Object oneness: The essence of the topological approach to perception

Jun Zhang

Department of Psychology, University of Michigan, Ann Arbor, MI, USA

How perceptual organization occurs was the central question of Gestalt psychology, and the concise mathematical articulation of the Gestalt principles remains a century-old challenge. Chen (in the target paper), in summarizing two decades of pioneering research in the topological approach to visual perception, has presented a large body of empirical evidence (using both traditional Gestalt-style experiments and modern neural imaging technology), and persuades us to think again about the psychological and computational processes underlying object perception. Chen's work has been much influenced and inspired by Gibson (1979), who advocated the importance of environment-based visual invariants and direct perception. However, Chen has gone beyond the Gibsonian proposition by bringing in formal mathematical statements to enumerate and express fundamental geometric invariants for early visual perception. The experiments in Chen (1982, 1985) supplied the empirical evidence that the topology of a stimulus configuration plays an important role in visual perception. To formalize his intuition about topological visual perception, Chen (following Zeeman, 1962) defined a tolerance relation on a discrete point set, and suggested the use of tolerance space topology to characterize global topological invariants in a visual configuration. To complete his theory, Chen (1983) proposed an information processing hierarchy for the perception of a visual figure, which involves the extraction of (in progressive order) topological invariants, projective invariants, affine invariants, and finally Euclidean invariants. This sequence coincides with Felix Klein's Erlangen program for the mathematical characterization of geometries as transformation groups acting on a space. Identifying and mapping these geometric invariants onto a perceptual hierarchy, with the perception of topological invariants at the forefront, elegantly fulfils the Gibsonian promise of a geometric theory underlying perceptual organization.

At the core of Chen's thesis is the primacy of figure–ground segregation for visual perception. To transform an image-based representation, which occurs at the retina, to an object-based representation, where figures and their background are all separated, requires the "binding" of points/locations into distinct regions ("chunks") each with a certain topology. The "glue" that enables this binding is what motivates Chen's topological approach. Starting from a point set representing the visual input on the retina, Chen asks how to properly partition the visual space into regions/chunks solely based on large-scale topological properties.

Recall that topology on a point set deals with issues such as continuity, connectedness, neighbourhood, surroundedness, etc., and involves notions like closure, interior, boundary, etc. Insofar as objects occupy space separately from one another and from their surrounds, distinct topological relationships arise whenever there are occlusion relationships among the objects and/or between an object and its background. Take the favourite stimulus of Chen's: an object with a hole. The importance of such a stimulus is that, depending on whether the boundary of the hole belongs (or is perceived as belonging) to the exterior as opposed to the interior part of the "hole", one's percept switches from that of a doughnut (with a hollow centre that unveils its background) to that of a solid disk (in front of a continuous background). This type of ownership of boundary/border is, in set-theoretic language, a question of whether or not a set is defined to include its boundary ∂, i.e. whether the set is closed or open. This is clearly a distinction at the level of topology.

Chen's topological approach advocates a global-to-local order of processing, in the sense that processing of an object's topological property takes precedence over identifying an object's features—that is to say, the establishment of object as a whole, or object "oneness", precedes the identification of specific features belonging to an object. Chen's proposal is provocative, yet carefully reasoned. His information processing hierarchy placed the extraction of topological (and Euclidean) invariants as the first (and last) step. Though it might appear counterintuitive since one would expect the Euclidean properties of a visual image to be recorded right at the outset, Chen argued that Euclidean invariants, such as rigidity (invariant under mental translation and rotation), really are tag-on properties of an already-segregated visual object and are therefore computed after a stimulus is treated as a topological whole. This view challenges traditional computer vision algorithms, where object segregation is based on the identification and binding of features.

Chen's topological proposal has far-reaching consequences for computational algorithms of object oneness, and thus is worth scrutinizing. The following comments will examine Chen's specific suggested use of tolerance space topology on a discrete set, and discuss an alternative approach based on the topology of continuous spaces, i.e., a topological manifold. Moving from a discrete to a continuous setting allows one to conveniently impose differentiability conditions, thereby turning a topological manifold into a differentiable one with a fibre bundle structure. The two central concerns from Chen's topological visual perception, namely the characterization of object oneness and the characterization of shape-changing transformations, will be shown to admit a natural interpretation under the fibre bundle/Riemannian manifold model of visual perception.

Tolerance topology on discrete sets

Chen, following Zeeman's (1962) influential paper, proposed to use a particular type of discrete topology, called the tolerance space topology, to characterize the global topological properties of objects. A tolerance on a point set is a binary relation (i.e., among any two elements/points of the set) that is both reflexive and symmetric, but not necessarily transitive. The absence of a requirement for transitivity makes a tolerance relation different from an equivalence relation. This is an important distinction, because equivalence (reflexive, symmetric, and transitive) relations are the starting point for many common topics of topology, such as the quotient operation; in order to obtain nontrivial global topological properties on discrete sets, one is forced to use this tolerance relation (in lieu of the equivalence relation) to represent perceptual "indistinguishability".

Despite it being a topology on a discrete set, tolerance topology allows the definition of paths, connectedness, holes, and dimensionality. The space of tolerance relations, the tolerance space, is identified as the mathematical characterization of the stimulus configuration. Chen then invokes the algebraic topological notion of homotopy group, first suggested in the context of visual perception by Zeeman (1962) and Zeeman and Buneman (1968), to characterize the tolerance structure among the stimulus points. Specifically, a simplicial complex (i.e., a complex made of simplexes) can be constructed with its vertex points being the points in the original point set. Edges connect pairs of points that are within a given tolerance. For three distinct points, if all pairwise distances are within the tolerance, they form a triangular face; accordingly, they are indistinguishable from one another under this tolerance. The same holds for four, five, ... distinct points, and so on. In this model, distinguishability is characterized by missing edges, faces, etc., in the higher dimensional simplexes that make up the complex. When embedded into the Euclidean space, the dimensionality of the complex increases linearly with the total number of points in the stimulus configuration. Though not a problem in principle, the structure of this simplicial complex, and the resulting homology group H, may become extremely complicated and difficult to compute except for very few dots in the configuration. It remains a challenge to demonstrate that, with the criteria for spatial tolerance becoming either more relaxed or more stringent, a change of H would parallel the change in the resulting percept. In short, Chen's tolerance topology relies on the fundamental assumption of discreteness of visual stimulus configuration as input to vision perception. Even though it may appear as a simplifying assumption, the discrete set approach may turn out to suffer severe computational disadvantages.

Topological manifold and fibre bundle

An alternative to the tolerance topology idea of Chen (borrowed from Zeeman) is to introduce a manifold structure on visual space, so that visual perception takes place on a topological manifold. A (topological) manifold is formed by continuously pasting together pieces of Euclidean space. The only requirement imposed on the point set is Hausdorff separability, namely, for any two distinct points, there exist disjoint open sets that each point is contained by. This continuity property about visual inputs allows one to set up Cartesian coordinate systems (called "charts") at each point on the manifold, so that neighbouring points can be specified using these coordinates. Different charts centred on the same point are related to each other via a coordinate transformation. When a certain collection of charts covers the entire manifold, it is called an "atlas".

Topological manifold captures the basic architecture of information processing by the visual system: Neurons earlier in visual processing stream respond to inputs from restricted regions of the visual space, and that the entire visual space is covered by the overlapping receptive fields (charts) of the neuronal ensemble. In order to compare and contrast features extracted from nearby points, one needs to provide for a proper calculus on the topological manifold. This is achieved by supplying additional (differentiable) structure to make a topological manifold a differentiable one. One may, on a differentiable manifold, perform covariant (intrinsic) comparisons of vectors located at neighbouring points, accomplished through a geometric entity called an affine connection. If the manifold is further endowed with a metric tensor, then it becomes a Riemannian manifold, which admits a unique (called Levi-Civita) connection.

The argument that visual perception involves a stimulus manifold describable in terms of a Riemannian manifold has traditionally appeared in the study of binocular space and depth perception (for example, Indow, 1982, 1991; Luneburg, 1947; Smith, 1959; Yamazaki, 1987). The idea of stimulus comparison in multidimensional perceptual space using covariant differentiation was also explored in Levine (2000). However, none were addressing the issue of object oneness, namely the binding of contiguous locations on the base manifold into a topological whole. In a radical departure from these traditional approaches, Zhang and Wu (1990) used Riemannian geometry to characterize neural processes mediating the segregation of figure–ground relationships and the topological layout of the visual space. Based on identifying the tangent space of the visual manifold as that of motion (directional) selective neuronal responses, and that object oneness is reflected as the intrinsic constancy (through parallel transport) of tangent vectors across neighbouring points, the Levi-Civita connection Γ of the visual manifold is established. Solving for the metric tensor g yields the following expression of the Riemannian metric:

$$\mathbf{g} = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{bmatrix} \begin{bmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{bmatrix}$$

where f denotes the grey-level intensity function (of a two-dimensional image denoted by spatial coordinates x,y) and that the subscripts denote respective

partial derivatives. So any image f induces, through its second derivative (Hessian), a metric tensor and a resulting Levi-Civita connection. It was proven (Zhang & Wu, 1990) that the Riemann-Christoffel curvature R of this connection Γ is identically zero, so parallel translations of a vector are path-independent—this in turn means that segregation of image regions (objects) is possible globally. Specifically, motion-based object segregation, in which image points "glue" together if they are a part of an object undergoing rigid translation (i.e., with spatially uniform image velocity), is represented as a region at which the tangent vectors are intrinsically constant (with vanishing covariant derivative).

Take the example of a random-dot kinematogram (Braddick, 1974), which highlights, on the one hand, the remarkable ease at which object oneness is established by our visual system and, on the other hand, the difficulty with any computation algorithm of object segregation based on feature analysis, as forcefully argued by Chen in the target paper. The image luminance of successive frames allows the motion system (directional sensitive neurons) to extract local features in terms of local movement directions. However, because of the aperture problem, the local directions and the direction of the global target displacement would not agree. Furthermore, because the background dots were also being randomly displaced, motion sensors respond to these regions as well, resulting in a nonuniform response map by motion sensors (Zhang, 1995). This nonuniform response map, or tangent vector field V, is to be compared and construed under the Riemannian metric g. Global topological properties are extracted by covariant differentiation of V (motion response map). The advantage of this Riemannian geometric framework is that the chicken-and-egg problem of whether to compute features or objects first is avoided—objects defined by the constancy of their physical features (e.g., velocity) across space necessarily give rise to an intrinsically constant vector field under an affine connection and, therefore, can be immediately segregated using geodesic coordinates (see Zhang, 1995 for more details). The emergence of a visual figure (target) is the result of simultaneously solving the aperture problem and the location-binding problem.

Though constructed in a continuous (rather than discrete) setting, the differential manifold (fibre bundle) model of visual perception resonates with Chen's basic argument about the primacy of spatial proximity in establishing object oneness—his idea that proximity takes precedence over similarity. This is because, in the language of differential manifold, proximity is simply the (geodesic) distance between points on the base manifold while similarity/dissimilarity is represented by the covariant difference of vectors (i.e., visual features) situated on different base points of the stimulus manifold. The former involves a unique, metric-compatible Levi-Civita connection while the latter may use any affine connection defined on the appropriate fibre. Chen's ideas about proximity taking precedence over similarity precisely expressed the distinction between points on the base manifold (related by proximity) and points in the feature space (related by similarity).

Characterizing topological deformation of an image

One of the questions raised by Chen's topological approach to visual perception is about the characterization of rubber-sheet (plastic) deformations of objects in an image—Chen referred to them as "shape-changing transformations". Correspondence of an object across different images, e.g., in apparent motion, may be established even when the object undergoes considerable deformation. Our visual system's ability to detect and recognize the same object despite a topological deformation (with limited extent) is often called "shape constancy". Though intuitively easy to describe, the precise manner and degree of visual distortion of an object, however, is hard to quantify mathematically based on computation of grey-level image properties alone. Previously, Leyton (1992) systematically investigated the underlying general linear transformation group and demonstrated how a combination of "stretch", "shear", and "rotation" operations (which form appropriate subgroups themselves) on the object's symmetric axis would result in different shapes that nevertheless would be recognized as being produced by the same object. However, none of these operators were generated by specific images themselves, and therefore given an arbitrary image, one does not know which operators to apply and what symmetric axes are appropriate at each image location. One needs a set of image descriptors or curvilinear image coordinates that these shape-changing transformation groups can apply *locally*.

One such descriptor was provided in Zhang (1994). It was based on computing the second derivative (Hessian) of the image function *f*. More precisely, the eigen-vectors of the image Hessian are computed at each image location and, assuming their smoothness, a flow field can be constructed using either of the eigen-directions. These two orthogonal flow fields will fill up a patch of the two-dimensional visual manifold; together they become the curvilinear coordinates that capture local invariant structure of the image function. An image is allowed to deform along either coordinate curves (i.e., the value of a pixel may be dragged by those flows). To avoid the problem of noncommutativity of the two directional vector fields, Zhang (1994) reparameterized the flow fields to make them bona-fide (i.e., mutually compatible) coordinate curves; this was done through forcing their Lie bracket operator to commute, a necessarily condition for orthogonal flow fields to be orthogonal coordinate curves. The infinitesimal transformation of a visual contour, embodied as the Lie derivative dragging the flow field along its path, the so-called "orbit" of a Lie group, quantifies topological transformations such that the invariance ("psychological constancy") of a contour under the transformation group is reflected as its being annulled by the action of the Lie derivative (Hoffman, 1966). The only freedom remaining, the so-called "gauge freedom", is with respect to the scaling of these image-dependent coordinates; this flexibility is important because we want the amount of deformation to have some arbitrary scales. Examples of selecting (i.e., fixing) a particular gauge for "good" or Gestalt images were presented in that paper—it turns out that the original Cartesian space where the image function is defined and the curvilinear coordinates where deformation is quantified are related through a conformal transformation. This computational theory and the associated algorithm for characterizing shape-changing transformation closely follow the spirit of the Lie Transformation Group (LTG) approach proposed by Hoffman (1966, 1968, 1970, 1989, 1994). While it is in no sense complete, hopefully it is a first step towards finding a representation of rubber-sheet deformation (of an image) that is parameterized by the image itself.

Conclusion

To summarize, Chen's research in topological visual perception forces the computational vision community to rethink the difficult problem of object oneness. As Chen cited "Everything is difficult at its very beginning"; it is particularly true if this beginning involves specifying a proper topology for visual perception. Whether to use the tolerance topology on discrete sets or topological (and differentiable) manifold of fibre bundles, future research will clarify the most suitable topological framework to precisely capture the notion of object oneness.

REFERENCES

Braddick, O. (1974). A short-range process in apparent motion. *Vision Research, 14*, 519–527.

Chen, L. (1982). Topological structure in visual perception. *Science, 218*, 699–700.

Chen, L. (1983). What are the units of figure perceptual representation. *Studies of Cognitive Science, No. 22*. School of Social Science, University of California, Irvine).

Chen, L. (1985). Topological structure in the perception of apparent motion. *Perception, 14*, 197–208.

Chen, L. (this issue). The topological approach to perceptual organization. *Visual Cognition, 12*, 553–637.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.

Hoffman, W. C. (1966). The Lie algebra of visual perception. *Journal of Mathematical Psychology, 3*, 65–98. Errata *4*, 348–349.

Hoffman, W. C. (1968). The neuron as a Lie group germ and a Lie product. *Quarterly of Applied Mathematics, 25*, 423–441.

Hoffman, W. C. (1970). Higher visual perception as prolongations of the basic Lie transformation group. *Mathematical Biosciences, 6*, 437–471.

Hoffman, W. C. (1989). The visual cortex is a contact bundle. *Applied Mathematics and Computation, 32*, 137–167.

Hoffman, W. C. (1994). Conformal structures in perceptual psychology. *Spatial Vision, 8*, 19–31.

Indow, T. (1982). An approach to geometry of visual space with no a priori mapping functions: Multidimensional mapping according to Riemannian metrics. *Journal of Mathematical Psychology*, *26*, 204–236.
Indow, T. (1991). A critical review of Luneburg's model with regard to global structure of visual space. *Psychological Review*, *98*, 430–453.
Levine, D. N. (2000). A differential geometric description of the relationships among perceptions. *Journal of Mathematical Psychology*, *44*, 241–284.
Leyton, M. (1992). *Symmetry, causality and mind*. Cambridge, MA: MIT Press.
Luneburg, R. K. (1947). *Mathematical analysis of binocular vision*. Princeton, NJ: Princeton University Press.
Smith, A. A. (1959). The geometry of visual space. *Psychological Review*, *66*, 334–337.
Yamazaki, T. (1987). Non-Riemannian approach to geometry of visual space: An application of affinely connected geometry to visual alleys and horopter. *Journal of Mathematical Psychology*, *31*, 270–298.
Zeeman, E. C. (1962). The topology of the brain and visual perception. In M. K. Ford (Ed.), *Topology of 3-manifolds and related topics* (pp. 240–256). Englewood Cliff, NJ: Prentice-Hall.
Zeeman, E. C., & Buneman, O. P. (1968). Tolerant spaces and the brain. In C. H. Waddington (Ed.), *Theoretical Biology: Vol. 1. Prolegomena* (pp. 140–151). Birmingham, AL: Aldine Publishing Co.
Zhang, J. (1994). Image representation using affine covariant coordinates. In Y.-L. O, A. Toet, D. Foster, H. J. A. M. Heijmans, & P. Meer (Eds.), *Shape in picture: Mathematical description of shape in grey-level images* (NATO ASI Series F: Computer and Systems Sciences, Vol. 126, pp. 353–362). Berlin: Springer-Verlag.
Zhang, J. (1995). Motion detectors and motion segregation. *Spatial Vision*, *9*, 261–273.
Zhang, J., & Wu, S. (1990). Structure of visual perception. *Proceedings of National Academy of Sciences USA*, *87*, 7819–7823.

Please address all correspondence to: Jun Zhang, Department of Psychology, University of Michigan, Ann Arbor, MI 48109, USA. Email: junz@umich.edu