Introduction to the Special Section on Graph Algorithms in Computer Vision

Sven Dickinson, *Member*, *IEEE*,
Marcello Pelillo, *Member*, *IEEE Computer Society*, and Ramin Zabih, *Member*, *IEEE*

1 Introduction

In a letter to C. Huygens of 1679, G.W. Leibniz expressed his dissatisfaction with the standard coordinate geometry treatment of geometric figures and maintained that "we need yet another kind of analysis, geometric or linear, which deals directly with position, as algebra deals with magnitude" [1]. In fact, Leibniz initiated the study of the so-called "geometry of positions" (geometria situs) which, as L. Euler clearly put it in his famous 1736 Königsberg bridges paper which had to mark the beginning of graph theory, "is concerned only with the determination of position, and its properties; it does not involve measurements nor calculations made with them" [2]. After about two centuries, this study developed into two of the richest branches of modern mathematics: graph theory and combinatorial topology.

Mutatis mutandis, an analogous discontent is nowadays being felt among many researchers working in computer vision, a field that is currently dominated by purely geometric methods, who are increasingly making use of sophisticated graph-theoretic concepts, results, and algorithms. Indeed, graphs have long been an important tool in computer vision, especially because of their representational power and flexibility. However, there is now a renewed and growing interest toward explicitly formulating computer vision problems as graph problems. This is particularly advantageous because it allows vision problems to be cast in a pure, abstract setting with solid theoretical underpinnings and also permits access to the full arsenal of graph algorithms developed in computer science and operations research. Graph-theoretic problems which have proven to be relevant to computer vision include maximum flow, minimum spanning tree, maximum clique, shortest path, maximal common subtree/subgraph, etc. In addition, a number of fundamental techniques that were designed in the graph algorithms community have recently been applied to computer vision problems. Examples include spectral methods and fractional rounding.

• S. Dickinson is with the Department of Computer Science, University of Toronto, 6 King's College Rd., Toronto, Ontario, Canada M5S 3G4. E-mail: sven@cs.toronto.edu.
• M. Pelillo is with the Dipartimento di Informatica, Università Cá Foscari di Venezia, Via Torino 155, 30172 Venezia Mestre, Italy. E-mail: pelillo@dsi.unive.it.
• R. Zabih is with the Department of Computer Science, Cornell University, 4130 Upson Hall, Ithaca, NY 14853. E-mail: rdz@cs.cornell.edu.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 114885.

In 1999, we organized *independently* three meetings explicitly devoted to graph algorithms and computer vision. These were the DIMACS Workshop on Graph Theoretic Methods in Computer Vision, held in May at Rutgers University, the IEEE Workshop on Graph Algorithms in Computer Vision (associated with ICCV '99), held in September in Corfu, and the special session on Graph-Theoretic Techniques in Computer Vision at ICIAP '99, the 10th IAPR International Conference on Image Analysis and Processing, held in Venice, also in September. We felt that this was no coincidence and that it was a sign of the growing interest in computer vision around these themes. Therefore, we decided to organize a journal special section devoted to this theme and sent off a proposal to the *IEEE Transactions on Pattern Analysis and Machine Intelligence* editor-in-chief, who accepted it with enthusiasm. Our goal in organizing this special section was to solicit and publish high-quality papers that bring a clear picture of the state of the art in this area. We aimed to appeal to researchers in computer vision who are making nontrivial use of graph algorithms and theory and also to interest theoretical computer scientists in the graph problems that arise in vision.

Late in 1999, we issued a call for papers which resulted in 25 submissions and, after a careful review process, we accepted seven papers for publication, including five regular and two short papers. The papers were reviewed by computer vision researchers employing graph algorithms in their work, as well as graph algorithms researchers from the theoretical computer science community. This was the type of exchange we are trying to promote and we hope to expose others in the graph theory community to the application of graph algorithms to problems in computer vision.

The seven papers in this special issue fall into four categories. The first, graph partitioning, poses the problem of making cuts in a weighted graph according to an appropriate minimum weight criterion. Typical applications in computer vision include image segmentation or perceptual grouping. The second category is graph indexing, which addresses the problem of efficiently selecting a small number of candidate graphs (from a large database) that may account for a query graph. The third category, graph matching, attempts to compute correspondence between two graphs representing underlying image structure. Graph matching is common in problems ranging from object recognition to image registration. The final category, graph generalization, involves computing a prototype graph from a number of exemplar graphs, an important problem in object recognition and object modeling. Below, we briefly summarize the papers appearing in this issue.

2 Regular Papers

Yoram Gdalyahu, Daphna Weinshall, and Michael Werman address, in their paper "Self-Organization in Vision: Stochastic Clustering for Image Segmentation, Perceptual Grouping, and Image Database Organization," a graph partitioning problem that frequently arises in vision where the vertices represent data elements and the edge weights represent similarity. They use a stochastic algorithm to generate a set of $r\text{-way}$ cuts such that lower cost cuts are generated with higher probabilities. These cuts are generated by Karger's contraction algorithm. This effectively creates a new set of edge weights for each value of r, where the new edge weights incorporate nonlocal information, namely, the probability that an $r\text{-way}$ cut they generate will remove this edge. They define a typical $r\text{-way}$ cut as one that removes the edges that have a probability greater than 0.5 and then analyze the set of typical cuts to create a hierarchy of a few selected partitions. Their algorithm gains its robustness primarily from the manner in which the typical cuts are generated, which is stochastic and uses nonlocal information. The method is efficient and can be applied to diverse vision problems, including image segmentation and perceptual grouping.

In their paper "Globally Optimal Regions and Boundaries as Minimum Ratio Weight Cycles," Ian H. Jermyn and Hiroshi Ishikawa propose an energy function for image segmentation that includes information from both the boundaries and the interiors of regions. Their energy function takes the form of a ratio of terms, both of which are defined on the boundary. Information from region interiors is deduced via Green's theorem. They provide two graph algorithms which can efficiently compute the global minimum by using Karp's minimum mean weight cycle algorithm or Lawler and Meggido's minimum ratio weight cycle algorithm. These are two interesting graph algorithms that have not been previously exploited in vision. One of the algorithms proposed in this paper handles a somewhat restricted subclass of energy functions, but is easily parallelizable. The more general, but serial, algorithm is quite fast, typically requiring only a few seconds.

Stefano Berretti, Alberto Del Bimbo, and Enrico Vicario address, in their paper "Efficient Matching and Indexing of Graph Models in Content-Based Retrieval," the problem of content-based image retrieval. Motivated by a desire to include relational information in an image query, they adopt the attributed relation graph as an image query representation. This raises the critical problem of graph indexing, i.e., how to efficiently select a small number of model image graphs that are similar to the query image graph. Citing deficiencies in feature vector-based approaches, the authors propose the use of metric indexing as a means of organizing a large archive of model graphs. Under this scheme, model graphs are hierarchically clustered according to their distance from each other. To compute the distance between two graphs in the presence of distortion, i.e., solving the error-tolerant subgraph isomorphism problem, the authors present a new algorithm combining A^* search with a novel look-ahead estimate. A particularly attractive feature of the algorithm is its ability to accommodate user preferences, e.g., the balancing of feature relevance, during image retrieval. The proposed matching and indexing scheme is demonstrated on a content-based image retrieval application.

Next, in their paper "A Graph-Based Method for Face Identification from a Single 2D Line Drawing," Jianzhuang Liu and Yong Tsui Lee address the problem of line drawing interpretation, a classical computer vision problem with many important applications. The paper provides both a theoretical and a practical contribution. Borrowing from a theory recently introduced by Shpitalni and Lipson, they describe an approach for identifying the faces of a line drawing centered on the idea of finding the maximum weight cliques in a weighted graph. The graph is constructed in such a way that the nodes correspond to the "minimal potential faces" of the drawing, the weight on a node represents the number of edges comprising the face, and the edges express compatibility relations as imposed by a face adjacency theorem. The theoretical contribution of the paper is to show that this new formulation is equivalent to Shpitalni and Lipson's. The main advantage of the proposed formulation is that it allows the authors to develop a fast face identification algorithm; that is their practical contribution. The algorithm makes use of two efficient procedures: one which employs depth-first search to determine the set of minimal potential faces of a drawing and the other which finds all maximum weight cliques of a given graph. Experimentally, it turns out that the proposed algorithm is dramatically faster than Shpitalni and Lipson's method while obtaining precisely the same results.

In their paper, "Structural Graph Matching Using the EM Algorithm and Singular Value Decomposition," Bin Luo and Edwin R. Hancock formulate the inexact graph matching problem within a probabilistic framework. After developing a mixture model to express the probability of a match (or a mismatch) between a node in the data graph and a node in the model graph, they use the well-known expectation-maximization (EM) algorithm to maximize the mixture likelihood. In the expectation step of the algorithm, the a posteriori probability of the neighborhood matches conditioned on the current match is computed, whereas, in the maximization step, the best node assignments are computed by maximizing the expected log-likelihood function. The authors note that the expected log-likelihood function can be recast in a matrix framework and this allows them to realize the update procedure in the maximization step more efficiently using singular value decomposition. Experiments conducted on synthetic as well as real-world data confirm the effectiveness of the approach.

3 Short Papers

Josep Lladós, Enric Martí, and Juan José Villanueva address, in their paper "Symbol Recognition by Error-Tolerant Subgraph Matching between Region Adjacency Graphs," the problem of error-tolerant subgraph matching, assuming a region adjacency graph representation of both model and image. Following a review of both exact and inexact graph matching algorithms used in computer vision, they formulate the problem of inexact subgraph matching as a search for a minimum cost graph edit distance that aligns a distorted image subgraph with a model graph. Specifically, an initial correspondence between an image region and a model region is iteratively grown to accommodate neighboring regions. The cost of adding a neighbor to the correspondence is the cost of the string edit distance aligning the polygonally approximated outer boundaries of the graphs consisting of the matched regions and the neighbor region candidates. The approach is generally applicable to any region adjacency graph representation of an object (model) and has been successfully demonstrated on the domain of symbol recognition in hand-drawn documents.

In the final paper, "On Median Graphs, Properties, Algorithms, and Applications," Xiaoyi Jiang, Andreas Münger, and Horst Bunke consider the problem of extracting a representative model from a given set of graphs and propose extending the median concept to the domain of graphs. Given a set of graphs, the median is defined as the graph having the smallest sum of distances to all graphs in the set (note that this notion differs from the median graph concept used in graph theory). They distinguish between set median and generalized median graphs, the main difference being the set of graphs where the median is searched for. Clearly, both concepts require the notion of a distance between graphs and, in the paper, the authors introduce one based on edit operations. Since the computation of both types of median requires an exponential number of operations, the authors propose a heuristic based on genetic algorithms. The experimental results presented in the paper on both synthetic and real data show the usefulness of the median concept, the advantage of the generalized median over the set median, and the effectiveness of the genetic algorithm in finding good approximate solutions in reasonable time.

4 Conclusions

Graph algorithms have long been an integral part of computer vision research. Their recent resurgence, as witnessed by a flurry in workshop activity, is having an impact on a number of problems in object recognition, indexing, segmentation, and modeling. The application of graph algorithms to computer vision is growing as progress in segmentation and grouping provides more effective image abstractions. These abstractions, naturally represented as graphs, are then indexed and matched to stored, graphical models. This special section provides a sampling of this exciting convergence. We hope it will serve as a catalyst for further work and discussion in this area.

Acknowledgments

The guest editors would like to thank Kevin Bowyer for his advice and support in establishing this special section and Hilda Hosillos from the *TPAMI* editorial office for organizing the review process. They are also grateful to the reviewers for their careful work in evaluating the submissions.

References

[1] N.L. Biggs, E.K. Lloyd, and R.J. Wilson, *Graph Theory: 1736-1936.* Oxford, U.K.: Oxford Univ. Press, 1976.

[2] L. Euler, "Solutio Problematis ad Geometriam Situs Pertinentis," *Commentarii Academiae Scientiarum Imperialis Petropolitanae,* vol. 8, pp. 128-40, 1736 (translated in [1]).

Sven Dickinson received the BASc degree in systems design engineering from the University of Waterloo in 1983 and the MS and PhD degrees in computer science from the University of Maryland in 1988 and 1991, respectively. He is currently an associate professor of computer science at the University of Toronto. From 1995 to 2000, he was an assistant professor of computer science at Rutgers University, where he also held a joint appointment in the Rutgers Center for Cognitive Science (RuCCS) and membership in the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS). From 1994 to 1995, he was a research assistant professor at the Rutgers Center for Cognitive Science and, from 1991 to 1994, a research associate at the Artificial Intelligence Laboratory, University of Toronto. He has held affiliations with the MIT Media Laboratory (visiting scientist, 1992 to 1994), the University of Toronto (visiting assistant professor, 1994 to 1997), and the Computer Vision Laboratory of the Center for Automation Research at the University of Maryland (assistant research scientist, 1993 to 1994, visiting assistant professor, 1994 to 1997). Prior to his academic career, he worked in the computer vision industry, designing image processing systems for Grinnell Systems Inc., San Jose, California, 1983 to 1984, and optical character recognition systems for DEST, Inc., Milpitas, California, 1984 to 1985. His major field of interest is computer vision with an emphasis on shape representation, object recognition, and mobile robot navigation. Dr. Dickinson was cochair of both the 1997 and 1999 IEEE Workshops on Generic Object Recognition, held in San Juan, Puerto Rico, and Corfu, Greece, respectively, while in 1999, he cochaired the DIMACS Workshop on Graph Theoretic Methods in Computer Vision. In 1996, he received the US National Science Foundation CAREER award for his work in generic object recognition, and since 1998 has served as an associate editor of the *IEEE Transactions on Pattern Analysis and Machine Intelligence*. He is a member of the IEEE and the IEEE Computer Society.

Marcello Pelillo received the "Laurea" degree with honors in computer science from the University of Bari, Italy, in 1989. From 1988 to 1989, he was at the IBM Scientific Center in Rome, where he was involved in studies on natural language and speech processing. In 1991, he joined the Department of Computer Science at the University of Bari, Italy as an assistant professor. Since 1995, he has been with the Department of Computer Science at the University of Venice, Italy, where he is currently an associate professor. He held visiting research positions at Yale University, University College London, McGill University, Canada, the University of Vienna, Austria, and the University of York, England. His research interests are in the areas of pattern recognition, computer vision, and neural networks, where he has published more than 70 papers in refereed journals, handbooks, and conference proceedings. He has organized a number of scientific events, including the Neural Information Processing Systems 1999 Workshop on "Complexity and Neural Computation: The Average and the Worst Case" (Breckenridge, Colorado, December 1999). In 1997, he established a new series of international workshops devoted to energy minimization methods in computer vision and pattern recognition (EMMCVPR) and, in 2000, he was a guest coeditor of a special issue of the journal *Pattern Recognition* on this theme. He has been on the program committees of various international conferences and workshops and serves as an associate editor for the journal *Pattern Recognition*. Professor Pelillo is a member of the IEEE Computer Society, the International Association for Pattern Recognition, and the Pattern Recognition Society.

Ramin Zabih attended the Massachusetts Institute of Technology (MIT) as an undergraduate, where he received SB degrees in mathematics and computer science and the MSc degree in electrical engineering and computer science. After earning the PhD degree in computer science from Stanford University in 1994, he joined the faculty at Cornell University, where he is currently an associate professor of computer science. In 2001, he received a joint appointment as an associate professor of radiology at the Cornell Medical School. His research interests lie in early vision and in applications, especially in medicine. He has worked with graph alogrithms since 1997 and is best known for developing fast approximation algorithms for energy minimization that rely on graph cuts. He has consulted extensively with industry, primarily for Microsoft. He has also served on numerous program committees, including the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) in 1997, 2000, and 2001 and the International Conference on Computer Vision (ICCV) in 1999 and 2001. He is a member of the IEEE.