GestureExplorer: Immersive Visualisation and Exploration of Gesture Data

Ang Li
Monash University
Melbourne, Victoria, Australia
University of Queensland
Brisbane, Queensland, Australia
angli21@sigchi.org

Jiazhou Liu
Monash University
Melbourne, Victoria, Australia
jiazhou.liu@monash.edu

Maxime Cordeil
University of Queensland
Brisbane, Queensland, Australia
m.cordeil@uq.edu.au

Jack Topliss
University of Canterbury
Christchurch, New Zealand
jack.topliss@pg.canterbury.ac.nz

Thammathip Piumsomboon
University of Canterbury
Christchurch, New Zealand
tham.piumsomboon@canterbury.ac.nz

Barrett Ens
Monash University
Melbourne, Victoria, Australia
barrett.ens@monash.edu

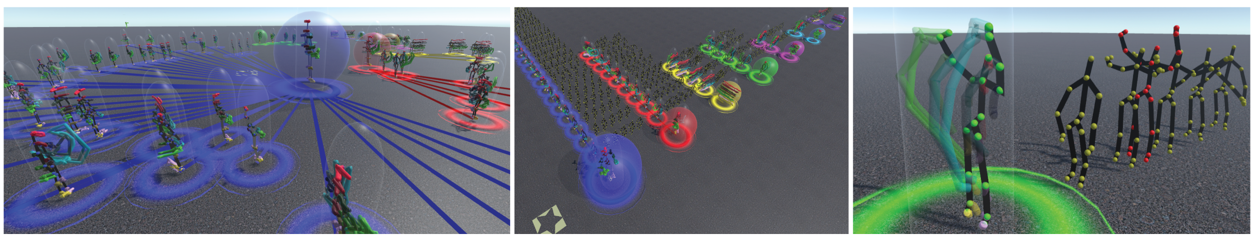


Figure 1: GestureExplorer supports immersive exploration of gesture data. Gestures are clustered by similarity and can be spatially arranged by similarity distance to each cluster (left), or in sorted order (middle). We provide several interactive features for exploring individual gestures such as trajectory visualisation, small multiples, and animation (right).

ABSTRACT

This paper presents the design and evaluation of GestureExplorer, an Immersive Analytics tool that supports the interactive exploration, classification and sensemaking with large sets of 3D temporal gesture data. GestureExplorer features 3D skeletal and trajectory visualisations of gestures combined with abstract visualisations of clustered sets of gestures. By leveraging the large immersive space afforded by a Virtual Reality interface our tool allows free navigation and control of viewing perspective for users to gain a better understanding of gestures.

We explored a selection of classification methods to provide an overview of the dataset that was linked to a detailed view of the data that showed different visualisation modalities. We evaluated GestureExplorer with two user studies and collected feedback from participants with diverse visualisation and analytics backgrounds. Our results demonstrated the promising capability of GestureExplorer for providing a useful and engaging experience in exploring and analysing gesture data.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI '23, April 23–28, 2023, Hamburg, Germany
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3580678

CCS CONCEPTS

• **Human-centered computing → Visualization techniques**; **Visual analytics**; **Virtual reality**.

KEYWORDS

gesture elicitation study, virtual reality, immersive analytics

ACM Reference Format:
Ang Li, Jiazhou Liu, Maxime Cordeil, Jack Topliss, Thammathip Piumsomboon, and Barrett Ens. 2023. GestureExplorer: Immersive Visualisation and Exploration of Gesture Data. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3544548.3580678

1 INTRODUCTION

Gesture elicitation studies (GES) are a popular way to elicit common gesture patterns that embody user preferences [52]. To analyse and group similar gestures, analysts have traditionally relied on video-recordings of gestures made by participants in response to a given set of referents. This method requires time-consuming manual analysis, which restricts their use to relatively small data sets [50]. However, boosted by motion capture technologies like Kinect[1], gestures can now be tracked and recorded as time series of 3D coordinates, making it possible for the data to be processed and analysed by machine.

[1]Kinect: https://azure.microsoft.com/en-us/services/kinect-dk/

Researchers have recently introduced tools that support the process of pattern analysis in GES through automated clustering algorithms and visualisation [12, 22, 24, 25, 49]. However, it is challenging for algorithms like *k*-means to make meaningful groups of high dimensional data like gesture data without human input [54]. Thus human-in-the-loop analysis is needed to refine and validate the clusters. Meanwhile, these existing GES tools based on desktops either use 2D projections or 3D views with fixed or limited change of view angles, which obscure some information from the original 3D gesture data. Another drawback is the limited available screen space for arranging a large set of gesture data.

The emergence of Immersive Analytics (IA) provides new opportunities for gesture data analysis. IA exploits emerging immersive technologies such as augmented reality (AR) and virtual reality (VR), to assist analysts to understand data in an engaging way [13]. Immersive environments preserve the 3D nature of data to reduce the cognitive effort of mentally manipulating 2D views [32]. The available large virtual space allows gestures to be arranged in 3D space for organisation and comparison, and can potentially leverage spatial memory to assist navigation between them [20, 23]. While research in IA has recently proliferated [18], resulting in many novel interaction techniques, toolkits, and applications [15], It has not yet been applied to the analysis of GES data.

In this paper, we present GestureExplorer, an immersive visualisation tool that uses 3D spatial arrangements to support gesture analysis and grouping in GES. As data exploration is an important part of this analysis process, GestureExplorer builds on prior gesture analysis tools by providing multiple interactive 3D gesture visualisations in a large virtual space to promote physical exploration. Our system contains numerous spatial visualisations for viewing and understanding 3D gesture data and leveraging spatial layouts to reveal relationships between gestures and clusters. We also include a wide range of interactive features for exploring various possible cluster configurations and making comparisons to define sensible groupings. Our spatial interface also affords interactions such as moving gestures between clusters and proposing embodied queries for finding particular gestures.

In summary, the contributions of this paper include:

- A novel interface with multiple interactive features and visualisations, exploiting spatial mappings for sense-making of gesture data.
- Novel combinations of clustering techniques and dimensional reduction algorithms for pattern identification of gesture data. And a benchmark of these combinations, evaluating their efficiencies.
- Two user studies with participants with varying experience in visualisation and IA to evaluate not only the visualisations and interactive features implemented, but also the performance of GestureExplorer in facilitating gesture analysis tasks in practice.

2 RELATED WORK

2.1 Gesture Elicitation Studies

Wobbrock et al. [51] coined the paradigm of gesture elicitation studies to explore user-defined gestures for interaction with new systems and applications in a way that reflects users' preferences [50, 52]. In a typical study, a set of desired functions, called 'referents' are given to participants, who will propose gestures to trigger each function. The researchers will then identify common gesture patterns from the collected data, with the aim of designing intuitive and discoverable gesture commands. Traditionally, the elicited gestures for review are recorded in an informal and unstructured way, such as in videos [27, 35]. Analysts must then manually inspect the videos and analyse the performed gestures, which takes a substantial amount of time, making GES poorly suited to large datasets. Furthermore, determining the similarity between gestures is highly subjective, and may result in groupings that do not accurately represent the users' preferences [49].

To resolve these drawbacks, Vatavu [49] came up with the first approach that exploits an objective distance algorithm (e.g., dynamic time warping or Euclidian distance) to aid agreement analysis among elicited gestures. Inspired by Vatavu's research [49], Dang and Buschek [12] made use of dynamic time warping with barycentre averaging [40] and proposed a method that not only finds the consensus among gestures but also computes an average gesture (barycentre) out of them. The computation of the average gesture enables clustering algorithms such as *k*-means clustering to be applied to the elicited gestures, which helps gesture designers and analysts identify groups of gestures that are in high consensus. However, two gestures may be considered similar by the automated approach, while being deemed semantically different from a human perspective. Therefore, human judgement needs to be introduced in the analysis process to help the algorithm make meaningful groupings [3]. Our tool builds on Dang and Buschek's analysis approach [12] with several new features that embody the human-in-the-loop principle [26] in a large immersive workspace.

2.2 Existing Visualisations for Gesture Data

Villarreal-Narvaez et al. [50] categorised 4 forms of visual representation for gesture data, ranging from formal to informal, structured to unstructured. The formal and structured representation, or 'ontology-based description', was frequently used in existing tools for GES analysis, allowing gestures embodied by such an ontology to be computationally analysed.

The intuitiveness of the visualisation of gestural data has a direct influence on the efficiency of pattern analysis in GES. A variety of attempts have been made in previous studies to visualise larger datasets. Some studies preprocess the dataset by grouping it into small clusters, which reduces the large amount of time and space required to investigate each gesture individually. Approaches for visualising these clusters include: 1) node representations for the clusters and links connecting nodes for the representation of the relationship between them [17, 25]; 2) a flat layout of all clustered data [9, 12, 22], in which a scatter plot or a density graph is drawn to provide an overview of the dataset; or 3) overlaying multiple items onto a single space [34] to allow swift identification of similarities and differences between them.

As for the visualisation of individual gesture data in a dataset, common approaches include: 1) drawing a trajectory of the motion [28, 38, 42]; or 2) picking frames of the gesture at fixed time intervals, then rendering the poses at the selected frames in a small-multiple plot [24, 42]. A variation to the latter method is to link a series of pre-processed gesture poses to represent that motion [12, 22]; 3) A third method is to provide playback animation of the gesture data represented by a skeleton ontology [12, 24, 25, 42]. Our tool provides all these approaches for users to choose from, and introduces a novel feature linking all (see Figure 1 right).

Building on these prior works, our immersive tool GestureExplorer investigates the use of a large virtual environment to provide a free perspective for inspecting with visualisations, to support an engaging experience for visual data exploration of gesture data.

2.3 Immersive Analytics

Research in IA [13] has demonstrated the benefits of applying emerging immersive display technologies in a range of areas, including road traffic data [41], biological data [10, 17, 19], geographic data [37, 43, 44], layout planning [21], programming [14], and even for the collaborative data analysis [29, 45].

Recent work has explored how interacting with embodied data constructs and arranging them in a user's surrounding virtual space can enhance data exploration and understanding. For instance, Cordeil et al. [11] proposed a system to explore multidimensional data in VR by manipulating virtual axes using natural interactions. Liu et al. [30] used a 'shelves' metaphor for manipulating arrangements of small multiples visualisations in 3D space. Hayatpur et al. [20] presented a visualisation system to trace data analysis steps in a virtual space utilising users' spatial ability.

Inspired by the potential of such embodied data exploration, our tool aims to similarly leverage the benefits of a large virtual space to support the exploration of gesture data. This work further overlaps with recent projects such as MIRIA [9] and AvatAR [42], which use Augmented Reality to support in-situ visual analytics of motion data. MIRIA provides abstract visualisation such as motion trajectories and heatmaps, while AvatAR bridges this by introducing a humanoid representation to visually retrace a person's activities in 3D space. In contrast to such tools aimed at providing environmental context through spatially situated visualisation, we aim to provide a more fully-featured visual analytics tool for exploring large gesture datasets in the context of GES analysis.

3 GESTUREEXPLORER: DESIGN RATIONALE

The main objective of a gesture elicitation study is to identify common patterns in gestures proposed by different users for a set of given gesture referents. This requires the analyst to inspect each referent one-by-one, to identify similar patterns, to semantically encode these patterns, and to group the gestures accordingly. Consensus can then be quantified for each referent. Traditionally, this process was done purely by manual inspection of data such as video recordings. However, Vatavu's initial exploration of algorithmic consensus metrics [49] opened the door to analysis tools that leverage computational processes [12, 49]. Thus our primary aim is to support semi-automated analysis features in a human-in-the-loop process to reduce the demands of this time-consuming process.

3.1 Feature Requirements

To support this overarching aim we identify several fundamental requirements for supporting data exploration and analysis for GES.

R1. **Computational measures to enhance sense-making** – Analysts rely on various measures to understand the overall agreement of collected gestures (e.g. Agreement score [51], Dissimilarity-consensus [49]) as well as to determine the proximity between gestures (e.g. Distance [24], Dissimilarity-variance [12, 49]).

R2. **Multiple views for gesture visualisation** – During data exploration, analysts often switch among different views to better understand data both individually (e.g. understanding the motion trend of a gesture) and overarchingly (e.g. identifying outliers in the dataset). In particular, GES tools should afford features that unfold the spatio-temporal nature of gesture data [12, 24]. In line with **R1**, different computational values could be embedded by various visualisations, which provide multiple perspectives [50].

R3. **Configurable and automated clustering algorithms to speed-up the pattern identification process** – The lengthy process of individually viewing and grouping a set of gestures manually has long been a barrier to the scalability of GES. To address this limitation, prior tools for GES have introduced automated clustering, using methods including *k*-means [12] and hierarchical clustering [49]. As it remains unknown which method is most useful for this purpose, GestureExplorer should provide multiple rationales to choose from, including the previously unexplored use of dimensional reduction algorithms.

R4. **Dynamic groupings for gestures** – In addition to **R3**, our system should give analysts the discretion to edit and refine the groupings of gestures that result from automated clustering [12, 24].

R5. **Comparison among gestures** – During the analysis process, analysts often need to make comparisons across gestures proposed by different participants and trials to understand differences and change their groupings [12, 24]. A variety of features should allow different types of comparison, for instance comparing spatial or temporal components of gestures, or comparing pairs of gestures versus comparing a single gesture against a group.

3.2 Key Concepts

Next we outline several key concepts that differentiate GestureExplorer from previous desktop tools.

C1. **Utilise physical space for the ability to 'explore'** – GestureExplorer focuses heavily on supporting the exploration aspect of the analysis process. Through parallax produced by body or head motions, the immersive view preserves depth cues of the gestures [33], allowing analysts to build an accurate mental representation of them. The large available space and 3D nature of a virtual environment provide an affordance for a metaphor of physical exploration among spatial data representations of gestures and their groupings [13, 16]. GestureExplorer takes motivation from recent work such as DataHop [20] and TimeTables [53], embedding similarity score of gestures with physical distance, turning the arrangement of data views in virtual space into a part of the analytical process.

C2. **Coordinated 2D and 3D views** – Following recent trends in Immersive Analytics [13] and building on basic gesture representations in MIRIA [9] and AvatAR [42], GestureExplorer goes beyond prior GES tools in supporting **R2**, by taking advantage of 3D space to arrange multiple linked 2D and 3D views. Whereas the immersive 3D view allows rich detail exploration and on-the-spot decision-making, 2D views can provide simplified abstract representations to compare clustering outcomes, or show a 2D overview map to allow more accurate distance estimations. Supporting both types of views provides analysts with flexibility in their analysis process.

C3. **Support engaging interactive gesture analysis** A central aspect of Immersive Analytics is the provision of engaging and embodied interactions to support analysis and decision-making [13]. GestureExplorer, likewise, provides abundant exploration space for visualisation views, where users can engage in a visceral experience of interaction, such as navigating among gestures at true scale, moving gestures between clusters, and engaging with multiple gestures simultaneously. Moreover, GestureExplorer takes inspiration from YouMove [4], enabling user-defined embodied search queries of the 3D gesture data. Such interactions exploit users' spatial abilities, which proved beneficial when exploring complex datasets [20].

4 DATA WRANGLING, PREPARATION, AND ANALYTICS

This section describes several steps that are required to wrangle and prepare the raw gesture data before applying the features that follow. All these steps are done within the tool and can be manually configured at any time.

4.1 Example dataset

For demonstration and testing, we use a dataset of full-body gestures collected by Vatavu [49]. This dataset contains 1312 gestures proposed by 30 children in response to 15 gesture referents, including "jump", "applaud", "draw a circle" and so on. Each gesture data contains the movement of 20 body joints and lasts varying frames. GestureExplorer provides built-in data structures and has wrangling methods implemented for Vatavu's dataset.

4.2 Average gesture with Dynamic Time Warping with Barycenter Averaging (DBA)

Inspired by GestureMap [12], we compute an average value among gestures. This computation is achieved by exploiting the DBA algorithm, initially proposed by Petitjean et al. [40]. The computed average gesture acts as a centroid for a cluster, enabling clustering algorithms such as k-means clustering to be applied to the dataset while maintaining most of the dataset's original dimensionality.

4.3 Data clustering

Clustering reduces the amount of manual work needed to identify similar gestures. We implement k-means and *mean shift* clustering algorithms to group gestures in a dataset.

K-means clustering. The first method we introduce is k-means. The original k-means approach requires a selection of centroids to initialise, which requires a time-consuming inspection of the entire gesture dataset before clustering [12]. To allow clustering to be provided upfront to support the exploration process, we use k-means++ [6] which requires users only to specify the number of initial centroids, while the choice of centroids will be made by the algorithm automatically.

Mean shift clustering. In contrast to k-means, which requires a human input of k to initialise, *mean shift* does the clustering based on a predetermined bandwidth. Two gestures will be grouped together if they are within each other's bandwidth. Gesture data are sampled into a fixed number of frames and flattened before they are fed to *mean shift* algorithm. Since the optimal bandwidth is unknown and may vary for different datasets, the bandwidth is determined by an estimation function implemented in scikit-learn [39] before clustering.

4.4 Dimensional reduction

Dimensional reduction methods are commonly used to reduce the complexity of high-dimensional data to make patterns more apparent. As prior implementations have not included dimensional reduction, we included this feature (based on participant feedback) to explore its applicability for gesture data. Before applying k-means or *mean shift* clustering, researchers can use either Principal Component Analysis (PCA) [48] or Metric Multi-Dimensional Scaling (MDS) [1] to first reduce the number of dimensions of the gesture data to 2. Treating these 2-dimensional values as x and y coordinates allows us to position gestures in the surrounding virtual environment (see PCA and MDS arrangements under Section 5.1.3).

4.5 Similarity and consensus among gestures

In GestureExplorer, we use the Dynamic Time Warping (DTW) algorithm to compute the similarity between two gestures [2, 8]. With the help of DBA, a cluster of gestures can be represented by the average value. Hence, we can measure the similarity between two clusters by computing the DTW distance between their average gestures. Likewise, we can find the similarity between a cluster and any given gesture. To find the consensus among gestures for a specific referent, we adopted the approach from GestureMap [12], which computes a consensus variance based on the mean of the sum of the DTW distance between each gesture and the average gesture of the dataset.

5 GESTUREEXPLORER: THE PROTOTYPE SYSTEM

We adopt an iterative methodology for design science research [31]. This process allows the designed features to evolve through extensive iteration during feature development, along with improvement based on feedback from multiple evaluation sessions (discussed in section 6). Our resulting design presents a substantial set of features to facilitate the analysis process in GES. These features cover the full range of analysis goals as defined in the visualisation analysis framework defined by Munzner [36].

Action	Visual Analytic Task	Feature in GestureExplorer
Analyse	Present	3D skeletons with static trajectories, Gesture animation, Small-multiples view, Gesture slider, Node-link view
	Derive	Gesture clustering, Average gesture
	Discover	PCA/MDS arrangement, Overview map
	Annotate	Gesture marking
Search	Explore	Global/Local arrangement, Overview map
	Browse	Cluster expansion, Line-up arrangement, Embodied search
	Locate	2D panel
	Lookup	Trajectory filter
Query	Compare	Trajectory stacking, Heat map, Close comparison, 2D panel, Change cluster
	Summarise	Overview map, Line-up arrangement, 2D panel

Table 1: Interactive features in GestureExplorer and their corresponding abstract action and visual analytic task defined by Munzner[36]

Our tool is implemented with an Oculus Rift S [2] in Unity 3D version 2020.3.22f1 [3]. The source code is publicly available and may be downloaded via GitHub: https://github.com/LeonLiAng929/ImmersiveGestureVisualizer.

Here, we introduce the interactive features in GestureExplorer. Throughout, we explain how these features benefit analysts, in the context of Munzner's Visualisation Analysis framework [36]. A summary of all features is shown in Table 1.

5.1 Analyse

Analyse is the highest level of action in Munzner's framework [36]. The aim of an analyst performing this action is to either understand the dataset or produce new information for later use. In the context of GES, analysts at this stage aim to gain a general sense of the gestures in the dataset, and to create an initial grouping for closer analysis.

5.1.1 Present. Present refers to the use of data visualisation to communicate information to the analyst [36]. In response to **R2** and **C2**, GestureExplorer provides multiple visual representations, introduced below, aimed at revealing spatial and temporal aspects of the data. The interactive environment **(**C3) allows analysts to follow Shneiderman's mantra of "overview first, then details on demand" [46] by looking at simpler views first and then expanding to detailed views when necessary. For instance, an initial scan of motion trajectories (Figure 2a) allows analysts quickly understand and compare gesture behaviours without the need for a close inspection of every gesture. For any gesture that does require closer inspection, analysts can expand its keyframes (Figure 2b) or play its animation (Figure 2c).

- **3D skeletons with static trajectories –** Typically, a skeleton is drawn out as nodes and bones. Each node represents a body joint, while bones connect the nodes to make the skeleton look human. This normalised and standardised visualisation enables further actions to be applied to it, such as animation [50]. As opposed to previous desktop tools [12, 24], we implement 3D body-scale skeletons. A skeleton consists of 20 nodes, each corresponding to a body joint (see Figure 2a). For each joint of a skeleton, a motion trajectory is drawn, with each joint in a different colour. The static trajectories provide an overview of spatio-temporal gesture data. We cover each skeleton with a semi-transparent cylindrical hull (Figure 2a) to provide a surface region for direct input.
- **Gesture animation –** Animation shows all the frames from the start to the end of a gesture sequentially. It is a fundamental feature that is implemented in almost all existing tools for gesture exploration [7, 10, 12, 22, 24, 25, 28, 38, 49]. The animation is looped, and the trajectories are revealed as the animation progresses (see Figure 2c).
- **Small-multiples view –** The small multiples view allows closer inspection by showing a sequence of keyframe poses. These poses are taken by dividing the normalised time duration into a set of regular intervals. Viewing these poses in a row provides an overview of how the gesture changes over time (Figure 2b). This view makes it convenient to observe

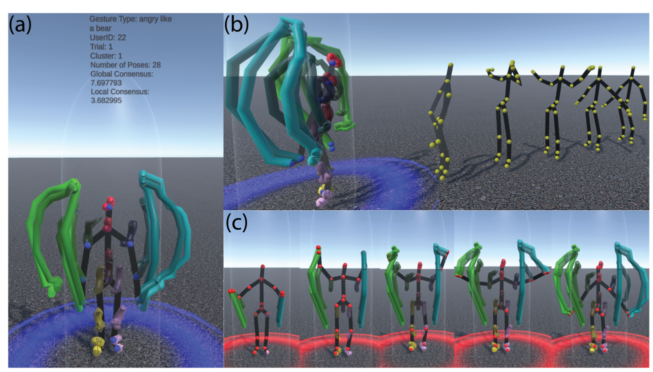


Figure 2: a) A 3D skeleton with static trajectories drawn around it. Trajectories for different body joints have distinct colours. Additional information about this gesture is displayed in a label above. b) The keyframes of a gesture are shown as small multiples behind its 3D skeleton-trajectory representation. c) Animation of a gesture, showing the continuous motion from left to right.

[2]Oculus Rift S: https://www.oculus.com/rift-s/
[3]Unity 3D: www.unity3d.com

the motion trend of a gesture and helps the user to understand the temporal nature of the gesture data [22, 24, 25, 38]. This can be viewed together with the animated skeleton overlaid on the small multiples, moving across the keyframes (see Figure 1 right).

- **Gesture slider** – Using a slider metaphor for time navigation, this feature allows users to control the time frame of the animation for closer inspection. We use the change of the horizontal orientation to control the time frame of gesture animation. Swinging the controller from right to left rolls back the animation.
- **Node-link view** – As the grouping process unfolds, a node-link view represents the clustering outcome. Each individual gesture is linked to its assigned cluster (introduced next, also see Figure 1 left). We distribute these in space to support physical exploration. The number of members in a cluster is encoded by the size of its surrounding bubble (Figure 3), which acts as an interactive surface for cluster interaction. For instance, if analysts apply the animation feature to a bubble, every gesture in the cluster will be animated. To associate each member gesture with its parent cluster, we assign them all the same colour (Figure 6). Visual connecting lines that reinforce these associations can be toggled on and off.

5.1.2 Derive. Derive creates new data based on existing data elements [36]. Analysts can derive sensible groupings based on the observation, as demanded in **R1** and **R3**. This is achieved by:

- **Gesture clustering** – To reduce the manual effort, GestureExplorer uses semi-automated clusterings to help analysts derive gesture groups. We employed a pipeline [4] connecting Unity to the python libraries, tslearn [47] and scikit-learn [39], to create the clusters, which are visualised at system start using the bubble metaphor described above. Users can choose either *k*-means or *mean shift* to cluster the data. In addition, GestureExplorer adds options for dimensional reduction, which is previously unexplored with gesture data and GES. Analysts may pre-process the data with the dimensional reduction rationales, as discussed in Section 5.1.3 below.
- **Average gesture** – To provide an overview of each cluster, we compute the barycentre average of all gestures contained. We embed this average skeleton and trajectory visualisation into the corresponding cluster bubble (Figure 4).

[4]https://docs.unity3d.com/Packages/com.unity.scripting.python@4.0

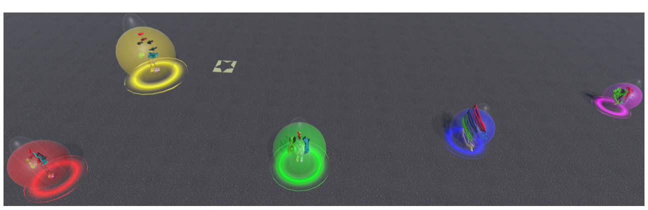


Figure 3: *K*-means++ clustering (*k*=5) results in 5 clusters, represented by uniquely coloured bubbles of different sizes. The size of a bubble reflects the number of gestures in the cluster. Each bubble contains a distinctive representation of the cluster average. The bubbles are arranged according to each cluster's DTW distance from the global average (explained in Section 5.2.1), encoded by their distance from the origin (denoted by a white star on the floor).

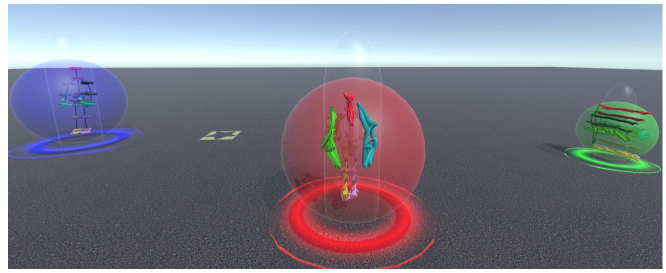


Figure 4: Embedded inside each cluster's main bubble node is a visualisation of the cluster's computed average gesture.

5.1.3 Discover. Discover refers to using visualisations to uncover information that was previously unknown [36]. At this point, analysts have had an initial look at the gestures but have not learned about their relationships. Embodying **C1** and **C2**, GestureExplorer supports discovery with the following features:

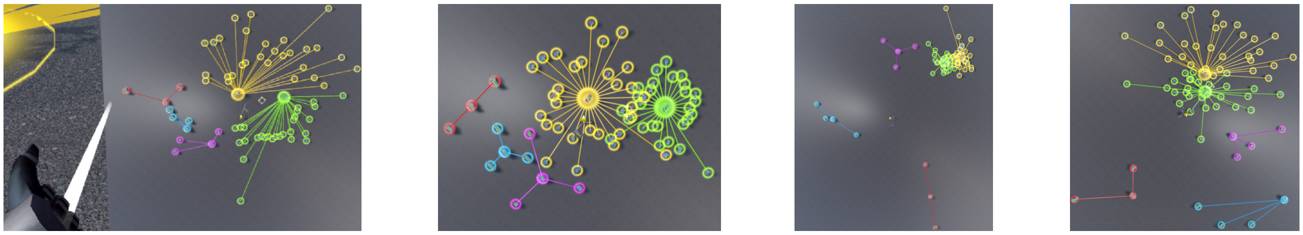
- ***PCA/MDS* arrangement** – In line with **R1**, the gesture data for a given referent is dimensionally reduced by *PCA* and *MDS* as described in Section 4.4. Analysts can observe the resulting distribution and use it to determine an initial value of *k* needed for *k*-means clustering. These arrangements also provide analysts with an alternate perspective on clustering results. For instance, one can initialise the k-means clustering with DBA, then visually inspect the result under the MDS or PCA arrangements to see how these differ (Figure 5c, 5d).
- **Overview map** – To assist accurate perception of distances and observation of outliers (see Section 5.2.1 below), we provide an *overview map* that shows a view from above, as shown in Figure 5 and Figure 7a. This map is attached to the left hand of the user. Users can press the left grip button to zoom in, and the right grip to zoom out.

5.1.4 Annotate. Annotate refers to the use of visual notation to highlight existing visualisations [36]. In our system, analysts can annotate data of interest during the initial discovery for later inspection (C3):

- **Gesture marking** – Introduced upon participants' feedback, gesture marking allows analysts to keep track of gestures of interest across different arrangements.

5.2 Search

Search is the middle-level action as defined by Munzner [36]. Its aim is to collect elements of interest for further operations. In the context of GES, analysts want to search for the most representative gesture pattern or anything that draws their interest after the initial clustering. This action features 4 different themes:



(a) (b) (c) (d)

Figure 5: Gesture arrangements. a) *Global* arrangement. Gestures are placed around the global origin point. The distance between a gesture and the origin denotes the similarity of that gesture to the average gesture of the entire dataset. b) *Local* arrangement. Gestures are placed relative to their assigned clusters. The distance between a gesture and its cluster denotes its within-cluster similarity. c) Principal Component Analysis (*PCA*) arrangement. d) Multi-dimensional Scaling (*MDS*) arrangement.

5.2.1 Explore. Analysts can start with an overview of everything to explore gestures of interest when they do not have any particular search targets [36]. In response to **R1**, **R2**, **C1**, and **C2**, GestureExplorer facilitates the exploration of clusters by using spatial distance to represent similarity (DTW distance) between gestures and clusters. This encoding reveals which items are closer to the average. In coordination with the overview map (Section 5.1.3), analysts can easily identify outliers in both the overall dataset and the derived clusters.

- ***Global* arrangement –** Under a *global* arrangement, all gestures and clusters are arranged around the global origin point of the workspace. We compute the average gesture for the entire dataset and measure the similarity of each gesture to this average gesture. Each gesture is then arranged around the origin at a corresponding distance. A similar procedure is followed in GestureMap [12], however, GestureExplor provides a baseline for interpreting these distances by representing the gestures at human scale. Unlike GestureMap, we also apply this concept to cluster averages, which are arranged similarly relative to the global average (Figure 5a).
- ***Local* arrangement –** Under a *local* arrangement, each gesture is arranged around its assigned cluster at a distance corresponding to its similarity with the cluster average (Figure 5b).

5.2.2 Browse. Browse allows users to search targets by characteristics [36]. In line with **R1**, **R2**, **C1**, and **C3**, analysts can browse gestures by 3 different characteristics, including their corresponding clusters, similarity to their cluster, and the particular behaviour they possess.

- **Cluster expansion –** C*luster expansion* allows analysts to browse gestures belonging to a given cluster. Analysts may trigger-select a cluster to expand or collapse its gestures to reduce visual clutter.
- **Line-up arrangement –** The *line-up* arrangement affords browsing gestures by their local similarity. This feature was introduced as per participants' feedback in the first round of user study 1 (discussed in Section 6.1). Under this arrangement, gestures are sorted based on their similarity to the average gesture of the assigned cluster, and then lined up in a row behind their respective clusters (Figure 7a). The clusters are arranged likewise based on their similarity with the dataset. This arrangement allows the analyst to easily confirm the similarity between numerous gestures when glancing along the line (Figure 7b). One can also traverse the lineup to identify a potential 'cutoff' point where a subset of gestures is dissimilar enough to be placed in a new cluster.

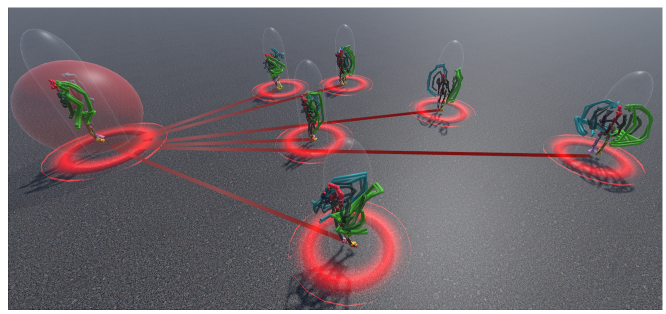


Figure 6: Member gestures of the red cluster after expansion. Each gesture is linked to the parent cluster's bubble node.

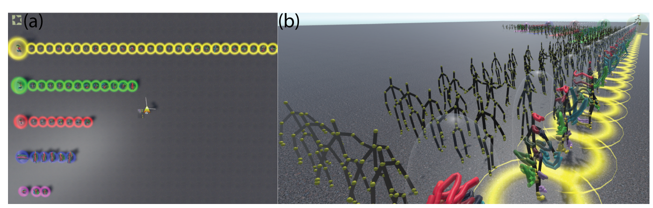
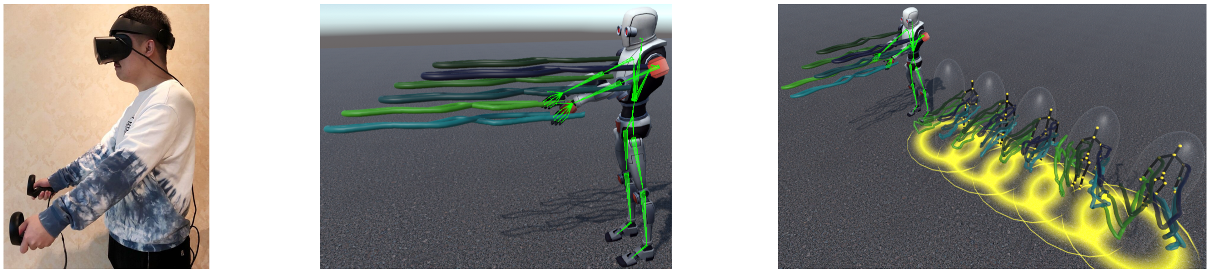


Figure 7: a) *Line-up* arrangement, in which gestures are lined up in order of similarity with the cluster average. Clusters are sorted with regard to their consensus with the entire dataset. b) The line-up arrangement with the expanded small multiples views.

- **Embodied search –** Embodied search supports browsing gestures by their motion. This feature was introduced after the first round of user study 1. It is similar to the search feature in YouMove [4], where a user-proposed gesture is used to search for existing body-exercising tutorials that perform similar movements. However, unlike YouMove [4],



(a) (b) (c)

Figure 8: Embodied search. a) User proposing an embodied gesture query with the controllers. b) The proposed query in the virtual environment, consisting of forward then backward motion with arms extended. c) The outcome of the search, 4 gestures whose similarity is within the defined tolerance. The gestures are lined up and sorted by their similarity to the proposed gesture.

which is an AR tool using motion capture cameras, our tool is based on VR and can only track hand movements via the 2 controllers.
In our implementation, the user is able to propose embodiment queries by holding the grip button on both controllers (Figure 8), and then match similar gestures in the dataset. The user-proposed gesture comprises movements of 6 body parts: left wrist, right wrist, left elbow, right elbow, left shoulder and right shoulder. We leverage inverse kinematics [5] to interpolate the movements for the remaining arm joints. We apply the DTW to identify gesture similarity within a predefined tolerance. The returned gestures are sorted by their similarity.

5.2.3 Locate. Locate is useful in the case where analysts have a clear target of interest but are unknown of its location [36]. In line with **C2** and **C3**, we support this via:

- **2D panel** – As per participants' feedback after the second round of user study 1, we provide a new 2D panel view each time a different clustering is applied. As shown in Figure 9, gestures on the panel are placed in a fixed order, allowing users to compare outcomes of different clustering settings. The frame surrounding a 2D gesture inherits the colour of the corresponding cluster. The 2D panel is coordinated with the 3D spatial view, allowing analysts quickly locate and interact with the 3D gesture visualisations via the panel. Analysts can apply any of the features described in Section 5 to the 2D gestures on the panel where applicable and then view the result on the corresponding 3D visualisation. Trigger-selecting a 2D gesture and then pressing the grip button teleports analysts to the position of the corresponding 3D gesture.

5.2.4 Lookup. When analysts have a target in mind and know where to find it, they can simply look for it [36]. GestureExplorer enables interactive lookup (**C3**) at the level of body parts, whose movement is represented by trajectories.

- **Trajectory filter** – As previous studies have reported visual occlusion when having too many trajectories drawn in the scene at once [28], we design a filtering feature to alleviate visual clutter and occlusion. As shown in Figure 10a, the trajectory filter has a skeletal shape. trigger-selecting the joints on the skeleton allows analysts to toggle on or off the trajectory of the chosen body part.

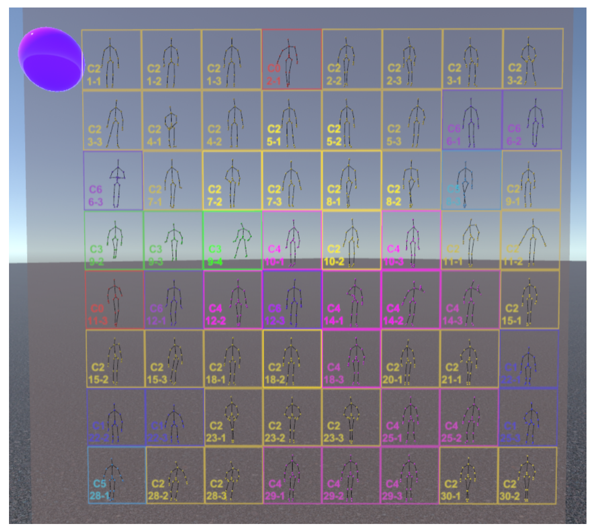


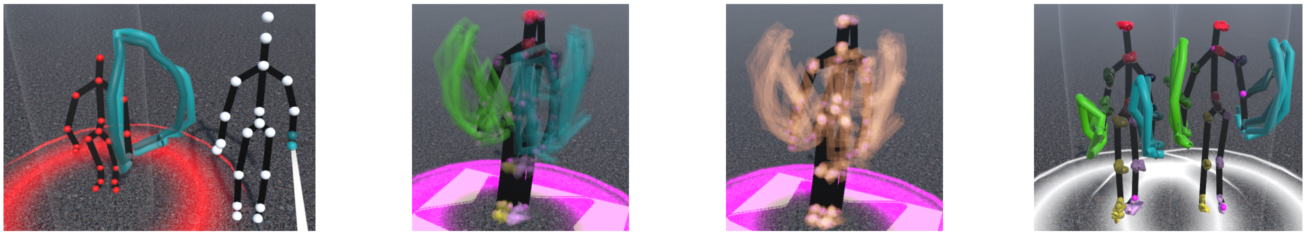
Figure 9: A 2D panel showing the clustering result. Users can quickly locate and interact with a 3D gesture object via its corresponding 2D gesture on the panel.

5.3 Query

Query is the lowest level of action according to Munzner [36]. The goal at this level is to finalise and refine data returned from a search. In the case of GES, Analysts need to refine the gestures and clusters searched for, to increase the accuracy of the identified gesture pattern.

5.3.1 Compare. Compare happens among multiple targets [36]. GestureExplorer supports comparisons among gestures in various ways, embodying **R2**, **R5**, **C2**, and **C3**:

- **Trajectory stacking** – Analysts can quickly identify similarities and differences among several gestures by stacking



(a) (b) (c) (d)

Figure 10: a) Trajectory filter, shown to the right of the skeleton-trajectory. In this figure, only trajectories for one hand of the gesture are selected, indicated by the coloured joints. b) Trajectory stacking of several gestures allows analysts to easily spot differences in gesture trajectories. c) Heat map for the stacked trajectories. Areas with stronger highlights are passed through by more gestures. d) Two gestures side-by-side for close comparison, the user can apply other features to the compared gestures for further exploration.

their trajectories (see Figure 10b). This approach borrows from tools in other domains [12, 34]. For instance, Dream-Lens [34] stacked different designs of a 3D model, allowing their combined opaqueness to reveal common design features. Likewise, using our feature, one can determine how similar the stacked gestures are by the transparency of the overlapped trajectories. When combined with Trajectory Filter (Section 5.2.4), the analyst can isolate the stacked trajectories of chosen body parts.

- **Heat map** – The heat map (see Figure 10c) is designed to enhance the contrast of the overlaid trajectories. We customise a semi-transparent shader that performs additive blending on the overlaid area, making the corresponding area more visible.
- **Close comparison** – Introduced in response to feedback from the first round of user study 1, this feature allows analysts to place a set of selected gestures side-by-side for close investigation. In combination with the presentation features described in Section 5.1.1, analysts can discern the nuances between these gestures.
- **Change cluster** – After a comparison, analysts may wish to group similar gestures together. Analysts can trigger-select any gestures, then trigger-select a cluster to move them there. When a gesture gets re-assigned from one cluster to another, the average gesture of both clusters will be updated (**R4**), along with the gesture's relative spatial position. Analysts can alternatively create a new cluster to hold the selected gestures or clusters. This feature introduces a necessary human element to the gesture grouping process.

In light of **C2**, comparisons can also be made among 2D panels. Suppose analysts process the dataset by different clustering configurations, as introduced in Section 5.1.2, multiple panels will be generated to register gesture information for each configuration. Analysts could then snap the panels side-by-side and annotate (Section 5.1.4) differences for further investigation.

5.3.2 Summarise. A summary involves all possible targets [36]. For instance, one may conclude their analysis by reviewing to the overview map (Section 5.2.1), the lineup arrangement (Section 5.2.2), and the 2D panel (Section 5.2.3) to observe the refined clusters and distil a final set of representative gesture patterns for the current referent.

5.4 Navigation and UI Interaction

Finally, we briefly explain the UI elements that allow users to access features. Users can move the right thumbstick to make a snap turn and move the left thumbstick to continuously move around gesture views. Or, they can trigger-select on the ground to teleport in the larger space. We implement an immersive menu (see Figure 11), containing interactive buttons representing features described in Section 5. Selection is powered by raycast of the controllers. Users can trigger-select a feature on the menu and trigger-select a cluster or a gesture to apply the feature to them. For further details, a video demonstrating the use of GestureExplorer in 3 scenarios is available in the supplementary materials.

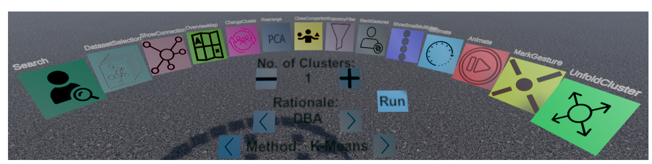


Figure 11: Feature menu in GestureExplorer, users can configure a clustering on the gesture dataset or select and apply features on the menu to the interactable gesture and cluster objects.

6 USER STUDIES

We conducted two user studies. The first study focused on evaluating the usability of features, based on feedback from PhD students and academics with experience in IA, data visualisation and HCI. The second study aimed to evaluate the tool with a diverse group of participants in a more realistic scenario, using a task that required participants to cluster a large set of gestures, to identify potential outliers, and to edit the clusters into appropriate matching groups of gestures. The virtual environment in GestureExplorer had a cloudy sky, and the trajectory visualisations were generated with a random colour for each body joint (see Figure 12). These were revised after

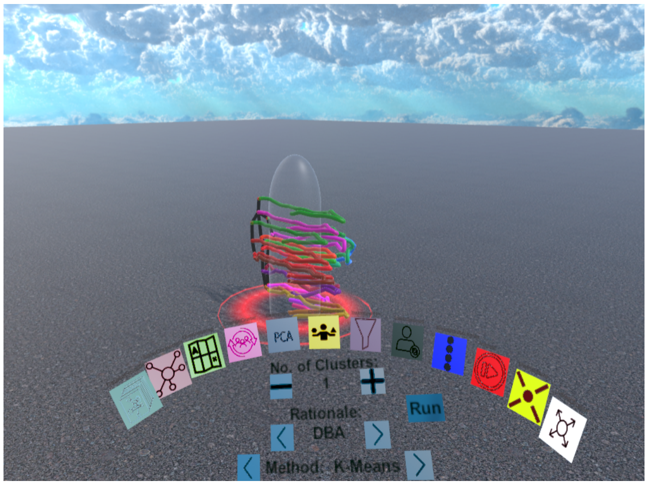


Figure 12: The version of GestureExplorer used in the user studies. Trajectories were assigned with randomly generated colours and the scene had a cloudy sky. After the user studies, we assigned similar colours to trajectories of body joints on the same body part and cleared the clouds in the sky.

the user studies and could be reflected by other figures presented in this paper.

6.1 Study 1: Iterative Feature Evaluation

Using an iterative design in 2 rounds, we conducted a usability test for the implemented visualisations and features. This allowed us to update several features based on participant feedback before the second round. After the first round, we polished the UI, improved the usability of existing features and introduced new features such as the *line-up arrangement* (5.2.2), *close comparison* (5.3.1) and *embodied search* (5.2.2) features based on collected feedback. Other features described in Section 5 are the final implementations. After the second round, we introduced the *2D panel* (5.2.3).

6.1.1 Participants. The study included participants from 3 universities across 3 countries to evaluate our tool. Each round featured 6 participants. Three participants from the first evaluation were invited to evaluate the improved prototype, alongside 3 newly invited participants in the second round. Out of the total 9 participants, 5 were male and 4 were female, with ages ranging from 21 to 49 years. All participants were academics ranging from PhD students to professors, having varying levels of expertise in Immersive Analytics, data visualisation and HCI. All participants owned or had access to a VR headset. We compiled and tested the system with several different brands of VR headsets, including Oculus Quest 2, HP Reverb, Samsung Odyssey, and Oculus Rift S.

6.1.2 Procedure. Due to the impact of the COVID-19 pandemic, all user studies were conducted remotely online via one-to-one meetings on Zoom. Each session lasted around an hour. Through the webcam and shared screen, we were able to communicate with the participants, and monitor their behaviour in GestureExplorer like in a face-to-face study.

This study aimed to collect expert feedback on our implemented features and visualisations. As a backdrop for training and using of these features, we used the task of exploring a single gesture referent, which was representative of how we expected such a tool would be used by researchers. For this we used a snippet of the dataset from Vatavu's gesture elicitation study [49], which contained the full set of 64 gesture proposals by 30 different participants for the referent "Angry Like a Bear".

Overview and training — Participants first got a clone of the tool on their local machine. Via a slide presentation shared on a video conference call, they were then introduced briefly to the background of our research, the visualisations, UI, and features implemented in our tool, as well as how to apply features to visualisation objects.

Experiencing the tool — Participants were asked to explore the example dataset with a given set of tasks using the features implemented in our tool. First, they were instructed to cluster the dataset and report the cluster with the most gestures. Then, they were asked to expand a cluster and use *gesture animation*, *gesture slider*, and *small-multiples* features on an individual gesture object, then to briefly describe the motion of the gesture.

Next, participants were asked to select and stack several gestures and identify the similarities and differences of the left-hand movement among the stacked gestures, with the help of *heat map* and *trajectory filter*. Then, with the *overview map* open and the *global arrangement* applied, participants were asked to identify the cluster that had the highest consensus with the entire dataset.

Lastly, participants were asked to switch the arrangement to *local*, identifying an outlier in the cluster they previously reported and assigning it to another cluster. In addition, participants in the second round were introduced to the improvements we had made, then spent some time experiencing them freely.

Questionnaire and feedback — We collected the demographic information of our participants via questionnaires. They were also asked to fill in an assessment questionnaire featured with Likert-based questions (scale 1-5) and some short-answered questions about various aspects of our tool.

6.1.3 Questionnaire Results. Overall, the responses we collected from the feedback questionnaire were positive. Participants generally agreed that our tool helped them in gaining an understanding of gesture data, and the various gesture visualisations could help them better understand not only the individual gestures but also the relationships between gestures and clusters.

As shown in Figure 13a, among the four implemented visualisations ("trajectory visualisation", "small multiples", "bubble representation", and "average gesture"), most participants considered the "Small-Multiples" visualisation "extremely useful" as it presented an insightful and concise overview for a gesture. The least favoured visualisation was the "Average Gesture", which was rated "slightly useful" by 3 participants. Participants who gave a low rating to this visualisation reasoned that because the average gesture was a purely computed value, the positions of the body joints sometimes changed abruptly, reducing the continuity of the trajectories and hence failed to provide a meaningful overview of a cluster. Some participants in the first round of studies considered the gesture trajectories visually "overwhelming", as the relative proportion of the trajectories was too large and often occluded the 3D skeleton.

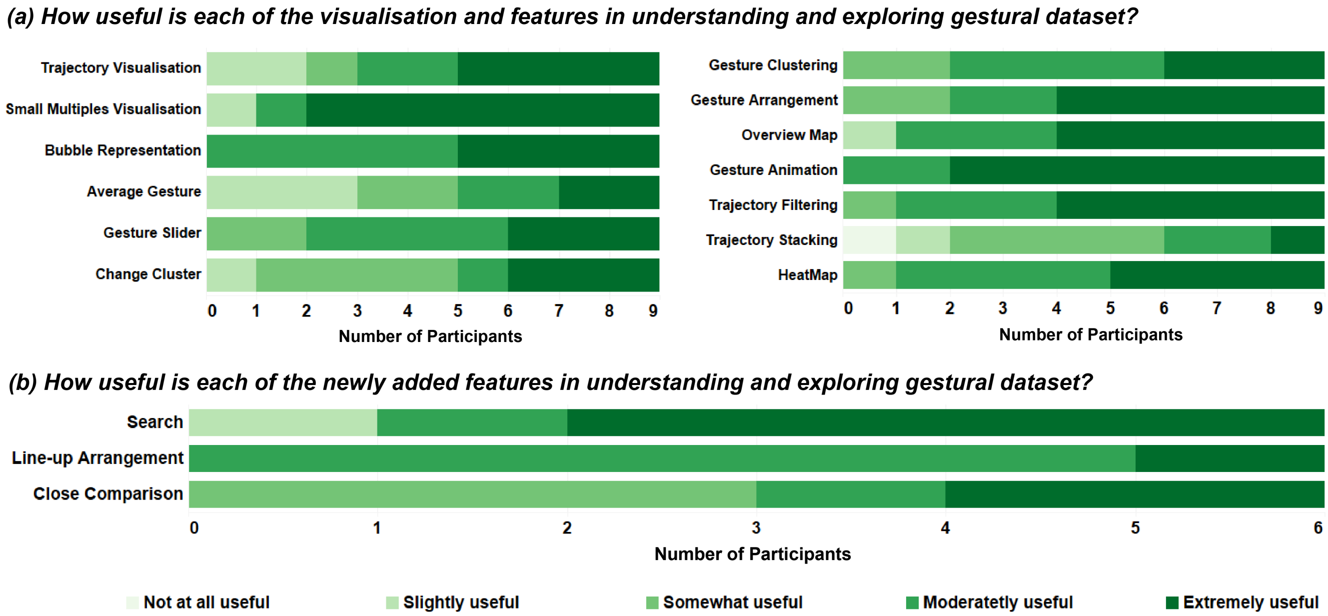


Figure 13: Participant responses. a) Combined responses from both study phases on the 4 visualisations and the interactive features. b) Responses from the second study phase on the newly introduced features.

This was no longer an issue in the second round, after we adjusted the proportions.

A majority of participants rated most features as either "Moderately Useful" or "Extremely useful" (Figure 13a). One exception was Trajectory Stacking", as participants found it difficult to identify similarities or differences among stacked gestures without the enhanced highlighting provided by the "Heat Map". The second exception was "Change Cluster". This was due to an issue with the cluster recalculation when a gesture was moved and was fixed for the second feedback round where it was rated favourably.

Most participants liked the newly introduced features (Figure 13b). In particular, the search feature was favoured highly. Participants commented, "This is a really cool feature and a very good use of the immersive environment," and "It is a very promising feature, as I've ever used such a search feature that involves physical movement/mimicking of actions."

Out of 9 participants, 7 indicated that they would prefer to use our immersive tool over a similar desktop tool for gesture exploration. However, one participant with expertise in interactive data visualisation commented, "I would not replace the traditional tool with this. I would instead use this tool as a complement to the traditional tools, especially for visualising the data." Another participant added, "Since I have a background in 'traditional' data visualisation, I actually preferred to see the clustering result in a 2D interface while viewing individual gesture data in the immersive tool". Legacy bias may exist as the two participants were both experienced in data visualisation on desktops. Nonetheless, they appreciated the immersive view of gestures in our tool.

6.2 Study 2: Clustering Task Evaluation

The first study allowed us to test and refine the features of GestureExplorer, however, did not provide us with insights into the performance of the tool in realistic tasks. To provide more holistic feedback on our prototype interface, we conducted the second study with a more heterogeneous group of participants. For this we simulated the analysis process of an open elicitation study [51], where analysts are required to identify distinct behaviours in the dataset and create sensible groupings of matching gestures.

We disabled the use of *gesture slider* (5.1.1), and *embodied search* (5.2.2) in the study, as these features may have hindered the timely progression of the task.

To determine which clustering method to use in our study, we ran a pilot evaluation of the clustering algorithms and various dimensional reduction rationales as introduced in Sections 4.3 and 4.4. The full evaluation is included in the supplementary materials. From these results, *mean shift* was found to have a worse performance than *k*-means clustering. We therefore disabled *mean shift* (5.1.2) and allowed participants to use only *k*-means clustering. All other features were enabled for the training and study.

6.2.1 Participants. The second study included 10 newly invited participants of diverse backgrounds, 6 male and 4 female, with ages ranging from 21 to 39. These participants were all students, recruited from various majors, varying between bachelor's, master's and PhD levels. Half of the participants had very high familiarity with VR, and the remainder had various degrees of experience with VR, including one with no familiarity. Only a few of them had some knowledge of gesture elicitation studies.

6.2.2 Procedure. Each session in the study lasted 90 minutes. Participants spent 40 minutes in training, 40 minutes in conducting the actual task, and the last 10 minutes in filling questionnaires. Observations were made throughout the session.

Dataset — To simulate a realistic task that could be completed within the duration of a typical VR study, we created a representative dataset using gesture data from Vatavu's study [49]. As this dataset contained many referents with unusually high agreement, due to didactic titles (e.g. "Hands up") aimed at understanding by children, we combined gestures from 3 different referents, including "Crouch", "Hands up" and "Applaud", but with their referent labels hidden to participants. Each referent contained 30 gestures, all proposed by different people, for a total of 90 gestures. Among these referents, "Crouch" was significantly different from the other two referents, while "Hands up" and "Applaud" both had a similar hands-raising motion but to different extents.

This mixture was representative of the variation that might be found in a typical study completed with 30 participants. It provided a variety of contrasting gestures to provide a challenging task and allowed us to evaluate the tool's performance in identifying both gesture patterns that were distinct from others as well as those that looked similar to each other. For training we created a separate dataset using 23 gestures from the referent "Angry like a bear". The training dataset was pre-clustered to ensure every participant would begin at the same place.

Training — Participants first were introduced briefly to the background and the purpose of our research, as most participants were not familiar with gesture elicitation studies. Then, they were asked to put on the VR headset and load the training dataset. Participants were introduced to the visualisations, UI, enabled features, and how to interact with these. The training ended with participants completing a simple task, in which they were required to find a gesture that we purposely put into a wrong cluster and to put it back into the correct one. They were suggested to make observations of gestures under different arrangements via the *overview map*, which is a useful technique to spot outliers.

Open elicitation task — Next, participants were asked to find gesture patterns in the dataset and group the gestures accordingly, simulating what analysts would do in an open elicitation study. During this task, participants were allowed to use the features of GestureExplorer freely. No instructions were given unless participants got stuck or asked for hints. Participants were expected to group the dataset into at least 3 clusters corresponding to the 3 referents. However, there exist sub-patterns of gestures within the same referent, and we left it up to the participants whether to split these sub-patterns into different clusters. Lastly, we measured the accuracy of participant-refined clusters using a metric we defined, as discussed next.

Accuracy metric — We defined a quantitative metric to help measure participants' grouping accuracy (we initially defined this metric to evaluate the clustering methods in the pilot evaluation mentioned in Section 6.2). We first labelled each gesture with its referent in the original dataset. Then we compared the referent labels between every pair of gestures in each resulting cluster. We considered a pair as correct if both gestures had the same label. To reflect the overall accuracy achieved by participants, we divided the total number of correct pairs found by the total number of gesture pairs present in all clusters.

Questionnaire and feedback — Participants were invited to answer a demographic and a feedback questionnaire, similar to Study 1 (Section 6.1).

6.2.3 Results. The accuracy of participants' refined datasets had a mean of 84.2% with a standard deviation of 13.3%. Half of the participants reached an accuracy above 90% at the end of the study. This is a leap compared to the accuracy at the beginning after a k-means clustering was initialised, whose accuracy varied roughly from 40% to 60% depending on the initial value of k and the selected pre-processing rationale.

As shown in Figure 14, we received positive feedback from participants in general. Despite their diverse backgrounds and little direct expertise, participants were in general consensus that the features in GestureExplorer not only helped them gain a better understanding of gesture data, but also allowed them to complete the task without much instruction.

We noticed some interesting cases where feedback for the implemented visualisations and features differed from the opinions received in Study 1. For instance, in the second study, the "Trajectory Visualisation" for gestures surpassed the "Small-Multiples" to become the most favoured visualisation, as it provided a quick way to understand the behaviour of a gesture and to compare gestures. Although "Small-Multiples" was deemed as the second most "extremely helpful" visualisation by participants, 3 of them considered it "not useful at all", possibly due to their limited experience in data visualisation.

Among the implemented features, "Gesture Animation" received unanimous approval (with similar results in the first study). All participants reckoned it was "extremely helpful" during the study. Many also commented it was the most "intuitive" way to inspect gestures in detail. Participants also acknowledged the usefulness of "Gesture Clustering", which dramatically saved time in identifying gesture patterns.

In response to the various gesture arrangements offered in GestureExplorer, the "Line-up" arrangement received the most positive feedback. "I really love this arrangement compared to the others, it is the simplest way to organise gestures, keeping them arranged compactly while allowing me to find the most similar gestures and potential outliers in a cluster at the head and the tail of the line respectively", said a participant. The newly introduced "PCA" and "MDS" arrangements were the second and third favoured arrangements. Many said the two arrangements were helpful to identify potential clusters when viewed on the *overview map* at the start of the study.

6.3 Discussion

In this section, we discuss the results of the user studies. In the first study, we validated the usability of GestureExplorer for gesture analysis and sense-making. Though many changes have been made based on participant feedback, we identified three important future improvements:

(1) Allow animation of user-proposed gestures in *search*, as they are currently visualised only via trajectories. The user can then replay the gesture in comparison with the animation of similar gestures in the dataset.
(2) The overview map could be replaced with a multidimensional scaling plot on a wall, to avoid occlusion of the scene caused by the handheld map.

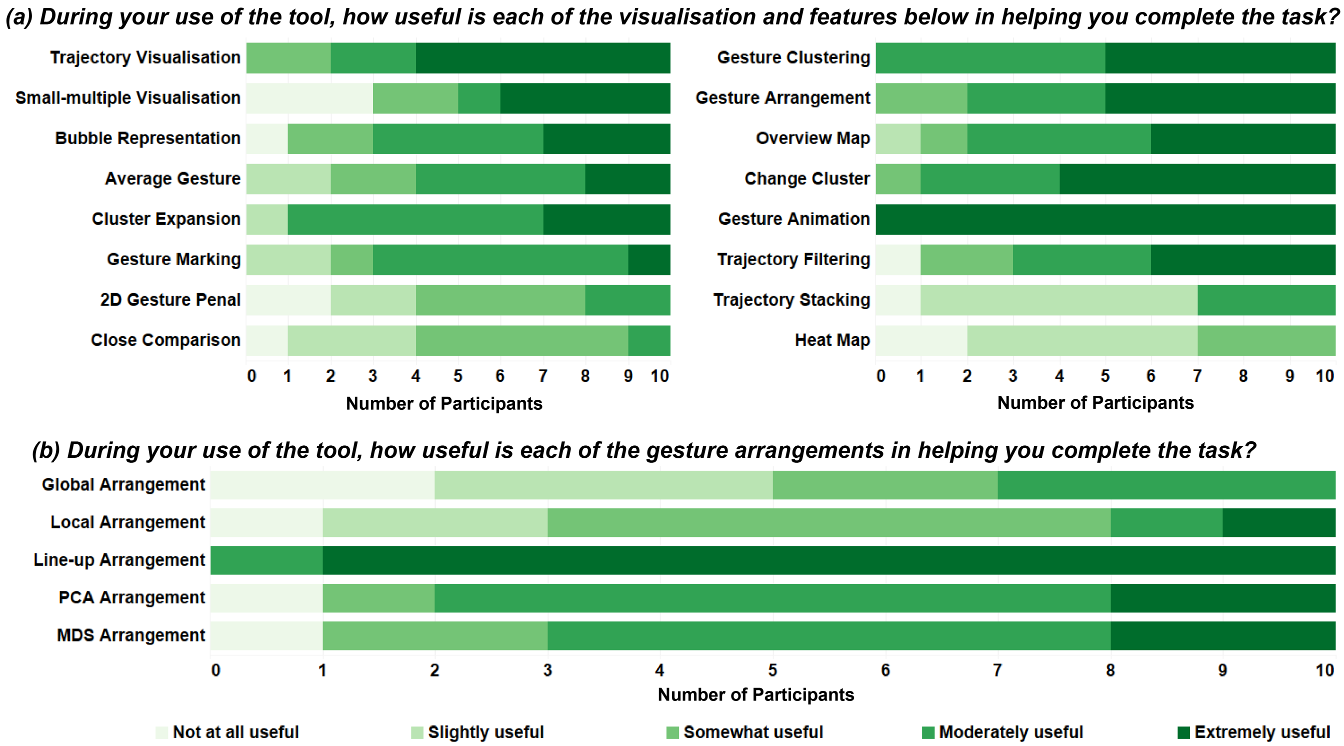


Figure 14: Participants’ responses on a) visualisations and interactive features, and b) the 5 gesture arrangements.

(3) Introduce a customised arrangement that lets the user organise gestures into different zones or meaningful positions, enhancing the sense-making process.

In the second study, with a more realistic gesture-grouping scenario, GestureExplorer demonstrated promising applicability for such gesture analysis tasks in practice. The ability of novice participants to complete this task unguided, is reflected by the high accuracy achieved. Participants reported a steep learning curve upon their initial use of the tool. It was challenging especially for participants without a relevant background. Nonetheless, as participants became more familiar with the tool, it was encouraging to see them being able to use GestureExplorer in their own ways. We report on notable behaviours in the following.

All participants initially approached the task by observing the overall distribution of gestures on the *overview map*, which was similar to the technique they were taught in the training session for outlier identification. Under either *PCA* (Figure 15a) or *MDS* arrangement (Figure 15b), most of the participants identified 2 potential clusters, while a few found 3 or more. Next, participants initialised a k-means clustering based on their initial observation and started refining the resulting clusters.

Each participant had their own way to identify distinct patterns. Though some showed minor differences, we were able to roughly divide their behaviours into two main analysis strategies:

(1) **Trajectory based analysis -** Some participants started refining a cluster by viewing trajectories or animations of all the gestures within. Although these participants had minimal experience in data analytics, they were able to make fast but rough comparisons by looking at the shape of trajectories, while animation offered an intuitive way to inspect gestures in detail. A few participants played the animation for the average gesture of each cluster. The average gesture of the cluster for "Crouch" had a smooth animation,

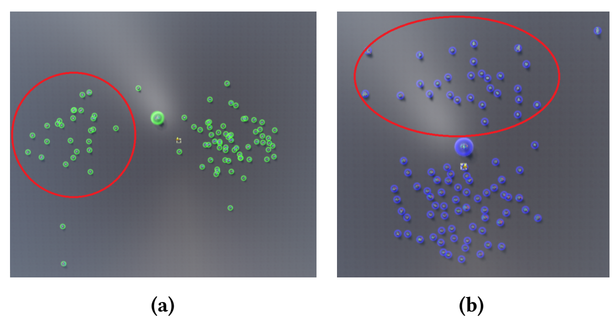


Figure 15: Gesture data for the open elicitation task of user study 2 under a) *PCA* arrangement and b) *MDS* arrangement. The dataset comprises gestures for "Crouch", "Hands up" and "Applaud". "Hands up" and "Applaud" have similar arm movements, while "Crouch" does not share any similarity. This is reflected in the *overview maps* above, which both show 2 distinct clusters of gestures with a small number of outliers. The cluster containing "Crouch" gestures (circled in red) is clearly separated from the other cluster, which contains a mixture of "Hands up" and "Applaud".

indicating low variance among the member gestures. The animation for the other cluster, which contained a mixture of "Applaud" and "Hands up" gestures, appeared "jerky" by comparison due to the different behaviours of the gestures. This behaviour prompted participants to investigate the cluster more closely, and eventually distinguish between "Hands up" and "Applaud" gestures.

(2) **Arrangement based analysis** - Other participants made greater use of the various gesture arrangements. While switching between different arrangements, participants first marked up outliers and marginal values for each cluster. Then, they made close comparisons between these gestures and gestures that were close to the cluster centroid. This approach also helped them find the two distinct behaviours in the mingled cluster.

During this process, a handful of participants were able to find sub-patterns in the "Crouch" cluster. They noticed in some gestures that the figure was standing back up again after the initial crouch. Hence they decided to split these gestures from the rest. Some other participants, despite also noticing this difference, considered the difference negligible and decided to keep those gestures within one cluster.

With patterns identified, participants proceeded to assign gestures into clusters with similar gestures. They did this in two ways. Some of them reset the value of k to the number of patterns observed and ran k-means clustering again, then sorted out the new clusters respectively, while others created an empty cluster using the "Change Cluster" feature, then assigned either "Hands up" or "Applaud" gestures to it. Many used the *line-up* arrangement for a final check, as it provided an organised view. Participants could simply go through each row of gestures to check if they all conform. Some participants stacked gestures for each cluster and viewed the alignment of the overlapped trajectories to determine if the cluster needed further refinement. However, one participant found an anomaly in the stacked gestures and attempted to select it, only to find out this feature was not supported, leaving us with an additional item for our list of future improvements.

7 CONCLUSION

In this paper, we proposed the first immersive, embodied tool with interactive features and visualisations that lets users explore large collections of recorded gesture data. We introduced various ways to pre-process and cluster gesture data. We evaluated GestureExplorer using two user studies. The first study validated the usability of GestureExplorer, and the second study demonstrated its ability to facilitate gesture analysis tasks in practice. We acknowledge that our studies are preliminary and investigated the use of the tool with a small number of participants. Future investigations, such as further evaluating the tool's immersive features like the Embodied Search, and comparison/field studies are needed to validate the tool for more generic, larger-scale applications.

One interesting outcome of our evaluation was a reflection on the potential benefits and limitations of immersive 3D data visualisations relative to 2D views. While 3D views provided a more 'complete' picture of the 3D gesture data, there were mixed preferences about the spatial arrangement of gestures in 3D virtual space. Some participants indicated the potential of using the immersive view as a companion to existing 2D tools. Conversely, additional 2D views can be integrated into the immersive space to provide the benefits of both in a single tool, for instance by placing a large 2D dashboard on one wall of the virtual space.

Nonetheless, we see value in the availability of a large 3D space for allowing participants to navigate among 3D gesture representations as part of the exploration process. The potential of this approach is reflected in the participants' favourable ratings of the Line-up arrangement. Our implementation was motivated by DataHop [20], which proposed the use of space to help users track their analysis history. However, there is further research needed to investigate the potential benefits of kinaesthesia and spatial memory to enhance such analysis tasks in a large virtual space.

Another aspect deserving further exploration is the potential to further integrate egocentric analysis into immersive systems such as GestureExplorer. Participants were particularly enthused by such a feature for proposing embodied search queries. Other features could further involve user activity in the analysis process, for instance by guiding users in recreating various gestures to better understand the body motions involved. Such features would be enhanced by additional tracking information of other body parts such as the torso, feet, head and fingers, as well as other physiological data such as electromyography (EMG) readings to understand muscle activity.

We would also like to investigate GestureExplorer with other types of gesture data, such as hand-tracking data, or user motion over larger areas of space. These future investigations may reveal other potential applications of GestureExplorer beyond gesture elicitation studies, which motivated this work.

REFERENCES

[1] 2005. *Metric and Nonmetric MDS.* Springer New York, New York, NY, 199–225. https://doi.org/10.1007/0-387-28981-X_9

[2] 2007. *Dynamic Time Warping.* Springer Berlin Heidelberg, Berlin, Heidelberg, 69–84. https://doi.org/10.1007/978-3-540-74048-3_4

[3] Daniel Abadi, Rakesh Agrawal, Anastasia Ailamaki, Magdalena Balazinska, Philip A. Bernstein, Michael J. Carey, Surajit Chaudhuri, Jeffrey Dean, AnHai Doan, Michael J. Franklin, Johannes Gehrke, Laura M. Haas, Alon Y. Halevy, Joseph M. Hellerstein, Yannis E. Ioannidis, H. V. Jagadish, Donald Kossmann, Samuel Madden, Sharad Mehrotra, Tova Milo, Jeffrey F. Naughton, Raghu Ramakrishnan, Volker Markl, Christopher Olston, Beng Chin Ooi, Christopher Ré, Dan Suciu, Michael Stonebraker, Todd Walter, and Jennifer Widom. 2014. The Beckman Report on Database Research. *SIGMOD Rec.* 43, 3 (2014), 61–70. https://doi.org/10.1145/2694428.2694441

[4] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2013. YouMove: Enhancing Movement Training with an Augmented Reality Mirror. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) *(UIST '13)*. Association for Computing Machinery, New York, NY, USA, 311–320. https://doi.org/10.1145/2501988.2502045

[5] A Aristidou, J Lasenby, Y Chrysanthou, and A Shamir. 2018. Inverse Kinematics Techniques in Computer Graphics: A Survey: Inverse Kinematics Techniques in Computer Graphics. *Computer graphics forum* 37, 6 (2018), 35–58.

[6] David Arthur and Sergei Vassilvitskii. 2007. K-Means++: The Advantages of Careful Seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (New Orleans, Louisiana) *(SODA '07)*. Society for Industrial and Applied Mathematics, USA, 1027–1035.

[7] Daniel Ashbrook and Thad Starner. 2010. MAGIC: A Motion Gesture Design Tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 2159–2168. https://doi.org/10.1145/1753326.1753653

[8] Donald J. Berndt and James Clifford. 1994. Using Dynamic Time Warping to Find Patterns in Time Series. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining* (Seattle, WA) *(AAAIWS'94)*. AAAI Press, 359–370.

[9] Wolfgang Büschel, Anke Lehmann, and Raimund Dachselt. 2021. MIRIA: A Mixed Reality Toolkit for the In-Situ Visualization and Analysis of Spatio-Temporal Interaction Data. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI ’21)*. Association for Computing Machinery, New York, NY, USA, Article 470, 15 pages. https://doi.org/10.1145/3411764.3445651

[10] D. Coffey, F. Korsakov, M. Ewert, H. Hagh-Shenas, L. Thorson, A. Ellingson, D. Nuckley, and D. F. Keefe. 2012. Visualizing Motion Data in Virtual Reality: Understanding the Roles of Animation, Interaction, and Static Presentation. *Computer graphics forum* 31, 3pt3 (2012), 1215–1224. https://doi.org/10.1111/j.1467-8659.2012.03114.x

[11] Maxime Cordeil, Andrew Cunningham, Tim Dwyer, Bruce H. Thomas, and Kim Marriott. 2017. ImAxes: Immersive Axes as Embodied Affordances for Interactive Multivariate Data Visualisation. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST ’17)*. Association for Computing Machinery, New York, NY, USA, 71–83. https://doi.org/10.1145/3126594.3126613

[12] Hai Dang and Daniel Buschek. 2021. *GestureMap: Supporting Visual Analytics and Quantitative Analysis of Motion Elicitation Data by Learning 2D Embeddings*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445765

[13] Tim Dwyer, Kim Marriott, Tobias Isenberg, Karsten Klein, Nathalie Riche, Falk Schreiber, Wolfgang Stuerzlinger, and Bruce H. Thomas. 2018. *Immersive Analytics: An Introduction*. Springer International Publishing, Cham, 1–23. https://doi.org/10.1007/978-3-030-01388-2_1

[14] Barrett Ens, Fraser Anderson, Tovi Grossman, Michelle Annett, Pourang Irani, and George Fitzmaurice. 2017. Ivy: Exploring Spatially Situated Visual Programming for Authoring and Understanding Intelligent Environments.

[15] Barrett Ens, Benjamin Bach, Maxime Cordeil, Ulrich Engelke, Marcos Serrano, Wesley Willett, Arnaud Prouzeau, Christoph Anthes, Wolfgang Büschel, Cody Dunne, Tim Dwyer, Jens Grubert, Jason H. Haga, Nurit Kirshenbaum, Dylan Kobayashi, Tica Lin, Monsurat Olaosebikan, Fabian Pointecker, David Saffo, Nazmus Saquib, Dieter Schmalstieg, Danielle Albers Szafir, Matt Whitlock, and Yalong Yang. 2021. *Grand Challenges in Immersive Analytics*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3446866

[16] Barrett Ens and Pourang Irani. 2017. Spatial Analytic Interfaces: Spatial User Interfaces for In Situ Visual Analytics. *IEEE Comput. Graph. Appl.* 37, 2 (mar 2017), 66–79. https://doi.org/10.1109/MCG.2016.38

[17] N. Férey, P. E. Gros, J. Hérisson, and R. Gherbi. 2005. Visual Data Mining of Genomic Databases by Immersive Graph-Based Exploration. In *Proceedings of the 3rd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia* (Dunedin, New Zealand) *(GRAPHITE ’05)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1101389.1101418

[18] A. Fonnet and Y. Prié. 2021. Survey of Immersive Analytics. *IEEE Transactions on Visualization and Computer Graphics* 27, 3 (2021), 2101–2122. https://doi.org/10.1109/TVCG.2019.2929033

[19] Yuki Fujiwara, Kohei Ichikawa, Haruo Takemura, et al. 2011. A multi-application controller for sage-enabled tiled display wall in wide-area distributed computing environments. *Journal of Information Processing Systems* 7, 4 (2011), 581–594. https://doi.org/10.3745/JIPS.2011.7.4.581

[20] Devamardeep Hayatpur, Haijun Xia, and Daniel Wigdor. 2020. DataHop: Spatial Data Exploration in Virtual Reality. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST ’20)*. Association for Computing Machinery, New York, NY, USA, 818–828. https://doi.org/10.1145/3379337.3415878

[21] Dominik Herrmann, Jan Reinhardt, Robert Krueger, Guido Reina, and Thomas Ertl. 2017. *Immersive Visual Analytics for Modular Factory Layout Planning*.

[22] Yueqi Hu, Shuangyuan Wu, Shihong Xia, Jinghua Fu, and Wei Chen. 2010. Motion track: Visualizing variations of human motion data. In *2010 IEEE Pacific Visualization Symposium (PacificVis)*. 153–160. https://doi.org/10.1109/PACIFICVIS.2010.5429596

[23] Jan-Paul Huttner and Susanne Robra-Bissantz. 2017. An Immersive Memory Palace: Supporting the Method of Loci with Virtual Reality.

[24] Sujin Jang, Niklas Elmqvist, and Karthik Ramani. 2014. GestureAnalyzer: Visual Analytics for Pattern Analysis of Mid-Air Hand Gestures. In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction* (Honolulu, Hawaii, USA) *(SUI ’14)*. Association for Computing Machinery, New York, NY, USA, 30–39. https://doi.org/10.1145/2659766.2659772

[25] S. Jang, N. Elmqvist, and K. Ramani. 2016. MotionFlow: Visual Abstraction and Aggregation of Sequential Patterns in Human Motion Tracking Data. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 21–30. https://doi.org/10.1109/TVCG.2015.2468292

[26] Waldemar Karwowski. 2006. *International encyclopedia of ergonomics and human factors*. Vol. 1. Taylor & Francis.

[27] Lawrence H. Kim, Daniel S. Drew, Veronika Domova, and Sean Follmer. 2020. User-Defined Swarm Robot Control. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI ’20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376814

[28] Simon Kloiber, Volker Settgast, Christoph Schinko, Martin Weinzerl, Johannes Fritz, Tobias Schreck, and Reinhold Preiner. 2020. Immersive analysis of user motion in VR applications. *The Visual computer* 36, 10-12 (2020), 1937–1949. https://doi.org/10.1007/s00371-020-01942-1

[29] Benjamin Lee, Maxime Cordeil, Arnaud Prouzeau, and Tim Dwyer. 2019. FIESTA: A Free Roaming Collaborative Immersive Analytics System. In *Proceedings of the ACM International Conference on Interactive Surfaces and Spaces* (Daejeon, Republic of Korea) *(ISS ’19)*. ACM, New York, NY, USA, 335–338. https://doi.org/10.1145/3343055.3360746

[30] Jiazhou Liu, Arnaud Prouzeau, Barrett Ens, and Tim Dwyer. 2020. Design and Evaluation of Interactive Small Multiples Data Visualisation in Immersive Spaces. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (VR ’20)*. IEEE, Atlanta, Georgia, USA, 588–597. https://doi.org/10.1109/VR46266.2020.00-23

[31] Salvatore T March and Gerald F Smith. 1995. Design and natural science research on information technology. *Decision Support Systems* 15, 4 (1995), 251–266.

[32] Kim Marriott, Jian Chen, Marcel Hlawatsch, Takayuki Itoh, Miguel A. Nacenta, Guido Reina, and Wolfgang Stuerzlinger. 2018. *Immersive Analytics: Time to Reconsider the Value of 3D for Information Visualisation*. Springer International Publishing, Cham, 25–55. https://doi.org/10.1007/978-3-030-01388-2_2

[33] Kim Marriott, Jian Chen, Marcel Hlawatsch, Takayuki Itoh, Miguel A. Nacenta, Guido Reina, and Wolfgang Stuerzlinger. 2018. Immersive Analytics: Time to Reconsider the Value of 3D for Information Visualisation. In *Immersive Analytics*. Springer International Publishing, Cham, 25–55.

[34] Justin Matejka, Michael Glueck, Erin Bradner, Ali Hashemi, Tovi Grossman, and George Fitzmaurice. 2018. *Dream Lens: Exploration and Visualization of Large-Scale Generative Design Datasets*. Association for Computing Machinery, Paper 369. https://doi.org/10.1145/3173574.3173943

[35] Meredith Ringel Morris. 2012. Web on the Wall: Insights from a Multimodal Interaction Elicitation Study. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces* (Cambridge, Massachusetts, USA) *(ITS ’12)*. Association for Computing Machinery, New York, NY, USA, 95–104. https://doi.org/10.1145/2396636.2396651

[36] Tamara Munzner. 2015. Why: Task Abstraction. In *Visualization Analysis and Design*. A K Peters/CRC Press, 43–65.

[37] Rhys Newbury, Kadek Ananta Satriadi, Jesse Bolton, Jiazhou Liu, Maxime Cordeil, Arnaud Prouzeau, and Bernhard Jenny. 2021. Embodied gesture interaction for immersive maps. *Cartography and Geographic Information Science* 48, 5 (2021), 417–431. https://doi.org/10.1080/15230406.2021.1929492 arXiv:https://doi.org/10.1080/15230406.2021.1929492

[38] Masaki Oshita. 2019. Motion Volume: Visualization of Human Motion Manifolds. In *The 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry* (Brisbane, QLD, Australia) *(VRCAI ’19)*. Association for Computing Machinery, New York, NY, USA, Article 7, 7 pages. https://doi.org/10.1145/3359997.3365684

[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[40] François Petitjean, Alain Ketterlin, and Pierre Gançarski. 2011. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition* 44, 3 (2011), 678–693. https://doi.org/10.1016/j.patcog.2010.09.013

[41] Arnaud Prouzeau, Anastasia Bezerianos, and Olivier Chapuis. 2016. Towards road traffic management with forecasting on wall displays. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*. ACM, Niagara Falls, Ontario, Canada, 119–128. https://doi.org/10.1145/2992154.2992158

[42] Patrick Reipschläger, Frederik Brudy, Raimund Dachselt, Justin Matejka, George Fitzmaurice, and Fraser Anderson. 2022. AvatAR: An Immersive Analysis Environment for Human Motion Data Combining Interactive 3D Avatars and Trajectories. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI ’22)*. Association for Computing Machinery, New York, NY, USA, Article 23, 15 pages. https://doi.org/10.1145/3491102.3517676

[43] Kadek Ananta Satriadi, Barrett Ens, Maxime Cordeil, Tobias Czauderna, and Bernhard Jenny. 2020. Maps Around Me: 3D Multiview Layouts in Immersive Spaces. *Proc. ACM Hum.-Comput. Interact.* 4, ISS, Article 201 (Nov. 2020), 20 pages. https://doi.org/10.1145/3427329

[44] Kadek Ananta Satriadi, Barrett Ens, Tobias Czauderna, Maxime Cordeil, and Bernhard Jenny. 2021. *Quantitative Data Visualisation on Virtual Globes*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445152

[45] Willy Scheibel, Matthias Trapp, and Jürgen Döllner. 2016. Interactive Revision Exploration using Small Multiples of Software Maps.. In *VISIGRAPP (2: IVAPP)*. 133–140.

[46] B. Shneiderman. 1996. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*. 336–343. https://doi.org/10.1109/VL.1996.545307

[47] Romain Tavenard, Johann Faouzi, Gilles Vandewiele, Felix Divo, Guillaume Androz, Chester Holtz, Marie Payne, Roman Yurchak, Marc Rußwurm, Kushal Kolar, and Eli Woods. 2020. Tslearn, A Machine Learning Toolkit for Time Series Data. (2020).

[48] Michael E Tipping and Christopher M Bishop. 1999. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society. Series B, Statistical methodology* 61, 3 (1999), 611–622.

[49] Radu-Daniel Vatavu. 2019. *The Dissimilarity-Consensus Approach to Agreement Analysis in Gesture Elicitation Studies*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300454

[50] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob O. Wobbrock. 2020. A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies?. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 855–872. https://doi.org/10.1145/3357236.3395511

[51] Jacob O. Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A. Myers. 2005. Maximizing the Guessability of Symbolic Input. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. Association for Computing Machinery, New York, NY, USA, 1869–1872. https://doi.org/10.1145/1056808.1057043

[52] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-Defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1083–1092. https://doi.org/10.1145/1518701.1518866

[53] Yidan Zhang, Barrett Ens, Kadek Satriadi, Arnaud Prouzeau, and Sarah Goodwin. 2022. TimeTables: Embodied Exploration of Immersive Spatio-Temporal Data. 599–605. https://doi.org/10.1109/VR51125.2022.00080

[54] Arthur Zimek, Erich Schubert, and Hans-Peter Kriegel. 2012. A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 5, 5 (2012), 363–387. https://doi.org/10.1002/sam.11161 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/sam.11161