Computer Vision and Image Understanding
Vol. 76, No. 1, October, pp. 1–5, 1999
Article ID IV990797, available online at http://www.idealibrary.com on IDEAL®

GUEST EDITORS' INTRODUCTION

Perceptual Organization in Computer Vision: Status, Challenges, and Potential

1. HISTORICAL SUMMARY

In a classic experiment, Smith [1] presented a perceptually random stimulus to subjects who were asked to reproduce it. The reproduced set was then used as a stimulus for a second set of subjects who were also asked to reproduce what they were shown. After about 12 cascades the final reproduction had a definite structure to it. The subjects had (gradually) imposed organization on an entirely random stimulus! The experiment indicates the importance of organization in human perception. The human visual system values organization to such a degree that it attempts to find (or impose) it even when none actually exists. Clearly this heuristic behavior has evolved in a world which, generally, is very much structured. By perceptual organization we refer to this ability of a vision system to organize detected features, or primitives, in images based on for instance, Gestaltic criteria. To put it another way, perceptual organization can be defined as the ability to impose structural regularity on sensory data, so as to group sensory primitives arising from a common underlying cause. This sort of organization then permits the formation of object hypotheses with minimal domain knowledge and, therefore, minimal restrictions.

The importance of finding organization in sensory data has long been recognized by researchers in human vision, especially the Gestalt psychologists. However, until relatively recently, the roles of structure and organization have been minimal in computer vision systems. Nevertheless, perceptual organization has been identified as one of the insufficiently emphasized areas in computer vision, lying as it does in the "middle ground" between low-level and high-level processing.

Early work in perceptual organization in computer vision dates back to Marr [2] (curvilinear groupings), Witkin and Tenenbaum [3], and Lowe [4]. It was then that the well-known principle of nonaccidentalness, also known as the principle of common cause or the coincidence explanation, was postulated for perceptual organization. This principle states that it is highly unlikely for organized arrangements of image features to arise by chance and hence, the occurrence of such is significant. Since the work of Lowe, who showed that even simple organizations, such as parallel lines and rectangles, can drastically prune the recognition search tree, there have been a number of contributions that demonstrate the importance of perceptual organization for various vision tasks, e.g., object recognition [5, 6], stereo [7], motion [8, 9], image databases [10], building detection [11–13], and change detection [14]. In addition, there has been an increase in the study of perceptual organization as a task in itself, similar to studies in shape from X. Indeed, it can be argued that a reasonable computational model of perception can be built around the notion of repeated detection and classification of organized structure. Again, however, despite these observations, the full potential of perceptual organization in computer vision is far from being realized.

2. STATE OF THE COMMUNITY

In computer vision, the term "perceptual organization" has been used by various researchers in various contexts, at different levels of vision processing, and with respect to different feature types. This practice has blurred the meaning of the term "perceptual organization." To restore focus to this domain, in [15] we proposed a classificatory structure and a nomenclature, based on the sensor signal dimensionality, level of abstraction, and module inputs and outputs. That is, perceptual groupings differ from one another with respect to the types of constituent features being organized and the dimensions over which the organizations are sought [3, p. 521]. We used these two factors as two axes in our classificatory structure, depicted in Table 1. One axis represents the dimensions over which organization is sought: 2D, 3D (or $2\frac{1}{2}$D), 2D plus motion, and 3D ($2\frac{1}{2}$D) plus motion. The other axis denotes the feature types to be organized, stratified by layers of abstraction: signal level, primitive level, structural level, and assembly level.

The signal level pertains to organizing the raw signal, for example, gray-level images in 2D, range images in $2\frac{1}{2}$D, motion sequences in 2D plus motion, and range sequences in 3D plus motion. The next two levels (primitive and structural) are based on the "dimensionality" of the feature with respect to the domain of organization. The criterion of dimensionality, although not strictly defined here mathematically, refers to the number of parameters that are needed to define a feature. For example, in a 2D static image a contour segment is a one-dimensional manifold while a ribbon is two-dimensional.

The primitive level deals with organizing features extracted from the signal level into lower dimensional manifestations in the organizing field. For example, constant curvature segments

1077-3142/99 $30.00

TABLE 1
Classificatory Structure for Perceptual Organization

	2D	3D($2\frac{1}{2}$D)	2D + time	3D + time
Assembly level	Structures found below Large, regular arrangements [5, 16–18]	Structures found below Large, regular arrangements	Structures found below Coherent motion grouping, articulated motion grouping [19]	Structures found below Coherent motion grouping, non-rigid motion grouping
Structural level	Edge & region primitives Ribbons, corners, merges, polygons, closed regions [5, 7, 11, 20–24, 28–37]	Co-parametric surfaces, boundaries Parallel, continuous patches, tetrahedral vertex combinations [25]	Flow types & boundaries Flow type grouping, correlated motion grouping [26, 27]	Flow streams & boundaries Groups of flow streams, correlated motion grouping
Primitive level	Regions, edge chains Surface faces, contour segments [20, 38, 39]	Surface patches & clusters Co-parametric surfaces, occlusion detection [25, 40]	Optic flow patches Swirls, vortices, sinks, sources [41–43]	3D flow patches Vortices, swirls, sinks, sources
Signal level	Dots, pixels Dot clusters, edge chains, regions, texture patches [38, 44–50, 52–56]	3D points, range Surface patches, discontinuities, point clusters Range segmentation work	Moving points/pixels Coherent pixel, motion groups [50, 51]	Moving 3D points Coherent motion groups

Note. Each cell has three rows. The first row lists some of the typical features to be organized at this level and dimension set. The second row lists some typical output organizations from modules at this level and dimension. The third row lists some of the representative work in this area.

and region boundaries built from edge maps are 1D manifolds embedded in 2D; surfaces are 2D manifolds in 3D. Hence, constant curvature segments and surfaces constitute primitive-level organizations in their respective domains.

At the structural level the organized features have the same dimensionality as that of the space in which they are being organized. Ribbons and closed regions are 2D manifestations in 2D and, therefore, represent structural-level features for 2D organization.

The assembly level is concerned with further organizing the structural-level features. Organizations such as parallel sets of ribbons or boxes constitute the assembly level for 2D grouping.

We use this classificatory matrix (Table 1) to group the past work in perceptual organization. The information in each cell of the matrix shown in Table 1 is arranged as follows. The first row lists some of the typical features to be organized at this level and sensor signal dimensionality. The second row lists some typical output organizations from modules at this level and dimension. The third row lists some of the representative work in this area. None of these lists are exhaustive; this is just a sampling to convey a statistical impression.

Based on the review of the recent work listed in Table 1 and the pre-1993 work reviewed in [15], we offer the following observations.

1. The work in the past six years (1993–1999) continues to follow the pattern established prior to 1993. Most of the work in perceptual organization has been at 2D at the signal, primitive, and structural levels, with increased emphasis on the structural level. Perceptual organization work in 2D and motion has not yet seriously ventured past the signal and primitive levels.

2. Most work in perceptual organization for 2D images has concentrated on extracting continuous contours by grouping image pixels, using primarily the properties of proximity and continuity [30, 38, 47, 52, 53, 55]. Of the work in perceptual organization with extended primitives, such as lines or arcs, the effort has been mostly to form simple, small groups of primitives such as parallels [22], convex outlines [31], ellipses [23, 28], and rectangles [7, 23, 24]. We believe this is partly due to the rarity of fast computational frameworks in which to form large assembly-level groups.

3. GENESIS OF THIS SPECIAL ISSUE

On June 26, 1998, the First International Workshop on Perceptual Organization in Computer Vision was held in Santa Barbara, California, in conjunction with CVPR–98.[1] This meeting was sponsored by the PAMI Technical Committee of the IEEE Computer Society and was a huge success. Approximately 60 attendees heard 21 paper presentations. To encourage intellectual interaction, talks were limited to about 10 min. each, with nearly as much time for questions and discussion following. Never did the discussion exhaust before the time for the next presentation arrived. The closing session was a spirited discussion of more than an hour on the status of perceptual organization as a research topic, its role in vision systems, and where we should go (as a subcommunity) from here.

[1] The proceedings of the workshop are available online at http://marathon.csee.usf.edu/~sarkar/pocv.html.

An outgrowth of the workshop is this special issue. Although the special issue was planned to coordinate with the workshop, submissions were invited from anyone working in the broad area of perceptual organization in computer vision. We received 18 fine papers, of which 7 were selected, following a rigorous peer review process, to be included in this issue.

4. THE WORK IN THIS SPECIAL ISSUE

The papers in this special issue, like most contemporary work in perceptual organization, deal with 2D images. Syeda-Mahmood's texture saliency work can be placed at the 2D signal and primitive levels, because of her use of extended texture primitives. The contributions of Crevier and of Casadei and Mitter are at the 2D primitive level. Lee and Medioni provide a unifying framework (and an unpronounceable title) that cuts across the 2D signal, primitive, and structural levels. Saund's work operates primarily at the 2D structural level. Selinger and Nelson, as well as Amir and Lindenbaum, offer mechanisms that enable large 2D assembly-level organizations; the first using explicit object models while the second does not.

Each paper in this special issue offers novel ideas, some of which should find application in other contexts. Taken together, they offer a good picture of the contemporary work in perceptual organization. Syeda-Mahmood describes a novel texture saliency measure that attempts to duplicate human judgments of visual saliency. This texture saliency measure is then used to identify interesting regions in images.

Crevier provides a probabilistic underpinning to the problem of grouping collinear segments. Previous work in collinear grouping has used pairwise probabilistic characterization, whereas Crevier models the nonaccidental occurrence of N collinear segments.

Amir and Lindenbaum start from the interesting premise that background features are less organized and hence are more easily detected than foreground objects. Their grouping proceeds by sequentially identifying and eliminating the background features.

Lee and Medioni put their voting based algorithm for points, line segments, and junctions into an elegant framework based on saliency tensors. Their results demonstrate that grouping based on even the simple principles of proximity and collinearity can provide substantial leverage.

Selinger and Nelson demonstrate the importance of perceptual organization in object recognition with a working system. A hierarchy of grouping processes that culminates in context-patch groups is used to control the combinatorics of recognition.

Saund offers a mechanism incorporating both local feature associations and global stability information to group contours across occlusions.

Casadei and Mitter employ graph-theoretical techniques to extract closed chains of contour segments. Of particular interest is the novel strategy used to control the combinatorics of the search process.

A few of the recurrent themes emerging from the papers in this special issue, which we believe will prove important in the future, are (i) the statistical (probabilistic) modeling of salient feature associations, (ii) the use of models to drive the grouping (top–down), at least at the assembly level, and (iii) the control of the combinatorics of the grouping process using voting or search–prune strategies.

5. FUTURE DIRECTIONS

It is our conviction that the area of perceptual organization is poised for increased activity and interest as we realize the importance of a robust intermediate-level grouping process in the design of artificial vision systems operating with large numbers of object models. Based on the discussion at the Santa Barbara workshop and our own experience, we believe the following are some of the major fundamental issues in perceptual organization that will be important in the future. We are certainly not sufficiently intrepid to attempt to predict the future; we are just suggesting some promising avenues for future research into this most exciting and intriguing problem of perceptual organization.

1. *The role of learning in perceptual organization*: Is it possible to learn to group features in a domain? Is perceptual organization domain dependent? Can we automatically choose the right features to group for a particular domain?
2. *The use of object models*: The dependence of perceptual organization on object models is an open issue. It is commonly believed that object models should affect the grouping of low-level features, especially to form large groups. However, mechanisms to achieve this without reducing the solution to model matching are rare. Some recent efforts in this direction are those of Nelson and Selinger [5, 37] and Havaldar *et al.* [34]. In these formalisms the object matching problem is intertwined with the higher-level grouping processes.
3. *Perceptual organization in 3D*: The role of perceptual organization in understanding range image data will increase in importance as range acquisition devices become cheaper, portable (or mobile), and readily available. Vision system developers will want to extract meaningful features from range images for recognition, navigation, change detection, and more.
4. *Quality measures*: We must begin to grapple with objective means of evaluation of perceptual organization algorithms. Clearly, the value of a technique can best be assessed in context, perhaps only in the framework of a full system [57]. But, at the very least, we assert that the objective evaluation and comparison of these techniques demands community attention.
5. *Perceptual organization of motion sequences*: Frame-by-frame analysis has been the dominant mode in motion studies. Present day desktop computers are sufficiently fast and have enough memory to undertake the analysis of motion based on spatio-temporal volumes, which generally exhibit more structure and organization than do individual 2D frames.

We hope you enjoy the papers in this special issue and that they inspire, motivate, and facilitate your own efforts in computer vision generally, and in perceptual organization specifically. We also hope that you accept our thoughts, predictions, and recommendations in the spirit in which they are intended—as an exercise in thought provocation. The directions in which we as a research subcommunity actually move will be determined by the statistical aggregation of our individual needs, opportunities, and discoveries yet to be seen—and more than a little by the interests of our collective sponsors. But it is always nice to have some idea of where one is sailing, even if one is not entirely in command of the vessel!

Enjoy and best wishes in your own research endeavors!

REFERENCES

1. B. J. Smith, Perceptual organization in a random stimulus, in *Human and Machine Vision II* (A. Rosenfeld, Ed.), pp. 237–242, Academic Press, San Diego, 1986.
2. D. Marr, *VISION: A Computational Investigation into the Human Representation and Processing of Visual Information*, Freeman, San Francisco, 1981.
3. A. Witkin and J. Tenenbaum, On the role of structure in vision, in *Human and Machine Vision* (J. Beck, B. Hope, and A. Rosenfeld, Eds.), pp. 481–543, Academic Press, San Diego, 1983.
4. D. G. Lowe, *Perceptual Organization and Visual Recognition*, Kluwer Academic, Boston, 1985.
5. R. Nelson and A. Selinger, A cubist approach to object recognition, in *International Conference on Computer Vision*, 1998.
6. B. P. Modayur and L. G. Shapiro, 3D matching using statistically significant groupings, in *Proceedings of the International Conference on Pattern Recognition, 1996*, pp. 238–242.
7. R. Mohan and R. Nevatia, Using perceptual organization to extract 3-D structures, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 1989, 1121–1139.
8. M. Allmen and C. R. Dyer, Computing spatiotemporal relations for dynamic perceptual organizations, *Comput. Vision Graphics Image Process.* **58**, 1993, 338–350.
9. Y. L. Chang and J. K. Aggarwal, Line correspondences from cooperating spatial and temporal grouping processes for a sequence of images, *Comput. Vision Image Understand.* 1997, 186–201.
10. J. Malik, D. A. Forsyth, M. M. Fleck, H. Greenspan, T. Leung, C. Carson, Z. Belongie, and C. Bregler, Finding objects in image databases by grouping in *International Conference on Image Processing, 1996*, pp. 761–764.
11. C. Lin, A. Huertas, and R. Nevatia, Detection of building using perceptual grouping and shadows, in *Proceedings of the Conference on Computer Vision and Pattern Recognition, 1994*, pp. 62–69.
12. S. Krisgnamachari and R. Chellapa, Delineating buildings by grouping lines with mrfs, *Trans. Image Process.* **5**, 1996, 164–168.
13. O. Henricsson, The role of color attributes and similarity grouping in 3D building reconstruction, *Comput. Vision Image Understand.* 1998, 163–184.
14. S. Sarkar and K. L. Boyer, Quantitative measure of change based on feature organization: Eigenvalues and eigenvectors, *Comput. Vision Image Understand.* 1998, 110–136.
15. S. Sarkar and K. L. Boyer, Perceptual organization in computer vision: A review and a proposal for a classificatory structure, *IEEE Trans. Systems, Man, Cybernetics* **23**, 1993, 382–399.
16. K. Rao and R. Nevatia, Space description from imperfect and incomplete data, in *International Conference on Pattern Recognition, 1990*, pp. 414–426.
17. S. Sarkar, Learning to form large groupings of salient image features, in *Proceedings of the Conference on Computer Vision and Pattern Recognition, 1998*, pp. 780–786.
18. P. L. Rosin and G. A. W. West, Extracting surfaces of revolution by perceptual grouping of ellipses, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991*, pp. 677–678.
19. D. D. Hoffman and B. E. Flinchbaugh, The interpretation of biological motion, *Biol. Cybernet.* No. 42, 1982, 195–204.
20. D. G. Lowe, Three-dimensional object recognition from single two—dimensional images, *Artif. Intell.* **31**, 1987, 355–395.
21. J. D. McCafferty, *Human and Machine Vision: Computing Perceptual Organization*, Ellis Horwood, West Sussex, 1990.
22. A. Etemadi, J.-P. Schmidt, G. Matas, J. Illingworth, and J. Kittler, Low-level grouping of straight line segments, in *Proceedings of the British Machine Vision Conference*, pp. 119–126, Springer-Verlag, Berlin/New York, 1991.
23. S. Sarkar and K. L. Boyer, Integration, inference, and management of spatial information using Bayesian networks: Perceptual organization, *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 1993, 256–274. [Special section on probabilistic reasoning]
24. S. Sarkar and K. L. Boyer, A computational structure for preattentive perceptual organization: Graphical enumeration and voting methods, *IEEE Trans. Systems, Man, Cybernet.* **24**, 1994, 246–267.
25. R. B. Fisher, *From Surfaces to Objects: Computer Vision and Three Dimensional Scene Analysis*, Wiley, New York, 1989.
26. M. Shah, K. Rangarajan, and P. S. Tsai, Motion trajectories, *IEEE Trans. Systems, Man, Cybernet.* **23**, 1993, 1138–1149.
27. K. Gould and M. Shah, The trajectory primal sketch: A multi-scale scheme for representing motion characteristics, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1989*, pp. 79–85.
28. G. Roth and M. D. Levine, Geometric primitive extraction using a genetic algorithm, *IEEE Trans. Pattern Anal. Mach. Intell.* **1**, 1994, 901–905.
29. R. L. Castano and S. Hutchinson, A probabilistic approach to perceptual grouping, *Comput. Vision Image Understand.* **64**, 1996, 399–419.
30. L. Williams and A. R. Hanson, Perceptual completion of occluded surfaces, *Comput. Vision Image Understand.* **64**, 1996, 1–20.
31. D. W. Jacobs, Robust and efficient detection of salient convex groups, *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 1996, 23–37.
32. D. Geiger, K. Kumaran, and L. Parida, Visual organization for figure ground separation, in *Proceedings of the Conference on Computer Vision and Pattern Recognition, 1996*, pp. 156–160.
33. J. August, K. Siddiqi, and S. W. Zucker, Fragment grouping via the principle of perceptual occlusion, in *Proceedings of the International Conference on Pattern Recognition, 1996*, Vol. 1, pp. 3–8.
34. P. Havaldar, G. Medioni, and F. Stein, Perceptual grouping for generic recognition, *Int. J. Comput. Vision* **20**, 1996, 59–80.
35. T. F. Syeda-Mahmood, Data- and model-driven selection using parallel line groups, *Comput. Vision Image Understand.* **67**, 1997, 205–222.
36. A. Jascot-Descombes and T. Pun, A synchronous perceptual grouping: From contours to relevant 2D structures, *Comput. Vision Image Understand.* **66**, 1997, 1–24.
37. R. C. Nelson and A. Selinger, A perceptual grouping hierarchy for 3D object recognition and representation, in *DARPA Image Understanding Workshop, 1998*, pp. 157–163.
38. A. Amir and M. Lindenbaum, A generic grouping algorithm and its quantitative analysis, *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 1998, 168–185.
39. P. L. Rosin and G. A. W. West, Segmenting curves into elliptic arcs and straight lines, in *International Conference on Computer Vision, 1990*, pp. 75–78.

40. K. L. Boyer, M. J. Mirza, and G. Ganguly, The robust sequential estimator: A general approach and its application to surface organization in range data, *IEEE Trans. Pattern Anal. Mach. Intell.* **16**, 1994, 987–1001.
41. J. Zhong, T. S. Huang, and R. J. Adrian, Salient structure analysis of fluid flow, in *Proceedings of the Conference on Computer Vision and Pattern Recognition, 1994*, pp. 310–315.
42. A. Verri and T. Poggio, Against quantitative optic flow, in *International Conference on Computer Vision, 1987*, pp. 171–180.
43. C. F. Shu and R. C. Jain, Vector field analysis for oriented patterns, in *Proceedings of the Conference on Computer Vision and Pattern Recognition, 1992*, p. 673.
44. S. W. Zucker, Early orientation selection: Tangent fields and the dimensionality of their support, in *Human and Machine Vision* (A. Rosenfeld, Ed.), pp. 335–364, Academic Press, San Diego, 1986.
45. S. W. Zucker, Computational and psychophysical experiments in grouping: Early orientation selection, in *Human and Machine Vision* (J. Beck, B. Hope, and A. Rosenfeld, Eds.), pp. 545–567, Academic Press, San Diego, 1983.
46. S. W. Zucker, C. David, A. Dobbins, and L. Iverson, The organization of curve detection: Coarse tangent fields and fine spline coverings, in *Proceedings on 2nd International Conference on Computer Vision, 1988*, pp. 568–577.
47. A. Sha'ashua and S. Ullman, Structural saliency: The detection of globally salient structures using a locally connected network, in *International Conference on Computer Vision, 1988*, pp. 321–327.
48. N. Ahuja and M. Tuceryan, Extraction of early perceptual structure in dot patterns: Integrating region, boundary, and component gestalt, *Comput. Vision, Graphics Image Process.* **48**, 1989, 304–356.
49. L. Herault and R. Horaud, Figure ground discrimination: A combinatorial optimization method, *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 1993, 899–914.
50. J. Shi and J. Malik, Normalized cuts and image segmenation, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997*.
51. A. Shio and J. Sklansky, Segmentation of people in motion, in *Proceedings of the IEEE Workshop on Visual Motion, 1991*, pp. 325–332.
52. J. Elder and S. Zucker, Computing contour closure, in *European Conference on Computer Vision, 1996*, pp. 399–412.
53. G. Guy and G. Medioni, Inferring global perceptual contours from local features, *Int. J. Comput. Vision* **20**, 1996, 113–133.
54. H. Ishikawa and D. Geiger, Segmentation by grouping junctions, in *Proceedings of the Conference on Computer Vision and Pattern Recognition, 1998*, 125–131.
55. S. Casadei and S. K. Mitter, Hierarchical image segmentation—detection of regular curves in a vector graph, *Int. J. Comput. Vision* **27**, No. 3, 1998.
56. L. R. Williams and D. W. Jacobs, Stochastic completion fields: A neural model of illusory contour shape and salience, in *International Conference on Computer Vision, 1995*, pp. 408–415.
57. S. Borra and S. Sarkar, A framework for performance characterization of intermediate-level grouping modules, *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 1997, 1306–1312.

Kim L. Boyer
Sudeep Sarkar
Guest Editors