

COMPUTATIONAL PRIMITIVES OF VISUAL PERCEPTION

Yongzhen Huang, Kaiqi Huang, Tieniu Tan

National Laboratory of Pattern Recognition, Institute of Automation
Chinese Academy of Sciences, Beijing, P.R.China, 100190

ABSTRACT

Great stride has been made in psychological research about primitives of visual perception, which is important to computer vision and image processing. In this paper, we propose a computational model to imitate the primitives of visual perception based on the psychological theory of topological perceptual organization. First, we adopt geodesic distance based descriptor to describe an independent topological structure. Then, we consider the spatial relationship of two independent structures. Experiments on structures classification demonstrates that the propose model is consistent with the psychological theory. Further experiments on patches clustering prove that our approach can be used to enhance other algorithms.

Index Terms— visual perception, computational model, descriptor.

1. INTRODUCTION

What are the primitives of visual perception? It has been debated for decades in the field of visual perception. Basically, there are two schools of thoughts: the early feature-analysis theory and the early holistic registration theory.

One representative of the early feature-analysis theory is Treisman's feature integration theory [7], an significant psychological model dominant in visual attention. According to this theory, in the first step of visual processing, visual features are processed and represented with separate "feature maps", which are later integrated in a "saliency map" in order to direct attention to the most conspicuous areas. Another representative is Marr's primal sketch theory [5], which claims that the primitives of visual information representation are simple components of forms and their local geometric properties, e.g., line segments with slopes. Marr's theory has acted as the foundation of most computer vision algorithms for the past 30 years.

This work is supported by National Basic Research Program of China (Grant No. 2004CB318100), National Natural Science Foundation of China (Grant No. 60736018, 60723005), NLPR 2008NLPRZY-2, National Science Founding (60605014, 60875021), National Hi-Tech Research and Development Program of China (2009AA01Z318).



Fig. 1. An example contradicting with the early feature-analysis theory.

The early holistic registration theory, however, considers the visual perception as a global-to-local process. It is supported by the Gestalt psychology of perceptual organization which argues that the whole is more than the sum of its parts, and the operational principle of the brain is holistic, parallel, and analog, with self-organizing tendencies. Fig. 1 shows a classic example where a dalmatian dog sniffing the ground in the shade of overhanging trees. The dog can not be recognized by first identifying its parts (feet, ears, nose, tail, etc.), and then inferring the dog from those component parts. Instead, the dog is perceived as a whole, all at once.

Lin Chen's theory of topological perceptual organization [1, 2] is a view inherited from the early holistic registration theory and Gestalt psychology. It assumes that wholes are coded prior to their separable properties or parts. Chen's theory is based on the core idea that perceptual organization should be understood in the perspective of transformation and perception of invariance over transformation. Furthermore, the topological transformation [2] is proved to be an optimal choice in his theory which is defined as a continuous and one-to-one transformation to change the shape of an object without changing the split or adjacent relation of any pairs of points, e.g., a disc smoothly changing to a solid ellipse.

Chen's theory about visual perception is supported by Gestalt-style experiments as shown in Fig. 2. In these experiments, subjects are requested to judge whether two images are different after a short glimpse. The high correct response means that two shapes are easily distinguished, i.e., topologically different. The experimental results are consistent

with their prediction that it is hard for subjects to differentiate topologically equivalent structures. Therefore it is concluded that topological structures are the fundamental components of the visual vocabulary. Further, Chen’s theory is supported by modern neural imaging evidences [9] and physiological experiments on bees’ visual perception [3].

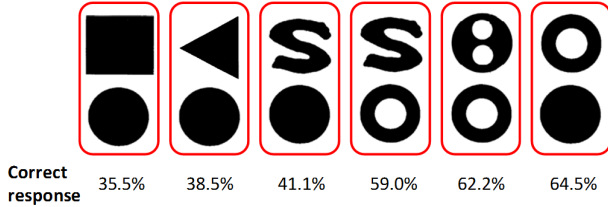


Fig. 2. Illustration of the psychological experiments in Lin Chen’s work. Subjects are requested to differentiate the pair of images in each column. See text for details.

In this paper, we propose a computational model to emulate Chen’s theory of topological perceptual organization. Experiments on structures classification demonstrate that the propose method outperforms classic traditional descriptors. Further experiments on patches clustering prove that our approach can be used to enhance other algorithms.

2. COMPUTATIONAL MODEL

Chen’s theory has opened new lines of research that is worthy of attention from not only visual perception but other related fields, e.g., computer vision and image processing. But it does not define the mathematical form to describe the topological properties. The most important contribution in this paper is the proposed computational model and its successful applications in computer vision.

According to Chen’s theoretical analysis [2], the topological properties include three important aspects: the connectivity, the holes and the spatial relationship between independent topological structures. To build the computational model, we firstly define a topology space. Afterward, we propose an effective method to quantitatively depict the spatial relationship between two independent topological structures.

In practical image processing, we utilize the distance between pairs of pixels to describe the topological structure of a part of an object. The Euclidean measure is apparently not a good candidate because a group of pixels can construct different shapes which depends on the connectivity relationship among them. Intuitively, the geodesic distance, or the shortest path, is a better choice to represent the connectivity and the holes.

Besides, to integrate scale information, we introduce the concept of tolerance space [8, 2]. In detail, points in a tolerance need not be restricted to fixed dots but movable in the

tolerance to generate perceptual organization. That is, two points are connective only if they are in a specific tolerance.

Based on the above analysis, we define the distance between two pixels as:

$$d(i, j) = g(d^*(i, j)), \quad (1)$$

$$d^*(i, j) = \begin{cases} d^o(i, j), & \text{if } d^o(i, j) < \xi \\ \infty, & \text{otherwise} \end{cases} \quad (2)$$

where $g(\cdot)$ denotes the operation of calculating geodesic distance, $d(i, j)$ is the topological distance between pixels i and j , $d^o(i, j)$ is the spatial Euclidean distance, ξ is the tolerance.

After defining the topology space, we construct a histogram to describe the topological structure of a part in an object. In detail, we utilize the quotient between $d(i, j)$ and $d^*(i, j)$ as the vote to construct the histogram, which is the feature vector of an independent topological structure.

The above defined histogram can describe an independent topological structure. But they can not reflect the relationship of two independent topological structures. As shown in Fig. 3, images in (a) and (b) are topologically different, showing the inside and outside relationship, respectively. But their quotient distance histograms are similar. Therefore we also consider the spatial relationship between independent topological structures.

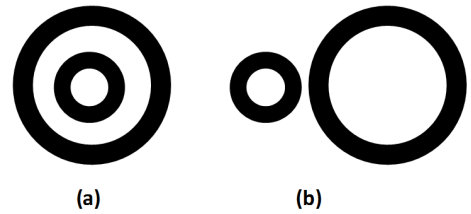


Fig. 3. An example of the inside and the outside relationship.

According to Chen’s theory [2], although the transformations of the shape occur in the production of illusory conjunctions, two independent topological constraints (the inside/outside relationship), remain invariant. We consider that the degree of containment is the key factors to decide the relationship of independent topological structures. The procedure is shown in Fig. 4. Firstly, we draw a group of lines starting from the center of the gravity of one structure, at an equal interval of a predefined angle. Then we count how many lines are across with the other structure. The counted number of crossing lines reflects the degree of a structure containing the other one. Finally, we augment the quotient distance histogram by the number of crossing lines. The augmented histogram is the final representation of the proposed computational model.

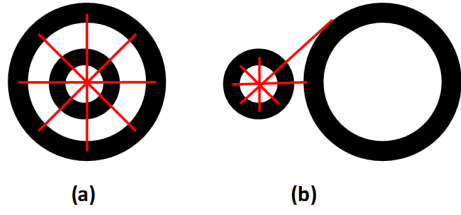


Fig. 4. Illustration of computing the degree of one structure containing the other structure. See the text for details.

3. EMPIRICAL STUDIES

To justify the effectiveness of the proposed computational model, we design two experiments. In the first experiment, we select six groups of graphs, and use the augmented histogram as the feature vector for classification. The benchmark is the SIFT descriptor [4]. In the other experiment, we apply the proposed model for patches clustering.

3.1. Structure Classification

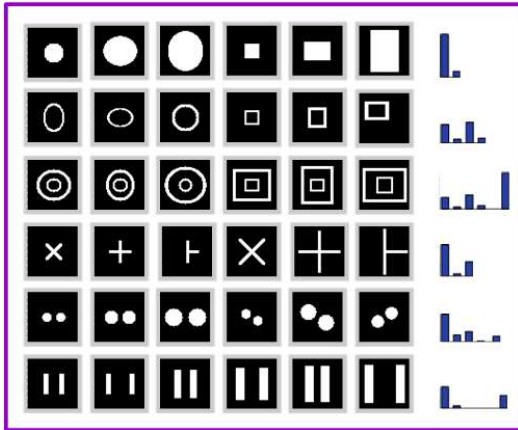


Fig. 5. Examples of artificial images. The histogram in each row is the mean augmented histogram corresponding to the images of the row.

In this experiment, we first test the discriminative ability of the augmented histogram on different topological structures. Fig. 5 shows some examples. We design six classes of topological structures. From top to bottom in Fig. 5, they are the round, the single ring, the double rings, the cross, the double holes and the parallel. There are about 30 images for each class. Note that the round images and the rectangle images share an identical topological structure, thus they are both called the round for convenience. Artificial shapes in the same row share an identical topological structure and their mean augmented histograms are almost equivalent. We compare the augmented histogram with the SIFT descriptor

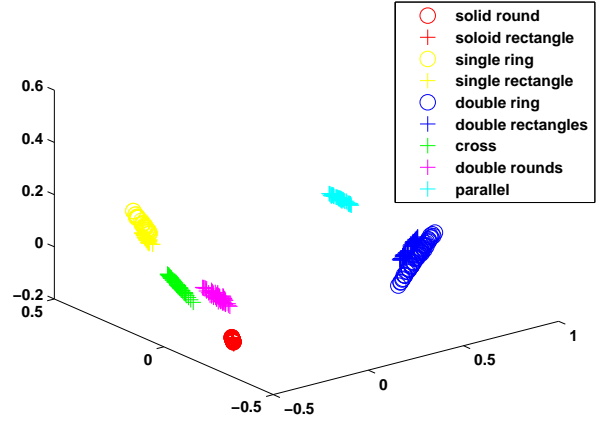


Fig. 6. (Please view in color) Performance of our method (augmented quotient distance) in preserving the geometric nature of the topological structure manifold. Each color correspond to a class of topological structures shown in Fig. 5.

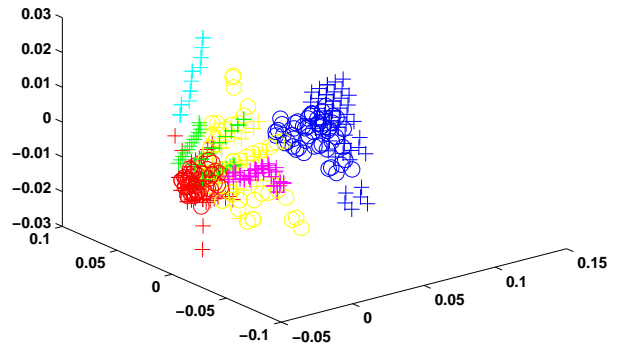


Fig. 7. (Please view in color) Performance of SIFT descriptor in preserving the geometric nature of the topological structure manifold. Each color correspond to a class of topological structures shown in Fig. 5.

[4]. To implement SIFT descriptor, we deem each image as a patch and use 8 directions for histogram calculation. Then the standard Isomap [6] algorithm is applied for dimensionality reduction to intuitively compare their ability of preserving topological properties as shown in Fig. 6 and Fig. 7.

Our approach greatly outperforms SIFT descriptor. In Fig. 6, different topological structures are effectively differentiated by our approach. In Fig. 7, some different topological structures of artificial images are mixed together by SIFT descriptor. The experimental results are consistent with Chen’s theoretical analysis and experimental results, and prove the effectiveness of the proposed method in classifying various

topological structures.

3.2. Clustering for Image Patches

Currently, the proposed model can be applied to describe the topological structure of artificial graphes. It is also possible to use the proposed computational model for real images.

Fig. 8 illustrates the patch selection and clustering procedure. There are 36 original image patches. In the first stage, we cluster all the original image patches. Patches in the first column are preserved cluster centers. In the second stage, only the above three rows will be kept because patches in the last three rows have no structure information. That is, these patches are out of our definition of basic topological shapes. The result of this procedure shows that meaningful patches are selected and redundant patches are removed.

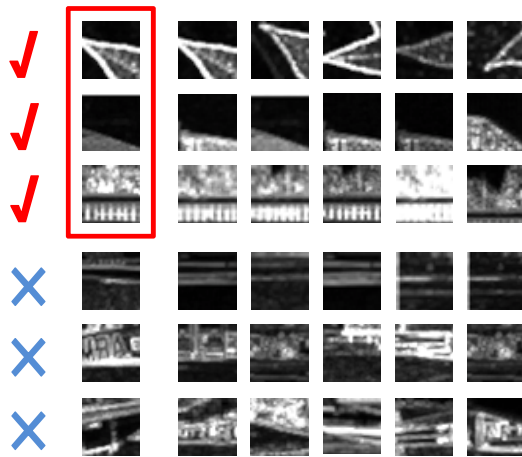


Fig. 8. Demonstration of patch clustering and selection procedure. The three patches in the red rectangle is the finally preserved cluster centers. See the text for detail.

4. CONCLUSION

In this paper, we have first briefly outlined the psychological research about primitives of visual perception and introduced Lin Chen's theory of topological perceptual organization. Based on this theory, we have proposed a computational model to describe the topological structure of a part of an object. In particular, we consider the connectivity, the holes and the spatial relationship between independent topological structures. Experiments on structures classification and iamge patches clustering demonstrates that the proposed model is consistent with Chen's psychological theory and can be applied to cluster and select meaningful image patches.

5. REFERENCES

- [1] L. Chen. Topological structure in visual perception. *Science*, 218:699–700, 1982.
- [2] L. Chen. The topological approach to perceptual organization. *Visual Cognition*, 12(4):553–638, 2005.
- [3] L. Chen, S. W. Zhang, and M. Srinivasan. Global perception in small brain : Topological pattern recognition in honeybees. *Proceedings of the National Academy of Science*, 100:6884–6889, 2003.
- [4] D. G. Lowe. Distinctive image features from dcale-invariant key-points. *International Journal of Computer Vision*, 2(60):91–110, 2004.
- [5] D. Marr. Representing visual information: A computational approach. *Lectures on Mathematics in the Life Science*, 10:61-80, 1978.
- [6] J. B. Tenenbaum, V. Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality. *Science*, 290(22):2319–2323, 2000.
- [7] A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980.
- [8] E. C. Zeeman. The topology of the brain and visual perception. *Topology of 3-manifolds and related topics*, 3:240–256, 1962.
- [9] Y. Zhuo, T. G. Zhou, H. Y. Rao, J. J. Wang, M. Meng, M. Chen, C. Zhou, and L. Chen. Contributions of the visual ventral pathway to long-range apparent motion. *Science*, 299:417–420, 2003.