine: OSC pipelines, microcontroller firmware (Arduino/ESP/Teensy/STM32), device drivers, and bespoke calibration tools.

Across years, we also observe increased use of toolkit-style contributions that attempt to make pairing reusable (for instance, block-based tangible proxy toolkits [Feick *et al*., 2023a], retargeting toolchains such as HaRT-based methods (2023-2025), and published plugins/add-ons [Auda *et al*., 2021b]). This suggests a shift from isolated system demos to reproducible building blocks, though interoperability across toolkits remains limited.

Although headset capabilities have improved with higher resolution, better inside-out tracking, richer pass-through, lighter form factors, the reviewed studies repeatedly report persistent rather than vanishing technical challenges-especially tracking stability, co-registration drift, and end-to-end latency. This persistence helps explain why external tracking and markers remain common in 2024-2025 despite more capable consumer headsets. In parallel, the emergence of markerless and ML-based approaches introduces a different challenge profile: even when deployment becomes easier, researchers must mitigate noisy pose estimates, occlusions, and classification errors through filtering, redundancy, or interaction constraints.

The technological advances in the corpus are best characterized as: a broadening of platform diversity (VR HMDs, AR/MR headsets, and occasional non-HMD installations), a bifurcation of tracking strategies (high-fidelity external tracking vs. deployable markerless sensing), increasing sophistication of peripheral multisensory hardware, and consolidation on Unity-centered software stacks with novel integration layers as the primary contribution site. Importantly, the timeline indicates that year-to-year progress is less about eliminating pairing difficulties and more about expanding what is feasible and deployable, while keeping alignment and latency as the core technical constraints shaping system design.

5.3 Research Question 3 (Open Challenges and Future Directions)

The reviewed literature suggests that real-virtual object pairing is transitioning from proofs of feasibility toward broader, application-driven deployments (training, teleoperation, well-being, cultural heritage, social XR). However, the same body of evidence indicates that the field remains constrained by a small set of recurring technical and methodological bottlenecks. Below, we synthesize the main open challenges and outline future directions grounded in the challenges reported across the accepted studies.

Robust co-registration beyond the lab. A dominant open problem is achieving stable, low-effort co-registration between physical props, hands, and virtual counterparts under real-world variability. Many works still depend on controlled lighting, constrained workspaces, markers, or external motion capture to ensure alignment, while more deployable approaches (e.g., ad hoc surfaces and markerless tangibles) introduce new noise and failure modes [Mollyn *et al.*, 2025; Fan *et al.*, 2025; Jain *et al.*, 2023]. The future direction is twofold: hybrid tracking pipelines that fuse inside-out headset sensing with lightweight object instrumentation when needed, that can be selective tags, intermittent re-localization, opportunistic depth sensing; and interaction designs that gracefully tolerate residual error, including perceptual compensation; such as conservative redirection bounds; adaptive alignment, and task-aware snapping strategies.

End-to-end latency as a perceptual and control constraint. Even when spatial errors are small, temporal misalignment between vision, touch, and proprioception remains a key limiter, especially for fine contact events, texture rendering, and fast tool interactions [Normand *et al.*, 2024; Matthews *et al.*, 2023b]. Future work should treat latency not merely as an engineering metric but as a perception-aware budget: systems can measure and allocate delay across sensing, networking, actuation, and rendering, and then explicitly adapt interaction parameters (gain schedules, predictive filtering, contact hysteresis) to keep delays below perceptual thresholds for the intended task.

Safety and comfort in embodied, high-agency scenarios. As pairing systems move into higher-energy interactions (for example, locomotion, exertion controllers, in-vehicle VR, large props, robots, or dynamically redirected encountered-type devices), safety becomes a first-class challenge [Auda *et al.*, 2021b; Zhu *et al.*, 2023; Zhou and Popescu, 2025; Yeo *et al.*, 2025; Senk *et al.*, 2025]. A clear future direction is formal safety envelopes for pairing: runtime monitoring of user posture, reachable zones, collision risk, and device state, coupled with fail-safe modes (soft resets, safe retraction, speed limiting, emergency stop patterns). This is particularly critical for robot or vehicle-mediated haptics and for multi-user co-located setups.

Generalizability across users, tasks, and environments. Across studies, limitations repeatedly include restricted contexts (single task, limited object set, controlled room, trained participants), raising concerns about transferability to broader use [Dufresne *et al.*, 2024; Cherix *et al.*, 2025; Fouad *et al.*, 2025]. Future research should prioritize: cross-task benchmarks that test pairing under varied interaction primitives (grasp, press, slide, rotate, walk, handover), multi-environment evaluations (lighting, clutter, surface materials, occlusions), user-diverse studies that capture differences in proprioception, VR experience, and sensory sensitivity (in cases of sensitivity-focused protocols [Lebrun *et al.*, 2024]). These steps are needed to move from works in this setup to predictable performance envelopes.

Scaling pairing to many objects and continuous interaction. A recurring systems challenge is scaling from one well-tracked proxy to scenes containing many interactive objects. Retargeting and proxy-selection methods address this via resets and prop reuse [Matthews *et al.*, 2023b; Sun *et al.*, 2025; Liu *et al.*, 2025], and synthesis approaches attempt to reduce mismatch by reshaping virtual content [Fukumoto *et al.*, 2022]. Future directions include: scene-level pairing planners that allocate limited physical resources across objects, incorporate ergonomic costs, and adapt online to user intent; and content-aware authoring tools that recommend which virtual objects should be paired, retargeted, or synthesized based on expected interaction frequency and required fidelity.

Standardizing evaluation: what should be measured, and how? Because pairing spans perception, interaction, and engineering, studies use heterogeneous measures (presence, realism, performance, workload, sickness, thresholds), making comparisons difficult. Recent work that explicitly measures co-registration error and delay [Mubarrat *et al.*, 2024] points toward a needed direction: shared evaluation protocols that combine objective spatial/temporal metrics, task performance, and perceptual outcomes (comfort, plausibility, detection). A practical next step is to define minimum reporting standards for pairing systems (tracking method, calibration procedure, measured drift, measured latency, failure cases) so results become comparable and reproducible.

Broadening pairing beyond visuo-haptics: multisensory and social coupling. While many works focus on visuo-haptic alignment and illusion thresholds, there is an emerging trend toward multisensory pairing (thermal/airflow/motion, sound-driven narratives, real-world distraction mapping, well-being installations) and social XR (collaborative ownership, shared haptic props, virtual pets) [Tao and Lopes, 2022; Yeo *et al.*, 2025; Harley *et al.*, 2021; Auda *et al.*, 2021a; Huard *et al.*, 2022; Fouad *et al.*, 2025]. Future work should explore cross-modal consistency as a first-class design variable: how thermal, airflow, sound, and motion cues can compensate for haptic mismatch or exacerbate it, and how multi-user synchronization affects perceived co-presence when physical props cannot be identically realized for all participants.

Practical adoption: deployability, cost, and maintenance. Finally, many pairing systems remain research prototypes with non-trivial setup overhead (calibration rigs, external tracking, custom actuation). For broader adoption in training, healthcare, and industry, future directions include deployable hardware reference designs (modular mounts, repeatable calibration procedures), software packaging (toolkits with documented pipelines), and maintenance-aware design (battery life, robustness, sanitation for wearables/manikins, and field recalibration).

Overall, the evidence suggests that the field's next gains will come less from isolated improvements in realism and more from robustness, scalability, and standardization: pairing methods that tolerate real-world variability, quantify their spatial-temporal envelopes, and support diverse tasks and users. These directions also naturally expand the role of real-virtual pairing from visuo-haptic illusion studies toward reliable, multisensory XR systems for consequential application, such as training, teleoperation, healthcare, and social XR.

6 Conclusion

This systematic review examined how XR systems pair real and virtual objects to enable visuo-haptic consistency, passive/encountered haptics, tangible interaction, and illusion-based remapping. Across the analyzed studies, we identified a set of recurring real-virtual pairing solution archetypes, including one-to-one tracked proxies, prop reuse via retargeting and resets, dynamic redirection, markerless/everyday-object tangibles, and synthesis or perceptual manipulation approaches that reduce (or exploit) mismatch rather than eliminating it. Taken together, the evidence indicates that object pairing has matured from isolated demonstrations into a versatile design strategy supporting increasingly diverse applications, such as skills training in medical, industrial, and safety scenarios, collaborative and social XR, cultural heritage interaction, and multisensory wellbeing experiences.

From a technological perspective, the timeline suggests a clear expansion in both hardware and software sophistication: modern standalone headsets increasingly serve as the primary platform, while studies frequently integrate supporting sensing (depth cameras, wearables, mobile devices, or motion capture) when higher fidelity is required. Author-developed Unity-based applications dominate the ecosystem, reflecting both rapid prototyping advantages and the current lack of standardized, reusable pipelines for calibration, tracking, and co-registration. Despite these advances, the review shows that technical progress has not removed the core constraints of real-virtual pairing; instead, it has shifted them toward deployment-relevant concerns, such as robustness under environmental variability, end-to-end latency, ergonomic burden, and safety in high-agency interactions.

The experiments surveyed further reveal a methodological pattern: most evaluations rely on small-to-medium participant samples and controlled laboratory conditions, often focusing on presence, realism, performance, workload, and cybersickness. While such studies provide strong initial validation for specific systems and tasks, the broader picture highlights persistent gaps in generalizability and comparability. In particular, heterogeneous outcome measures and incomplete reporting of spatial/temporal error budgets limit cross-paper synthesis and hinder reproducibility. These issues reinforce the need for shared evaluation practices that combine objective co-registration metrics with task performance and perceptual outcomes, alongside minimum reporting standards for calibration procedures, tracking accuracy, latency, and failure cases.

Finally, the open challenges consolidated in this review point to the same conclusion: the next stage of progress is likely to be driven less by incremental realism improvements and more by scalable, deployable pairing. Key future directions include robust co-registration pipelines that fuse inside-out sensing with lightweight object instrumentation, perception-aware latency management and adaptive interaction strategies, formal safety envelopes for dynamic and robot/vehicle-mediated haptics, benchmarking and user-diverse evaluations to improve generalizability, and authoring tools and scene-level planners that scale pairing to many objects and continuous interaction. Addressing these directions can consolidate real-virtual object pairing as a dependable foundation for XR systems that are not only compelling in the lab, but also practical in real-world deployments.

Declarations

Authors' Contributions

JMT and FN contributed to the conception of this study and reviewed the manuscript. DO and MK performed the main part of the SLR. DO is the main contributor and writer of this manuscript, while MK and CG helped with the writing and reviewing as well.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets generated and/or analysed during the current study will be made available upon request.

References

Auda, J., Busse, L., Pfeuffer, K., Gruenefeld, U., Rivu, R., Alt, F., and Schneegass, S. (2021a). I'm in control! transferring object ownership between remote users with haptic props in virtual reality. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction*, SUI '21, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3485279.3485287.

Auda, J., Verheyen, N., Mayer, S., and Schneegass, S. (2021b). Flyables: Haptic input devices for virtual reality using quadcopters. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, VRST '21, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3489849.3489855.

Azmandian, M., Hancock, M., Benko, H., Ofek, E., and Wilson, A. D. (2016). Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences. In *Proceedings of the 2016 chi conference on human factors in computing systems*, pages 1968–1979. DOI: 10.1145/2858036.2858226.

Beever, L. and John, N. W. (2022). Leveled sr: A substitutional reality level design workflow. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 130–138. DOI: 10.1109/VR51125.2022.00031.

Botvinick, M. and Cohen, J. (1998). Rubber hands 'feel'touch that eyes see. *Nature*, 391(6669):756–756. DOI: 10.1038/35784.

Bozgeyikli, L. (2024). Real-virtual objects: Exploring bidirectional embodied tangible interaction with a virtual human in world-fixed virtual reality. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pages 147–156. DOI: 10.1109/VR58804.2024.00038.

Bozgeyikli, L. L. and Bozgeyikli, E. (2022). Tangiball: Foot-enabled embodied tangible interaction with a ball in virtual reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 812–820. DOI: 10.1109/VR51125.2022.00103.

Cai, S., Chen, Z., Gao, H., Huang, Y., Zhang, Q., Yu, X., and Zhu, K. (2025). Vibopneumo: A vibratory-pneumatic finger-worn haptic device for altering perceived texture roughness in mixed reality. *IEEE Transactions on Visualization and Computer Graphics*, 31(7):3957–3972. DOI: 10.1109/TVCG.2024.3391877.

Calandra, D., De Lorenzis, F., Cannavò, A., and Lamberti, F. (2023). Immersive virtual reality and passive haptic interfaces to improve procedural learning in a formal training course for first responders. *Virtual Reality*, 27(2):985–1012. DOI: 10.1007/s10055-022-00704-9.

Cao, W., Li, Z., He, Y., Liu, Z., Zhang, M., Lv, J., and Si, W. (2025). Cost-effective tangible rehearsal interface for microsurgical clipping of intracranial aneurysm. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pages 401–411. DOI: 10.1109/VR59515.2025.00064.

Chen, P.-Y., Tsai, C.-Y., Wang, W.-H., Lai, C.-J., Fan, C.-A., Lin, S. C., Chi, C.-C., and Chen, M. Y. (2023). Aircharge: Amplifying ungrounded impact force by accumulating air propulsion momentum. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3586183.3606768.

Cherix, R., Mugellini, E., and Lalanne, D. (2025). Impact of passive haptics on task performance: of the effect of technological evolution. In *Proceedings of the 2025 31st ACM Symposium on Virtual Reality Software and Technology*, VRST '25, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3756884.3765999.

Clavelin, G., Bouhier, M., Tseng, W.-J., and Gugenheimer, J. (2023). Exploring the perception of pain in virtual reality through perceptual manipulations. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3544549.3585674.

Cosco, F., Garre, C., Bruno, F., Muzzupappa, M., and Otaduy, M. A. (2012). Visuo-haptic mixed reality with unobstructed tool-hand integration. *IEEE transactions on visualization and computer graphics*, 19(1):159–172. DOI: 10.1109/TVCG.2012.107.

Degraen, D., Feick, M., Durdyyev, S., and Krüger, A. (2024). Prototyping surface slipperiness using sole-attached textures during haptic walking in virtual reality. In *Proceedings of the 23rd International Conference on Mobile and Ubiquitous Multimedia*, MUM '24, page 95–105, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3701571.3701589.

Dufresne, F., Nilsson, T., Gorisse, G., Guerra, E., Zenner, A., Christmann, O., Bensch, L., Callus, N. A., and Cowley, A. (2024). Touching the moon: Leveraging passive haptics, embodiment and presence for operational assessments in virtual reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3613904.3642292.

Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433. DOI: 10.1038/415429a.

Faltaous, S., Prochazka, M., Auda, J., Keppel, J., Wittig, N., Gruenefeld, U., and Schneegass, S. (2022). Give weight to vr: Manipulating users' perception of weight in virtual reality with electric muscle stimulation. In *Proceedings of Mensch Und Computer 2022*, MuC '22, page 533–538, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3543758.3547571.

Fan, X., Huang, X., and Xiao, R. (2025). Tangiar: Markerless tangible input for immersive augmented reality with everyday objects. In *Proceedings of the 2025 31st ACM Symposium on Virtual Reality Software and Technology*, VRST '25, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3756884.3766028.

Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers*, 35(3):379–383. DOI: 10.3758/BF03195514.

Feick, M., Biyikli, C., Gani, K., Wittig, A., Tang, A., and Krüger, A. (2023a). Voxelhap: A toolkit for constructing proxies providing tactile and kinesthetic haptic feedback in virtual reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3586183.3606722.

Feick, M., Kleer, N., Zenner, A., Tang, A., and Krüger, A. (2021). Visuo-haptic illusions for linear translation and stretching using physical proxies in virtual reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3411764.3445456.

Feick, M., Regitz, K. P., Tang, A., and Krüger, A. (2022). Designing visuo-haptic illusions with proxies in virtual reality: Exploration of grasp, movement trajectory and object mass. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3491102.3517671.

Feick, M., Zenner, A., Ariza, O., Tang, A., Biyikli, C., and Krüger, A. (2023b). Turn-it-up: Rendering resistance for knobs in virtual reality through undetectable pseudo-haptics. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3586183.3606787.

Fender, A. R., Roberts, T., Luong, T., and Holz, C. (2023). Infinitepaint: Painting in virtual reality with passive haptics using wet brushes and a physical proxy canvas. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3544548.3580927.

Fouad, A., Nam, H., Nguyen, A., Han, D., Kim, D., Cho, I., and Kim, K. (2025). Touching the virtual dog: Effects of active and passive haptic feedback on social presence and emotional bonding in virtual pet interaction. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 943–953. DOI: 10.1109/ISMAR67309.2025.00102.

Fukumoto, N., Isoyama, N., Uchiyama, H., Sakata, N., and Kiyokawa, K. (2022). An object synthesis method to enhance visuo-haptic consistency. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 422–430. DOI: 10.1109/ISMAR55827.2022.00058.

Geslain, B., Bailly, G., Haliyo, S. D., and Duboc, C. (2021). Visuo-haptic illusions for motor skill acquisition in virtual reality. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction*, SUI '21, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3485279.3485291.

Geslain, B., Besga, S., Lebrun, F., and Bailly, G. (2022). Generalizing the hand redirection function in virtual reality. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*, AVI '22, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3531073.3531100.

Gouton, M.-A., Dacremont, C., Trystram, G., and Blumenthal, D. (2023). Effect of perceptive enrichment on the efficiency of simulated contexts: Comparing virtual reality and immersive room settings. *Food Research International*, 165:112492. DOI: 10.1016/j.foodres.2023.112492.

Grabowski, A. (2021). Practical skills training in enclosure fires: An experimental study with cadets and firefighters using cave and hmd-based virtual training simulators. *Fire Safety Journal*, 125:103440. DOI: 10.1016/j.firesaf.2021.103440.

Grießel, R., Weber, B., and Weyers, B. (2025). Passive haptics role in vr-based ergonomic workplace assessment. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 1–11. DOI: 10.1109/ISMAR67309.2025.00050.

Günther, S., Rasch, J., Schön, D., Müller, F., Schmitz, M., Riemann, J., Matviienko, A., and Mühlhäuser, M. (2022). Smooth as steel wool: Effects of visual stimuli on the haptic perception of roughness in virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3491102.3517454.

Harley, D., P. Tarun, A., J. Stinson, B., Tibu, T., and Mazalek, A. (2021). Playing by ear: Designing for the physical in a sound-based virtual reality narrative. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '21, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3430524.3440635.

He, W., Li, X., Xu, S., Chen, Y., Sio, C.-I., Kan, G. L., and Lee, L.-H. (2024). Metadragonboat: Exploring paddling techniques of virtual dragon boating in a metaverse campus. In *Proceedings of the 32nd ACM International Conference on Multimedia*, MM '24, page 6335–6344, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3664647.3681078.

Huard, A., Chen, M., and Sra, M. (2022). Cardsvr: A two-person vr experience with passive haptic feedback from a deck of playing cards. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 538–547. DOI: 10.1109/ISMAR55827.2022.00070.

Iwamoto, K. (2022). Mixed environment of real and virtual objects for task training using binocular video see-through display and haptic device. In *2022 IEEE 20th International Conference on Industrial Informatics (INDIN)*, pages 374–379. DOI: 10.1109/INDIN51773.2022.9976123.

Jain, R., Shi, J., Duan, R., Zhu, Z., Qian, X., and Ramani, K. (2023). Ubi-touch: Ubiquitous tangible object utilization through consistent hand-object interaction in augmented reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3586183.3606793.

Johnson-Glenberg, M. C., Yu, C. S., Liu, F., Amador, C., Bao, Y., Yu, S., and LiKamWa, R. (2023). Embodied mixed reality with passive haptics in stem education: Randomized control study with chemistry titration. *Frontiers in Virtual Reality*, 4:1047833. DOI: 10.3389/frvir.2023.1047833.

Kahl, D., Ruble, M., and Krüger, A. (2022). The influence of environmental lighting on size variations in optical see-through tangible augmented reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 121–129. DOI: 10.1109/VR51125.2022.00030.

Kang, H., Kang, T., and Wallraven, C. (2023). Putting vision and touch into conflict: Results from a multimodal mixed reality setup. *IEEE Transactions on Visualization and Computer Graphics*, 29(12):5224–5234. DOI: 10.1109/TVCG.2022.3207241.

Kang, H., Yang, J., Ko, B.-S., Kim, B.-S., Song, O.-Y., and Choi, S.-M. (2024). Integrated augmented and virtual reality technologies for realistic fire drill training. *IEEE Computer Graphics and Applications*, 44(2):89–99. DOI: 10.1109/MCG.2023.3303028.

Kawazoe, A., Reardon, G., Woo, E., Luca, M. D., and Visell, Y. (2021). Tactile echoes: Multisensory augmented reality for the hand. *IEEE Transactions on Haptics*, 14(4):835–848. DOI: 10.1109/TOH.2021.3084117.

Kim, W. and Xiong, S. (2022). Pseudo-haptic button for improving user experience of mid-air interaction in vr. *Inter-*

national Journal of Human-Computer Studies, 168:102907. DOI: 10.1016/j.ijhcs.2022.102907.

Kurzweg, M., Weiss, Y., Ernst, M. O., Schmidt, A., and Wolf, K. (2024). Survey on haptic feedback through sensory illusions in interactive systems. *ACM Computing Surveys*, 56(8):1–39. DOI: 10.1145/3648353.

Lebrun, F., Haliyo, S. D., Auvray, M., Gueorguiev, D., and Bailly, G. (2024). Study of the impact of visual and haptic sensory sensitivities on the detection of visuo-haptic illusions in virtual reality. In *Proceedings of the 35th Conference on l'Interaction Humain-Machine*, IHM '24, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3649792.3649803.

Lécuyer, A. (2009). Simulating haptic feedback using vision: A survey of research and applications of pseudo-haptic feedback. *Presence: Teleoperators and Virtual Environments*, 18(1):39–53. DOI: 10.1162/pres.18.1.39.

Li, Y., Ye, H., Wu, S., Zhao, X., Liu, Y., Lv, L., Zhang, P., Zhang, X., and Zhou, Y. (2022). Mixed reality and haptic–based dental simulator for tooth preparation: Research, development, and preliminary evaluation. *JMIR Serious Games*, 10(1). DOI: 10.2196/30653.

Liu, Z., Wu, J., Wang, L., Li, X., and Im, S. K. (2025). Proxy importance based haptic retargeting with multiple props in vr. *IEEE Transactions on Visualization and Computer Graphics*, 31(7):3987–4002. DOI: 10.1109/TVCG.2024.3392743.

Matthews, B. J., Reichherzer, C., Thomas, B. H., and Smith, R. T. (2023a). Maskwarp: Visuo-haptic illusions in mixed reality using real-time video inpainting. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 767–768. IEEE. DOI: 10.1109/ISMAR-Adjunct60411.2023.00167.

Matthews, B. J., Thomas, B. H., Von Itzstein, G. S., and Smith, R. T. (2023b). Adaptive reset techniques for haptic retargeted interaction. *IEEE Transactions on Visualization and Computer Graphics*, 29(2):1478–1490. DOI: 10.1109/TVCG.2021.3120410.

Milgram, P. and Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329. Available at:`https://www.researchgate.net/publication/231514051_A_Taxonomy_of_Mixed_Reality_Visual_Displays`.

Mollyn, V., DeVrio, N., and Harrison, C. (2025). Eclipsetouch: Touch segmentation on ad hoc surfaces using worn infrared shadow casting. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*, UIST '25, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3746059.3747743.

Mubarrat, S. T., Chowdhury, S. K., and Fernandes, A. S. (2024). Evaluating visual-spatiotemporal co-registration of a physics-based virtual reality haptic interface. *IEEE Access*, 12:57017–57032. DOI: 10.1109/ACCESS.2024.3391186.

Murmu, C., Venkatakrishnan, R., Venkatakrishnan, R., Lin, W.-C., Robb, A. C., Pagano, C., and Babu, S. V. (2025). An empirical evaluation of how virtual hand visibility affects near-field size perception and reporting of tangible objects in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 31(11):9867–9877. DOI: 10.1109/TVCG.2025.3616829.

Nam, S., Kim, M., Kim, B., Kim, D., Seo, M.-H., and Lee, M. (2025). Exploring how prior knowledge and presence shape transfer of a reversed size-weight illusion from virtual to real. In *Proceedings of the 2025 31st ACM Symposium on Virtual Reality Software and Technology*, VRST '25, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3756884.3765985.

Normand, E., Pacchierotti, C., Marchand, E., and Marchal, M. (2024). How different is the perception of vibrotactile texture roughness in augmented versus virtual reality? In *Proceedings of the 30th ACM Symposium on Virtual Reality Software and Technology*, VRST '24, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3641825.3687738.

Ohara, D., Oliveira, R., Nunes, F., and Teixeira, J. M. (2025a). Visuo-haptic immersion made easy: A fast-track course on spatial anchors. In *2025 27th Symposium on Virtual and Augmented Reality (SVR)*, pages 1–6. DOI: 10.1109/SVR67689.2025.00012.

Ohara, D., Pierre, G., Nunes, F., and Teixeira, J. M. (2025b). Tactile perception in matched virtual reality: Evaluating virtual textures on a real surface. In *2025 27th Symposium on Virtual and Augmented Reality (SVR)*, pages 165–174. DOI: 10.1109/SVR67689.2025.00033.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., *et al*. (2021). The prisma 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, 372. DOI: 10.1136/bmj.n71.

Pillis, D. (2025). Towards temporal telepresence multisensory spatial intelligence: Reconstructing the human experience through teleabsence. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 1134–1141. DOI: 10.1109/VRW66409.2025.00225.

Pooladvand, S., Hasanzadeh, S., Takahashi, G., Park, K. J., and Marroquin, J. (2025). Simulating social pressure: Evaluating risk behaviors in construction using augmented virtuality. *IEEE Transactions on Visualization and Computer Graphics*, 31(5):2494–2503. DOI: 10.1109/TVCG.2025.3549877.

Sagheb, S., Liu, F. W., Vuong, A., Dai, S., Wirjadi, R., Bao, Y., and Likamwa, R. (2023). Geppetteau: Enabling haptic perceptions of virtual fluids in various vessel profiles using a string-driven haptic interface. In *Proceedings of the Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '23, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3569009.3572745.

Schild, J., Carbonell, G., Tarrach, A., and Ebeling, M. (2022). Vitawin - interprofessional medical mixed reality training for paramedics and emergency nurses. In *2022 IEEE 10th International Conference on Serious Games and Applications for Health(SeGAH)*, pages 1–8. DOI: 10.1109/SEGAH54908.2022.9978566.

Senk, V., Ghazanfari, M., Rasoulzadeh, S., Vasylevska, K., Kovacs, B. I., Mortezapoor, S., Vonach, E., Kovacic, I., Füssl, J., Kaufmann, H., and Königsberger, M. (2025). Hexa: Haptic-enhanced extended reality framework for material-informed architectural design. *Journal of Building Engineering*, 108:112707. DOI: 10.1016/j.jobe.2025.112707.

Song, T., Pabst, F., Eck, U., and Navab, N. (2025). Enhancing patient acceptance of robotic ultrasound through conversational virtual agent and immersive visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 31(5):2901–2911. DOI: 10.1109/TVCG.2025.3549181.

Stellmacher, C., Mathis, F., Weiss, Y., Loerakker, M. B., Wagener, N., and Schöning, J. (2024). Exploring mobile devices as haptic interfaces for mixed reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3613904.3642176.

Sun, A., Wu, J., Fan, R., Im, S. K., and Wang, L. (2025). Efficient and comfortable haptic retargeting with reset point optimization. *IEEE Transactions on Visualization and Computer Graphics*, 31(9):6588–6602. DOI: 10.1109/TVCG.2024.3523042.

Tao, Y. and Lopes, P. (2022). Integrating real-world distractions into virtual reality. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3526113.3545682.

Teng, S.-Y., Li, P., Nith, R., Fonseca, J., and Lopes, P. (2021). Touchfold: A foldable haptic actuator for rendering touch in mixed reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3411764.3445099.

Uhl, J. C., Schrom-Feiertag, H., Regal, G., Gallhuber, K., and Tscheligi, M. (2023). Tangible immersive trauma simulation: Is mixed reality the next level of medical skills training? In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3544548.3581292.

Vidal-Balea, A., Blanco-Novoa, O., Fraga-Lamas, P., and Fernández-Caramés, T. M. (2025). Off-cloud anchor sharing framework for multi-user and multi-platform mixed reality applications. *Applied Sciences*, 15(13):6959. DOI: 10.3390/app15136959.

Weiss, Y., Rasch, J., Fischer, J., and Müller, F. (2025a). Investigating the effects of haptic illusions in collaborative virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 31(11):10109–10119. DOI: 10.1109/TVCG.2025.3616760.

Weiss, Y., Villa, S., Schmidt, A., Mayer, S., and Müller, F. (2023). Using pseudo-stiffness to enrich the haptic experience in virtual reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3544548.3581223.

Weiss, Y., Villa, S., Ziarko, M., and Müller, F. (2025b). Manipulating stiffness perception of compliant objects while pinching in virtual reality. In *Proceedings of the 2025 31st ACM Symposium on Virtual Reality Software and Technology*, VRST '25, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3756884.3765988.

Xu, F., You, H., Zhou, T., and Du, J. (2025). Augmented reality robot manipulation with physical mockup and improved spatial cognition. *International Journal of Human-Computer Studies*, 205:103658. DOI: 10.1016/j.ijhcs.2025.103658.

Yeo, D., Kim, G., Oh, M., Park, J., Gim, B., Kang, S., Elsharkawy, A., and Kim, S. (2025). Attracar: Multisensory in-car vr with thermal, airflow, and motion feedback through built-in vehicle systems. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*, UIST '25, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3746059.3747642.

Yigitbas, E., Krois, S., Renzelmann, T., and Engels, G. (2022). Comparative evaluation of ar-based, vr-based, and traditional basic life support training. In *2022 IEEE 10th International Conference on Serious Games and Applications for Health(SeGAH)*, pages 1–8. DOI: 10.1109/SEGAH54908.2022.9978596.

Zenner, A., Ullmann, K., and Krüger, A. (2021). Combining dynamic passive haptics and haptic retargeting for enhanced haptic feedback in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2627–2637. DOI: 10.1109/TVCG.2021.3067777.

Zhang, T., Lin, W., Zhang, D., Pan, X., Latham, W., Grayson, K., Watson, Z., and Gillies, M. F. P. (2025). Sensing nature. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*, DIS '25, page 1–14, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3715336.3735424.

Zheng, W., Xiong, D., Li, J., Jiang, J., Weng, C., Zhou, J., and Fan, M. (2024). Investigating size congruency between the visual perception of a vr object and the haptic perception of its physical world agent. In *Proceedings of the 17th International Symposium on Visual Information Communication and Interaction*, VINCI '24, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3678698.3678706.

Zhou, Y. and Popescu, V. (2025). Dynamic redirection for safe interaction with ethd-simulated virtual objects. *IEEE Transactions on Visualization and Computer Graphics*, 31(5):2923–2932. DOI: 10.1109/TVCG.2025.3549169.

Zhu, H. Y., Chen, H.-T., and Lin, C.-T. (2023). The effects of virtual and physical elevation on physiological stress during virtual reality height exposure. *IEEE Transactions on Visualization and Computer Graphics*, 29(4):1937–1950. DOI: 10.1109/TVCG.2021.3134412.

Zhu, L., Jiang, X., Shen, J., Zhang, H., Mo, Y., and Song, A. (2022). Tapetouch: A handheld shape-changing device for haptic display of soft objects. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3928–3938. DOI: 10.1109/TVCG.2022.3203087.

Table 3. Highlighted papers selected from the final corpus based on the largest participant samples.

Authors	Participants	Summary	Main contributions
Gouton *et al*. [2023]	279	Compares performance and related outcomes between HMD-based VR and an immersive room/CAVE-like setup, examining whether perceptive enrichment improves efficiency in simulated contexts; includes a large participant sample.	Large-sample empirical evidence contrasting HMD VR and immersive-room setups for efficiency/experience trade-offs.
Pillis [2025]	115	Presents TeleAbsence as a direction for reconstructing richer human experiences via telepresence with multisensory/spatial intelligence, framing system concepts and reporting evaluation signals with a large sample.	Conceptual + technical framing of TeleAbsence with empirical signals from a sizable participant group, informing future multisensory telepresence research.
Kang *et al*. [2023]	90	Studies what happens when visual and tactile cues conflict in a multimodal MR setup, using controlled conditions to probe dominance, integration, and perceptual consequences under multisensory mismatch.	Empirical results + methodology for studying cross-modal conflicts in MR; actionable implications for XR setups that intentionally or inadvertently create cue conflicts.
Murmu *et al*. [2025]	80	Investigates how seeing (or not seeing) one's virtual hands impacts near-field size perception and the way users report properties of tangible objects while in VR, highlighting implications for perception and interaction design.	Evidence-based guidance for VR interaction/UI design regarding embodiment cues (hands) and perceptual reporting of physical/tangible targets.
Grabowski [2021]	74	Proposes and compares two virtual training systems for enclosure-fire nozzle person training: a CAVE and an HMD VR simulator. Includes end-user acceptance/usability, presence/realism, sickness, and stress considerations with 67 cadets + 7 firefighters.	Practical evidence and design discussion for multisensory/passive-haptic firefighter training simulators; comparative insights about CAVE vs HMD VR.
Pooladvand *et al*. [2025]	64	Examines whether simulated social pressure in an augmented-virtuality construction context can influence risk behaviors, using participant-based evaluation to assess behavioral changes and experience outcomes.	Evidence that XR can be used to probe and potentially train against risk behaviors driven by social pressure, with practical implications for safety training design.
Schild *et al*. [2022]	44	Describes an MR training setup for interprofessional education where paramedic students and emergency nursing trainees jointly train a shared burn-patient scenario, integrating a haptic mannequin and evaluating usability/presence/sickness and practical constraints; includes 30 paramedic students + 14 nursing trainees.	MR training approach integrating social XR + haptics mannequin plus evaluation findings and limitations (complexity, comfort, stability, emotion fidelity).
Bozgeyikli [2024]	40	Explores real-virtual objects enabling bidirectional tangible interaction between a user and a virtual human in world-fixed VR, focusing on embodied interaction and shared object manipulation; includes evaluation with 40 participants.	Interaction concept + implementation framing for real-virtual objects with virtual humans; evaluation-driven insights into bidirectional tangible interaction.
Bozgeyikli and Bozgeyikli [2022]	40	Proposes a foot-enabled tangible ball interaction for VR by pairing a tracked physical ball with a virtual ball; evaluates benefits for performance/experience in a room-scale soccer-like task.	System design for tangible foot interaction + empirical evidence on when physical-ball pairing improves outcomes; practical guidance for sports/rehab/game-like VR.
Li *et al*. [2022]	34	Presents the design of an MR+haptics dental simulator for tooth preparation and reports a preliminary face validity evaluation with 34 participants (novice vs skilled). Reports satisfaction/acceptance outcomes and group comparisons.	End-to-end MR+haptic simulator description + preliminary evidence of usability/acceptance and measures linked to overall experience.