Identifying the Perceptual Dimensions of Visual Complexity of Scenes

Aude Oliva (oliva@mit.edu)
Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139

Michael L. Mack (mackmic1@msu.edu)
Department of Computer Science, Michigan State University, East Lansing, MI 48824 USA

Mochan Shrestha
Department of Mathematics, Michigan State University, East Lansing, MI 48824 USA

Angela Peeper
Department of Psychology, Michigan State University, East Lansing, MI 48824 USA

Abstract

Scenes are composed of numerous objects, textures and colors which are arranged in a variety of spatial layouts. This presents the question of how visual complexity is represented by a cognitive system. In this paper, we aim to study the representation of visual complexity for real-world scene images. Is visual complexity a perceptual property simple enough so that it can be compressed along a unique perceptual dimension? Or is visual complexity better represented by a multi-dimensional space? Thirty-four participants performed a hierarchical grouping task in which they divided scenes into successive groups of decreasing complexity, describing the criteria they used at each stage. Half of the participants were told that complexity was related to the structure of the image whereas the instructions in the other half were unspecified. Results are consistent with a multi-dimensional representation of visual complexity (quantity of objects, clutter, openness, symmetry, organization, variety of colors) with task constraints modulating the shape of the complexity space (e.g. the weight of a specific dimension).

Introduction

Real-world scenes are composed of numerous objects, textures and colored regions, which are arranged in a variety of spatial layouts. Although natural images are visually complex, we are able to form a coherent percept amid numerous regions, and identify a complex scene at a glance (Potter, 1976), even in the face of visually degraded conditions (Schyns & Oliva, 1994). This presents the question of how a cognitive system may represent the level of complexity of a scene. Specifically, the following question motivated the experiment presented in this paper: can visual complexity be conceptualized along a single dimension? Or is visual complexity better represented as a multi-dimensional space where the axes might correspond to meaningful perceptual dimensions?

Visual complexity

The perception of visual complexity has been studied with natural texture images (e.g. Heaps & Handel, 1999; Rao & Lohse, 1993) and simple patterns (see Palmer, 1999 for a review). Heaps and Handel had participants rank texture images along several perceptual dimensions including complexity, connectedness, depth, orientation, repetitiveness, and structure. The authors defined complexity as "the degree of difficulty in providing a verbal description of an image". They observed that the complexity of a texture could be estimated along a one dimensional axis representing the degree of perceivable structure: textures with repetitive and uniform oriented patterns were judged less complex than disorganized patterns. This finding correlates with results in the domain of perceptual grouping by acknowledging that the presence of regularities (e.g., symmetry, repetition, similarity) simplifies a visual pattern (Feldman, 1997; Palmer, 1999; Van der Helm, 2000).

How can we represent the complexity of a stimulus like a scene, which has a high variability of parts and spatial layout organization? According to Heylighen (1997), the perception of complexity is correlated with the *variety* in the visual stimulus. Figure 1 illustrates two instances of *variety.* First, the perceived visual complexity can increase as a function of the quantity and range of objects. Second, the perceived visual complexity can increase as a function of the variety of materials and surface styles while the number of objects and surfaces remain constant. The representation of a real-world scene is likely to combine both levels of varieties (parts and surface styles). Intuitively, complex scenes should contain a larger variety of parts and surfaces styles, as well as more relationships between these regions than do simpler scenes.

A visual pattern is also seen complex if its parts are difficult to identify and separate from each other. Yet, paradoxically, when the parts are separated or conceptualized as a whole,

the valence of the complexity changes and the pattern becomes simpler (Heylighen, 1997). This suggests that the perceived complexity of an image also depends on the amount of perceptual grouping, a characteristic independent of the quantity of parts, an observer perceives in the scene. Additionally, the perception of visual complexity is likely to be dependent on the scale of observation (e.g. looking at a bookshelf or the books level), preexisting schemas and familiarity with the scene.

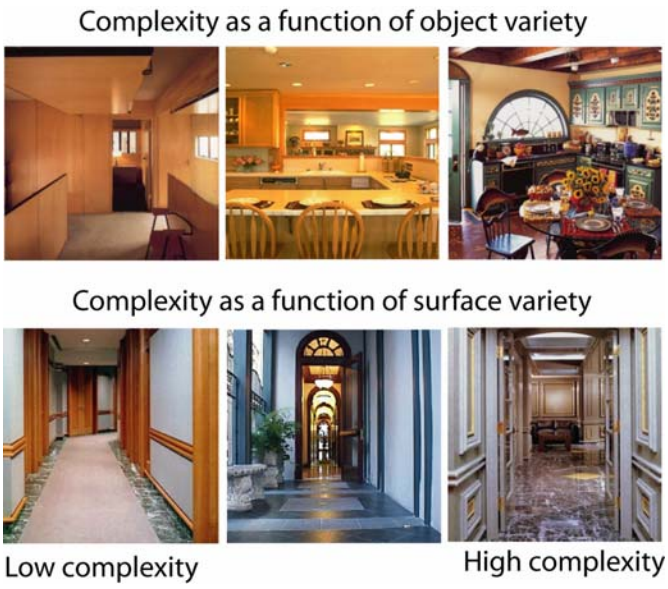

Figure 1: Illustration of how visual complexity evolves as a function of object variety (top) and surface variety (bottom).

If the perception of visual complexity is an interaction between the information in the image and task constraints, can we still identify a set of perceptual properties that participants consistently use to characterize visual complexity of real world scenes? The shape of the visual complexity representation could take three forms:
(1) Unique Perceptual Dimension: the properties of complexity are combined into one principal dimension, robust to subjectivity and task constraints. This is the case of the *naturalness* dimension in real world scenes (e.g. judging if a scene image is a natural or a man-made environment, Oliva & Torralba, 2001).
(2) Multi-dimensional Space Representation: most of visual complexity variability is explained by an identifiable number of perceptual dimensions. The weight of each dimension may vary with task constraints, but the principal dimensional vocabulary remains the same (Gardenfors, 2000). This seems to be the case of the representation of basic-level scene categories (e.g., beach, street, Oliva & Torralba, 2001).
(3) Flexible Space Representation: the properties that human observers use to represent the visual complexity of a particular scene vary with image characteristics (e.g., structure, clusters), tasks constraints, and attentional mechanisms. There is no specific vocabulary that is used for representing visual complexity.
These three levels of representation are not incompatible: for a particular task, the visual complexity space could be skewed towards a line (e.g. one perceptual property is dominant), but for a different task, the space of visual complexity might take into account multiple dimensions. The experiment presented below evaluates the format and content of the representation of visual complexity with the aim to tease apart the three levels of representation suggested above.

Experiment

The goal of the experiment is to study the representation of visual complexity while two groups of participants are told different definitions of visual complexity. Both groups performed a hierarchical grouping task with images of various levels of visual complexity. A hierarchical grouping task allows for identifying the explicit criteria participants used when they organize the pictures (see Oliva & Torralba, 2001) and helps to give a psychological interpretation of the axes provided by a multi-dimensional scaling algorithm (see Results section).

Method

Subjects Thirty-four students from an introduction to psychology course at Michigan State University participated in the study for course credits. Half were in the *control* group and the other half in the *structure* group.

Materials The present study used 100 pictures of indoor scenes. This subset was selected at random from a database of 1000 scenes previously ranked on their subjective visual complexity. The subset had the constraints to represent all levels of complexity along a scale from 1 to 100. The general scene database was originally composed from sources such as the web, magazines and various image databases. Since the volume of the space that a scene image represents is correlated with a given range of clutter (Torralba & Oliva, 2002), only scenes of a small volume range (indoors) were kept for this present study. Moreover, indoor scenes contain a greater variety of colors and objects in a variety of layouts compared to larger scaled environments (e.g. natural space, Oliva & Schyns, 2000).

Procedure The hierarchical grouping task was performed as follows (see Figure 2): starting with 100 pictures shown in a grid on a 23" Apple monitor, participants were asked to separate images into two groups on the screen, corresponding respectively to the most complex vs. the simplest scenes. In a second step, they were asked to split each group into two more subdivisions, and in a third step, split the four groups into two groups each, leading to a total of eight groups. For each subdivision, they were asked to follow a criterion corresponding to visual complexity

(simplicity) and give a verbal description of it. Participants could move each picture across boundaries at any stage, and see an enlarged version of the image by double clicking on it. Similarly to Heaps and Handel (1999), our *Control* group was told the following instruction: "Visual simplicity is related to how easy it will be to remember the image after seeing it for a short time. Visual complexity is related to how difficult it will be to give a verbal description of the image and how difficult it will be to remember the scene after seeing it for a short time." For the *Structure* group, the following instructions were given in addition to the control instructions: "Visual complexity is related to the structure of the scene and therefore, is not merely related to color or brightness. Simplicity is related to how you see that objects and regions are going well together. Complexity is related to how difficult it is to make sense of the structure of the scene". Both groups were forbidden to use a criterion related to the semantic class of the scene (e.g. kitchen) or the presence of a specific object or color.

Results

Table 1 summarizes a taxonomy corresponding to the most common criteria from the descriptions given by participants at the primary and secondary divisions. Each verbal description was recoded as a class of concepts. Some descriptions were a composition of concepts (e.g. pictures on the left seemed more *cluttered* whereas the ones on the right seemed more *open in space*), others were unique (e.g. *quantity* of objects). The percentage in Table 1 should be seen as an indicator of the strength of a perceptual property (most of the time used, often used or almost never used) and not as a fixed value, as variability among individual descriptions was high.

Table 1: Criteria of visual complexity used for the primary and secondary divisions and their % for both groups.

Criteria	Group:Structure	Group:control
Quantity of:		
object	*19*	*32*
detail	*8*	*8*
color	*2*	*19*
Quantity total	29	**59**
Clutter	**18**	5
Symmetry	**15**	2.5
Open Space	**18**	10
Organization	**13**	7
Contrast	<1	8

For the control group, where complexity was defined as a difficulty of verbal and visual recording, the criteria corresponding to *variety* and *quantity* of objects and color dominated the representation of complexity. In the second group where complexity was defined as relating to the structure of the scene, participants evenly used a set of criteria that the control group mentioned less frequently. The primary criterion of the structure group still concerns the quantity and variety of parts, participants referring either to the quantity of objects per se (19%), or the relationship between quantity of objects and spatial arrangement (18%, *clutter*). The other criteria were mostly concerned with spatial layout (symmetry, open space and organization {e.g. grid, centralized, cluster}).

For each condition, we investigated the consistency of the complexity ratings for the 100 images across subjects by computing a Spearman's rank-order correlation for each possible pairing of subjects (images within each subgroup were given the same complexity value, from 1 to 8). If participants were consistent, correlations among participants' rankings should be high. In both groups, Spearman's correlations were all statistically significant ($p < .01$) and were moderate to large in magnitude. Mean correlations of all the pair-wise comparisons were the same in the control and structure group, respectively, $r = 0.62$ and $r=0.61$; (stdev = 0.15 and 0.14).

Next, we applied a nonlinear dimensional reduction method (*Isomap*, Tenenbaum, de Silva, & Langford, 2000) onto a dissimilarity matrix constructed from participants' grouping for each condition (control and structure). To do so, a symmetric 100 x 100 matrix was constructed for each participant. Pairs of images placed in the same group versus in a different group were given respectively a score of 0 or a score of 1. Dissimilarity matrices from all participants from each condition were summed to create two pooled dissimilarity matrices. The Isomap analysis uses the dissimilarities of judgments given by human observers and provides a low dimensional visual representation of the mapping of proximities (i.e., distances) existing between images of various levels of complexity.

Figure 3 shows a two dimensional projection of the 100 images given by Isomap for the *Structure* group. The representation corresponds to the number of independent ways in which visual scenes can be perceived to resemble or differ in visual complexity. Although the dimensions per se are difficult to interpret and further experiments will be needed to assess more accurately the underlying dimensions of the space shown in Figure 3, it shows indeed a first principal direction corresponding to increasing "clutter" and quantity of objects. The second axis, illustrated in Figure 4, suggests an ordering along mirror symmetry and layout organization.

Albeit the correlation between the two first axes given by the Isomap representation for the structure and control group is nearly identical (0.98), the correlation between the ranks of images along the two second axes drops to 0.33 (see Figure 4), suggesting that participants used a different combination of criteria beside quantity while ranking the visual complexity of scenes. In the control group, participants were told that complexity was related to the difficulty of verbally describing an image. Consequently,

they estimated complexity almost exclusively based on the quantity and variety of objects and colors. In the structure group, participants were sensitive to spatial layout criteria, such as symmetry and open space.

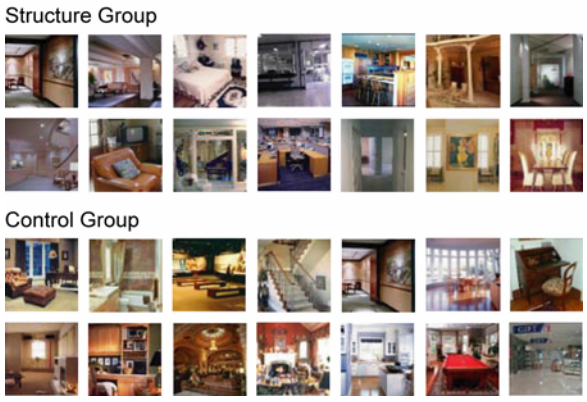

Figure 4: Sample of images projected onto the second principal dimension of *Isomap* for the structure group (top) and the control group (bottom). For the structure group, the images are organized from the top-left to the bottom-right following a property that resembles mirror symmetry. For the control group, the images are organized following a different combination of properties. These projections illustrate the differences in the criteria used between the two groups.

Conclusion

The goal of this study was to characterize the representation of visual complexity and its modulation by task constraints. The complexity ratings provided by observers on 100 pictures of (indoor) real-world scenes are consistent with a multi-dimensional representation of visual complexity.
Furthermore, the high correlations across participants for both groups (average of 0.61) suggest that, within each group, participants used a same (or similar) set of holistic perceptual dimensions to represent complexity. While the contribution of the dimensions are modulated by task constraints, visual complexity is principally represented by the perceptual dimensions of quantity of objects, clutter, openness, symmetry, organization, and variety of colors.

The dimensions of visual complexity listed in Table 1 are not exhaustive: one can imagine that the perceived complexity of scenes of a larger volume of space (e.g., urban environments) might require new dimensions better suited to representing these spaces (e.g., perspective). However, the fact that there exists a set of defined properties that most people are sensitive to is appealing for modeling the visual complexity, where each dimension would be represented as a combination of low-level (e.g. contours, junctions) and medium-level features (e.g. connectedness, symmetry, Mack & Oliva, 2004). Furthermore, finding the true meaningful axes in the space generated by a multi-dimensional scaling algorithm, as well as the status of these dimensions (separable, integral, Garner, 1974; Gardenfors, 2000; Maddox, 1992) will be the subject of a follow-up study.

Acknowledgments

This research was partly funded by a graduate research assistantship to M.L.M (NSF-IGERT training grant) and A.O. was partly funded by an NIMH grant (1R03MH068322-01**)**. We used the *Isomap* code in Matlab provided by J.B. Tenenbaum. The authors would like to thank Nancy Carlisle, Monica Castelhano, Zach Hambrick, Antonio Torralba as well as two anonymous reviewers for helpful comments about the paper. Correspondence can be addressed to A.O. (oliva@mit.edu), or M.L.M (mackmic1@msu.edu).

References

Gardenfors, P. (2000). *Conceptual spaces: the geometry of thoughts*. Bradford Books MIT Press.

Garner, W.R. (1974). The processing of information and structure. Potomac, MD: Erlbaum.

Feldman, J. (1997) Regularity-based perceptual grouping. Computational Intelligence, 13(4), 582-623.

Heaps, C., & Handel, C.H. (1999). Similarity and features of natural textures. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 299-320.

Heylighen F. (1997). *The Growth of Structural and Functional Complexity during Evolution*. F. Heylighen & D. Aerts (eds.).

Mack, M.L., & Oliva, A. (2004). The perceptual dimensions of visual simplicity. Presentation at the *4th Annual Meeting of Visual Sciences Society*, Sarasota, Florida.

Maddox, W.T. (1992). Perceptual and decisional separability. In Ashby, G.F., ed. *Multidimensional models of perception and cognition*, 147-180. Hillsdale, NJ: Lawrence Erlbaum.

Oliva, A., & Schyns, P.G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, 34, 72-107.

Oliva, A., & Schyns, P.G. (2000). Colored diagnostic blobs mediate scene recognition. *Cognitive Psychology*, 41,176-210.

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, 42,145-175.

Palmer, S.E., (1999). Vision Science: Photons to Phenomenlogy. MIT Press.

Potter, M.C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 509-522.

Rao, A.R., & Lohse, G.L. (1993). Identifying high-level features of texture perception. *Graphical Models and Image Processing*, 55, 218-233.

Schyns, P.G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial scale dependent scene recognition. *Psychological Science*, 5, 195-200.

Tenenbaum, J.B., de Silva, V., & Langford, J.C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290, 2319-2323.

Torralba, A., & Oliva, A. (2002). Depth Estimation from Image Structure. *IEEE Pattern Analysis and Machine Intelligence*, 24 (9), 1225-1238

van der Helm, P.A. (2000). Simplicity versus likelihood in visual perception: From surprisals to precisals. *Psychological Bulletin*, 126, 770-800.

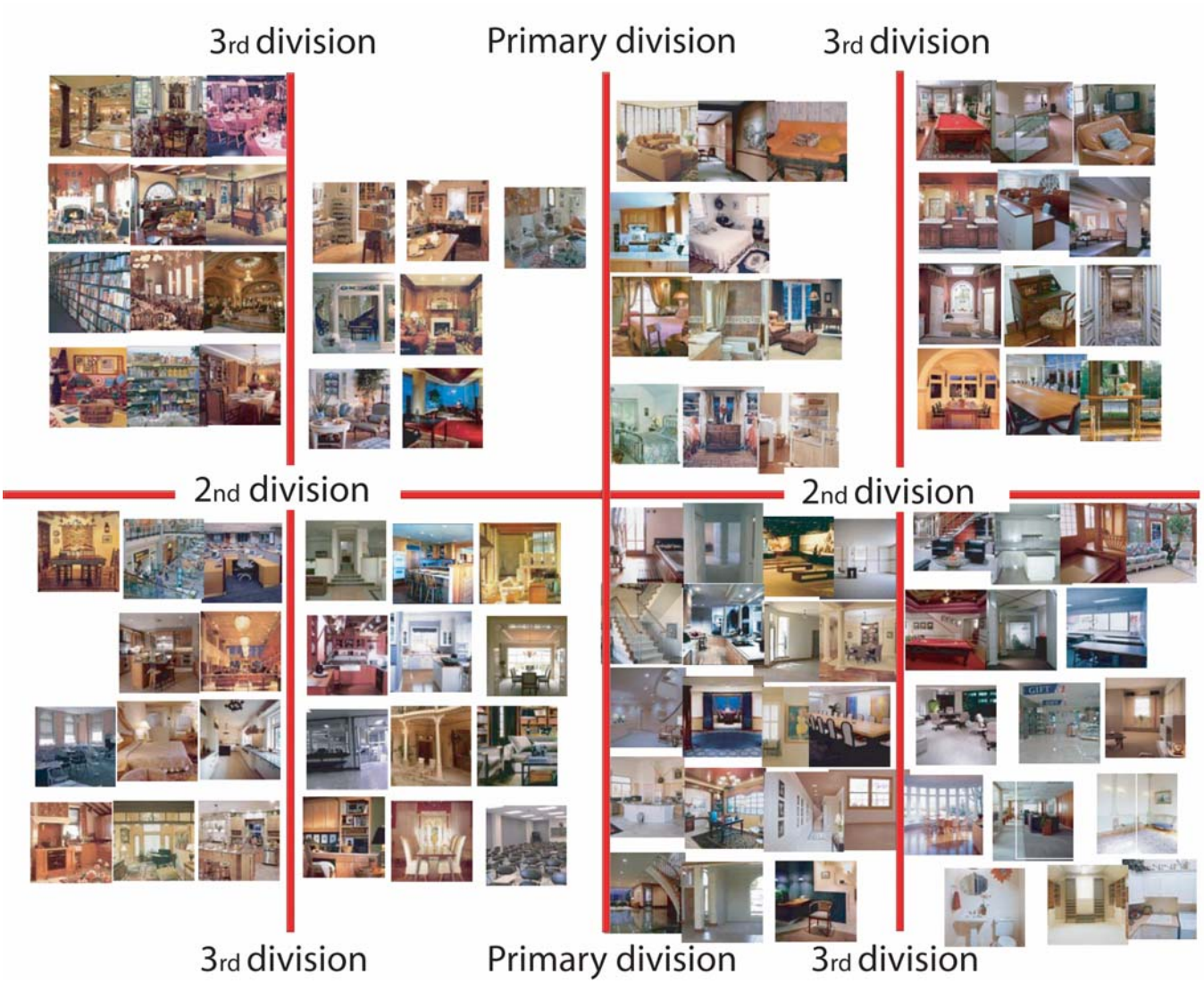

Figure 2: Illustration of the hierarchical grouping task after completion (organization made by subject 1 in the *Structure* group). Most complex scenes are in the top left corner, and most simple scenes are the bottom right corner.

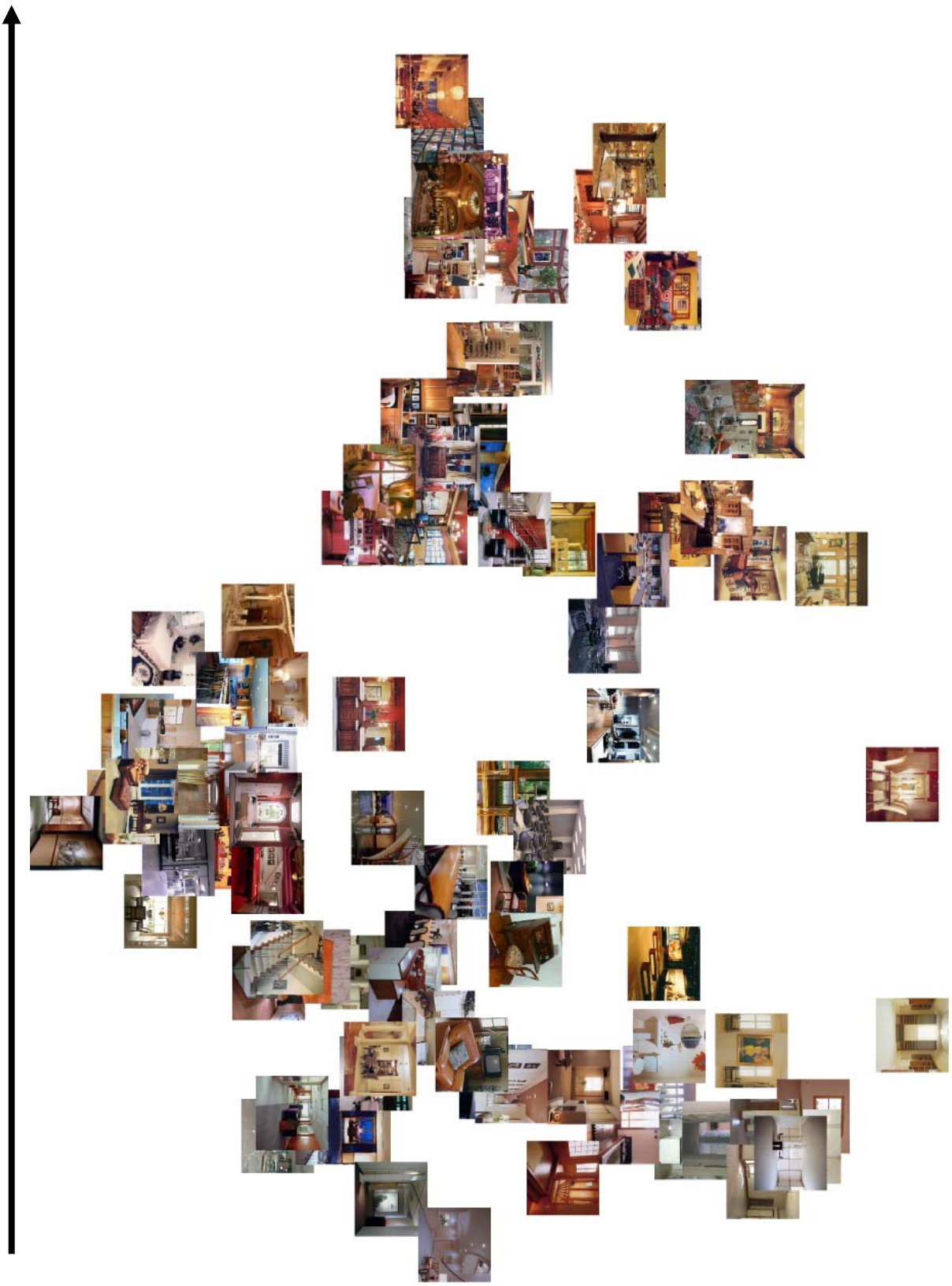

Figure 3: Representation given by Isomap for the structure group. The space shows on the arrow axis, a principal direction corresponding to increasing quantity of objects and clutter. The images that are far away from that direction are images that exhibit the highest amount of variability in how they were grouped in relation to other images. Scenes of medium and low level of clutter exhibit more variations along a second direction, possibly related to symmetry and spatial arrangement (cf. Figure 4).