A Computational Method for the Classification of Mental Representations of Objects in 3D Space

Samuel S. Sohn
Rutgers University, Piscataway, NJ, USA

Panagiotis Mavros[1]
Singapore-ETH Centre, Future Cities Laboratory, CREATE campus, 1 CREATE Way, #06-01 CREATE Tower, 138602, Singapore

Mubbasir Kapadia
Rutgers University, Piscataway, NJ, USA

Christoph Hölscher
Chair of Cognitive Science, ETH Zürich, Switzerland
Future Cities Laboratory, Singapore-ETH Centre, Singapore

Abstract

The structure mapping task is a simple method to test people's mental representations of spatial relationships, and has recently been particularly useful in the study of volumetric spatial cognition such as the spatial memory for locations in multilevel buildings. However, there does not exist a standardised method to analyse such data and structure mapping tasks are typically analysed by human raters, based on criteria defined by the researchers. In this article, we introduce a computational method to assess spatial relationships of objects in the vertical and horizontal domains, which are realized through the structure mapping task. Here, we reanalyse participants' digitised structure maps from an earlier study (N=41) using the proposed computational methodology. Our results show that the new method successfully distinguishes between different types of structure map representations, and is sensitive to learning order effects. This method can be useful to advance the study of volumetric spatial cognition.

2012 ACM Subject Classification General and reference → Metrics; General and reference → Experimentation; General and reference → Evaluation

Keywords and phrases mental representations of space, spatial cognition, structure mapping task, 3D space, volumetric space

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.20

Category Short Paper

Funding The research was conducted at the Future Cities Lab, Singapore-ETH Centre and supported by the National Research Foundation, Singapore under the CREATE programme, as well as NSF Awards: IIS-1703883, IIS-1955404, IIS-1955365, RETTL-2119265, and EAGER-2122119.

1 Introduction

For many species, navigation entails movement not only in the horizontal plane, but also, to some extent, in the vertical domain. Aerial species need to coordinate flight to find their nest or food on trees. Underwater species coordinate movement in different depths to seek shelter and resources. Terrestrial species may traverse undulating terrain or climb atop objects and surfaces. As humans, we also climb surfaces, live in multi-storey structures, and routinely organise objects and other information in vertical space.

[1] Corresponding Author

15th International Conference on Spatial Information Theory (COSIT 2022).
Editors: Toru Ishikawa, Sara Irina Fabrikant, and Stephan Winter; Article No. 20; pp. 20:1–20:8
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Previous research in spatial cognition has shown that people organise spatial information hierarchically [13], grouping objects based on proximity, visual, or semantic salience [3]. This process is also called “regionalisation”, and people use the resulting hierarchies, or regions, to plan spatial behaviour [13]. While extensive research has focused on the cognition of 2D surfaces (layout rooms, buildings, or neighbourhoods), recently research has turned to how people perceive, form mental representations of, and reason about 3D and volumetric spaces; here we are particularly interested in navigating through buildings.

Various methods have been developed to access and assess mental representations of space. These include onsite and offsite pointing, identifying novel shortcuts, sketch-mapping, and others [7]. To assess how people perceive, understand, and utilise 3D spatial relationships, previous research in volumetric spatial cognition has relied mostly on 3D pointing tasks [16, 9] or navigational tasks [6, 8, 4]. One limitation of pointing tasks is that they examine pairwise (or triplewise in the case of judgements of relative direction) spatial relationships; the overall spatial organisation of multiple locations is only indirectly assessed through the accuracy in multiple trials between separate locations. Similarly a limitation of multilevel navigation tasks is that the mental representation is confounded with the immediate spatial information available to the individual, e.g., the visibility of spaces, decision points, landmarks, and signs.

An alternative approach that captures in an abstract manner the spatial organisation of multiple objects is the structure mapping task (SMT). In the SMT, as it has been applied to study volumetric spatial cognition [1, 10], individuals are provided with a set of representations for physical locations (e.g. a small card or object) and are asked to place them on a two-dimensional (2D) surface in a manner that represents how these locations are organised in space.

In this study, we seek to develop a new, computational approach to analyse 2D representations of 3D spatial information. This approach is motivated by and applied to the analysis of structure mapping tasks. In the following sections we describe the analytical framework developed to distinguish between horizontally-biased and vertically-biased representations, and we apply this to behavioural data obtained from a previous study. Below, we summarise key aspects of the previous study, articulate in detail the mathematical description of the new metrics, and use them to assess the representations produced by human participants.

2 Methods

2.1 Data collection

Data collection was conducted as part of a larger experiment which is presented in more detail in [10]; here we include relevant information to assist readers understand the structure mapping task analysis. Participants (N = 41) learned the layout and the location of twelve goal locations (shops) spread across four floors of a large, complex, multilevel building (Figure 1).

Participants were randomly assigned in two spatial learning groups. The horizontal training group learned the locations by walking to all locations in a floor before moving to the next floor. By contrast, the vertical training group walked to the locations by first visiting all locations of a vertical cluster (using the escalators) before moving to the locations of the next cluster (Figure 1). This allowed us to test the effect of spatial knowledge acquisition mode (learning) on mental representation structure. After the training phase, they completed a structure mapping task which is the focus of this paper, and then proceeded with the rest of the experiment [10].

For the structure mapping task, 12 cards representing each goal were shuffled and given to them, with theinstruction *"how do you think they are arranged in space"* without at time limit. Later, they self-reported their sense of direction using SBSOD [5], prior familiarity with the building, as well as demographic information.

3 Analysis

Each goal location (landmark) from the structure mapping task corresponds to a recalled 2D position in each participant's representation of the building. The same landmark is also associated with a real 3D position computed from the digital floorplan of the building. However, the relationships between the recalled positions and real positions learned by participants remain unknown, e.g., whether their representations were tied to a particular perspective of the building. In order to uncover these relationships and assess their accuracy, we propose two analysis techniques catered to the structure mapping task.

Morphological analysis uses *recalled* positions to determine how well a participant's representation conforms to canonical organisations of *real* landmark positions, e.g., by the floor of the building. This requires that an experimenter make assumptions about which landmarks should be canonically grouped. On the other hand, functional analysis uses clustering algorithms to group a participant's recalled positions and determines what type of information the clusters are useful for distinguishing about real positions. This technique requires the experimenter to make assumptions about the utility of clusters instead of the clusters themselves, making it complimentary to morphological analysis.

3.1 Morphological Analysis

A canonical organisation of landmarks is defined as a set of clusters. For multilevel buildings, we consider two types of canonical clusters based on the floors of the building (labeled 1-4) and corridors (labeled A-D), which in our case are consistent between floors. Henceforth, we refer to floor-based clusters as horizontal clusters C^H and corridor-based clusters as vertical clusters C^V. For a given participant, we encoder their clusters $C \in \{C^V, C^H\}$ as a sequence of m clusters and each cluster C_i as a sequence of recalled 2D positions $C_{i,k}$. The morphological analysis quantifies two characteristics of a participant's representation for each type of canonical clustering C: the distinctiveness $\mathcal{D}(C)$ and the alignment $\mathcal{A}(C)$.

3.1.1 Distinctiveness

The distinctiveness metric $\mathcal{D}(C)$ measures the separateness between clusters by considering them as polygons (Eq. 1). In particular, we represent each cluster as a convex hull, i.e., the smallest convex polygon that encompasses all of its recalled positions. The separateness between clusters can then be computed using function $\Delta(C_i, C_j)$, which outputs 0 when the convex hulls of clusters C_i and C_j are intersecting and 1 otherwise. This means that two clusters are only considered separate when their recalled positions are visually separable by a line. The distinctiveness metric $\mathcal{D}(C) \in [0, 1]$ encodes the probability that a cluster is visually separable from all other clusters.

$$\mathcal{D}(C) = \frac{1}{m} \sum_{i=1}^{m} \prod_{j=1}^{m} \Delta(C_i, C_j) \;\Big|\; i \neq j \tag{1}$$

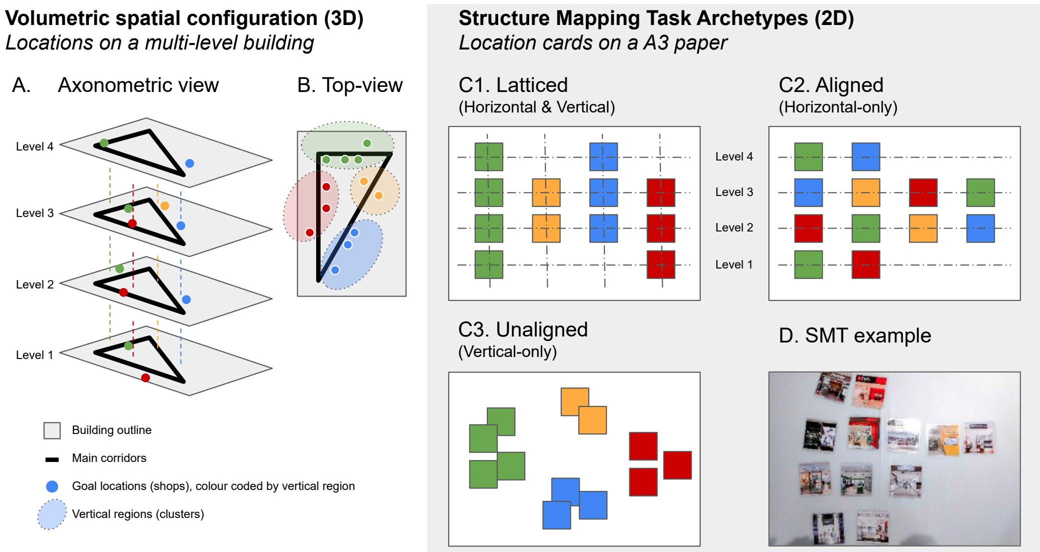

Figure 1 Schematic diagrams of the physical volumetric space, a large multilevel building, in axonometric-view (A) and top-view (B). There were 2 goal locations on level 1 (ground-floor), 4 on level 2, 4 on level 3, and 2 on level 4. (C) Three major archetypes of structure mapping task products (C1–3) that capture horizontal and vertical rationalisations of the same 3D spatial configuration.

3.1.2 Alignment

The alignment metric $\mathcal{A}(C)$ averages two terms: the eccentricity of clusters and the similarity between their orientations (Eq. 2). Both terms rely on Principal Component Analysis (PCA) [15] to compute the principal components $\Psi(C_i)$ and eigenvalues $\Lambda(C_i)$ of each cluster C_i using the distribution of its recalled positions. The first principal component $\Psi(C_i)_1$ for cluster C_i is the axis or unit vector along which the variance in recalled position $\Lambda(C_i)_1$ is maximized. The second principal component $\Psi(C_i)_2$ is orthogonal to the first, and it accounts for the next highest variance $\Lambda(C_i)_2$. In effect, the principal components act as the major and minor axes of an ellipse fitted to the cluster [14].

$$\mathcal{A}(C) = \frac{1}{2m} \sum_{i=1}^{m} \sqrt{1 - \frac{\Lambda(C_i)_2^2}{\Lambda(C_i)_1^2}} + \frac{1}{m(m-1)} \sum_{i=1}^{m} \sum_{j=i+1}^{m} \left|\Psi(C_i)_1 \cdot \Psi(C_j)_1\right| \tag{2}$$

The eccentricity term measures the average elongatedness of a cluster using the formula for the eccentricity of an ellipse [12]. For cluster C_i, this is computed as the square root of 1 minus the ratio between $\Lambda(C_i)_2^2$ and $\Lambda(C_i)_1^2$. This can be thought of as the ratio between the major and minor axes of an ellipse. The similarity term computes the average angular similarity between all unique pairs of clusters. The angular similarity between clusters C_i and C_j is formulated as the absolute value of the dot product between their first principal components, $\Psi(C_i)_1$ and $\Psi(C_j)_1$. The absolute value ensures that vectors facing in opposite directions are considered equal. Since alignment $\mathcal{A}(C)$ averages the eccentricity and similarity terms, it is maximized when clusters are both linear and parallel to each other.

3.1.3 Structure Mapping Task Archetypes

Distinctiveness is first measured for both horizontal and vertical clusters to assess whether participants are cognizant of them. For participants with distinctive representations, alignment is then measured to classify the appearance of the representation. When only one canonical type is distinctive, an *aligned* representation configures the clusters into either rows or columns, and an *unaligned* representation separates clusters into regions independent of any axis. It is possible for a representation to be aligned with respect to multiple canonical types. When these types have very little in common between their clusters, e.g. horizontal and vertical clusters, their axes of alignment can become orthogonal to each other. This causes horizontal clusters to become rows and vertical clusters to become columns or vice versa. We consider this special case as a *latticed* representation. We consider a *non-canonical* representation to be nondistinctive ($\mathcal{D}(C) < 1$) with respect to all canonical types. These archetypes align with the manual classification of structure maps done by Mavros et al. [10].

3.2 Functional Analysis

For participants with non-canonical representations according to the morphological analysis, functional analysis can offer an explanation. First, clusters are extracted from a participant's representation using a set of methods chosen by the experimenter. In our case, the methods are all instances of HDBSCAN [2] using different distance functions based on archetypes from the morphological analysis: Euclidean distance for clustering unaligned representations, Manhattan distance for latticed representations, and two variants of Manhattan distance for aligned representations, which weight the distances along the x- and y-axes with $\langle 0.8, 0.2\rangle$ and $\langle 0.2, 0.8\rangle$ before summing them. Manually annotated clusters by either the experimenter or participants can also be used as a method.

$$\mathcal{I}(C, \Phi) = \frac{1}{m} \sum_{i=1}^{m} \prod_{j=1}^{m} \Phi(C_i, C_j) \;\Big|\; i \neq j \tag{3}$$

The clusters C of each method are then judged by an informativeness metric $\mathcal{I}(C, \Phi)$ (Eq. 3) with respect to a function Φ, and only the best method is considered. $\Phi(C_i, C_j)$ is defined as any function that outputs 1 when two clusters C_i and C_j are separable according to a metric over *real* landmark positions (not *recalled* positions) and outputs 0 otherwise. This means that $\mathcal{I}(C, \Phi)$ evaluates the probability that a cluster can be distinguished from all other clusters according to function Φ, which is comparable to the formulation of distinctiveness $\mathcal{D}(C)$. We consider two metrics for Φ: the horizontal Euclidean distance along the xy-plane (used by Φ^H) and the vertical distance along the z-axis (used by Φ^V). Both functions Φ^H and Φ^V first compute the real centroid (i.e., the average real position) of landmarks in C_i and then compute whether all landmarks in C_i are closer to the centroid than those in C_j using the respective metric. With more data about the landmarks and the building, more metrics can be explored, e.g., the metabolic energy used when travelling between landmarks [11]. However, some complex metrics may be better suited for morphological analysis, because they are difficult to formulate computationally, e.g., the visual or semantic saliency of landmarks.

4 Results

Morphological analysis revealed that among the 41 participants, 7 people (17.1%) had aligned representations ($\mathcal{A}(C) \geq 0.9$), 5 people (12.2%) had unaligned representations ($\mathcal{A}(C) < 0.9$), 3 people (7.3%) had latticed representations, and 26 people (63.4%) had non-canonical representations ($\mathcal{D}(C^H) < 1$ and $\mathcal{D}(C^V) < 1$). For non-canonical representations, we apply functional analysis and compare $\mathcal{D}(C^H)$ with $\mathcal{I}(C, \Phi^H)$ and $\mathcal{D}(C^V)$ with $\mathcal{I}(C, \Phi^V)$. On average, the probability increased from distinctiveness to informativeness by 9.0% for horizontal and vertical comparisons, meaning that functional analysis offers a better explanation for non-canonical representations than morphological analysis. Furthermore, 9 people (22.0%) had a gain of at least 50% for one of the two comparisons, and 4 people (9.8%) had perfectly informative representations ($\mathcal{I}(C, \Phi) = 1$), which morphological analysis could not account for. Between aligned and informative representations, 7 people (17.1%) respected horizontal regionalisation and 6 people (14.6%) respected vertical regionalisation.

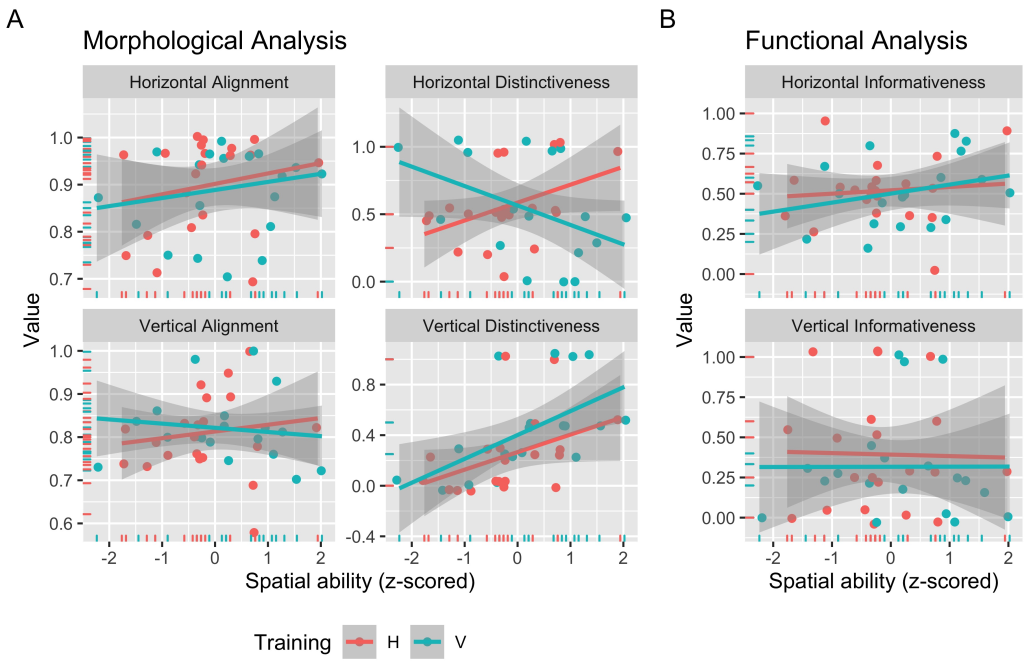

Figure 2 Scatter-plots of the morphological (A) and functional (B; i.e. bottom-up) analysis of the structure mapping task data, as a function of spatial ability (normalised SBSOD score). We observe that morphological analysis is sensitive to individual differences with regards to perception/externalisation of vertical relationships.

Figure 2 shows scatter plots of the 6 metrics (on the y-axis) from both analyses compared against the standardised spatial ability (SBSOD, on the x-axis) of participants. We observe an interaction between training and spatial ability, with respect to horizontal distinctiveness. Participants high in spatial ability trained horizontally (floor-by-floor) produced more horizontally distinctive structure maps; whereas high-spatial participants trained vertically had low horizontal distinctiveness. However, these participants produced higher vertical distinctiveness than high-spatial participants with horizontal training.

5 Discussion and Conclusion

In this article, we sought to establish a mathematical framework to capture spatial relationships from 2D structure maps. More specifically to measure spatial regionalisation of multilevel (i.e. 3D) environments. Our proposed method captures vertical versus horizontal alignment and distinctiveness of groups of objects —here landmarks in a building. We first demonstrated that this method can capture key aspects of structure mapping tasks, such as horizontal and vertical clustering (regionalisation). We then applied this methodology on human structure mapping task data, which helped identify a relationship between spatial abilities and the comprehension of horizontal and vertical spatial relationships.

Various objective methods have been developed to analyse externalisations (products) of mental representations of 2D space, e.g., measuring the angular error in a pointing task, or performing bi-dimensional regression to analyse configurational tasks like sketch maps [7]. Structure mapping tasks can also be analysed in similar terms when they represent two-dimensional configurations, such as objects on a table or locations on a city. To the best of our knowledge, the analysis of structure mapping data from 3D spatial configurations has not been sufficiently studied. As our example illustrates, 3D spatial configurations can be organised in both horizontal and vertical regions (or hierarchies) based on building floors, staircases, and other vertical spatial features. When people produce structure maps of such configurations, they have to perform a series of mental transformations, e.g., adopting an elevation-, top-, or other viewpoint and decoding spatial relationships in 3D space.

To conclude, in the present work we have proposed a novel mathematical framework to capture 3D spatial relationships from 2D structure maps. We applied this methodology on a set of human structure mapping task data, which helped identify a relationship between spatial abilities and the comprehension of vertical spatial relationships. Our results suggest that we can mathematically analyse the spatial configuration of 3D locations, by decomposing the two-dimensional structure maps into top-down as well as bottom-up spatial hierarchies that capture 3D spatial relationships (such as vertical columns and planes). This method provides a complementary approach to previously used classification by human raters [1, 10]. Further work will seek to compare how this method performs in comparison to human ratings. This method can be useful to advance the study of volumetric spatial cognition.

References

1. Simon J. Büchner et al. Path choice heuristics for navigation related to mental representations of a building. In *Proceedings of the Euro CogSci*, 2007.
2. R. Campello et al. Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 2013.
3. Helen Couclelis et al. Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology*, 7, 1987.
4. Yan Feng et al. Wayfinding behaviour in a multi-level building: A comparative study of HMD VR and Desktop VR. *Advanced Engineering Informatics*, 51, 2022.
5. Mary Hegarty et al. Development of a self-report measure of environmental spatial ability. *Intelligence*, 30:425–447, 2002.
6. Christoph Hölscher et al. Up the down staircase: Wayfinding strategies in multi-level buildings. *Journal of Environmental Psychology*, 26(4):284–299, 2006.
7. Rob Kitchin and Marc Blades. *The Cognition of Geographic Space*. I.B. Tauris, 2002.
8. Saskia F. Kuliga et al. Exploring Individual Differences and Building Complexity in Wayfinding: The Case of the Seattle Central Library. *Environment and Behavior*, 2019.
9. Yi Lu and Yu Ye. Can people memorize multilevel building as volumetric map? A study of multilevel atrium building. *Environment and Planning B: Urban Analytics and City Science*, 46(2):225–242, 2019. doi:10.1177/2399808317705659.
10. Panagiotis Mavros et al. Human Navigation in a Multilevel Travelling Salesperson Problem. *PsyArXiv Pre-prints*, pages 1–38, 2022.
11. M. Schwartz. Human centric accessibility graph for environment analysis. *Automation in Construction*, 127:103557, 2021.
12. G.B. Thomas. *Calculus and analytic geometry*. Addison-Wesley Publishing, 1968.
13. J.M. Wiener and Hanspeter A. Mallot. “Fine-to-Coarse” Route Planning and Navigation in Regionalized Environments. *Spatial Cognition and Computation*, 3(4):331–358, 2003.

14 Sudanthi Wijewickrema and Andrew Papliński. Principal component analysis for the approximation of an image as an ellipse. In *WSCG'2005, Plzen, Czech Republi*, 2005.

15 Svante Wold et al. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

16 Andreas Zwergal et al. Anisotropy of Human Horizontal and Vertical Navigation in Real Space: Behavioral and PET Correlates. *Cerebral Cortex*, 26(11):4392–4404, 2016.