SCIENTIFIC RESEARCH AT THE INSTITUTE OF INFORMATICS OF THE UNIVERSITY OF SZEGED 1998 – 2001

The Institute of Informatics was founded in 1990. It consists of three departments and a research group:

- Department of Foundations of Computer Science
- Department of Computer Science
- Department of Applied Informatics
- Research Group on Artificial Intelligence of the Hungarian Academy of Sciences

The Institute has a library which holds about 5000 Hungarian and English language volumes and subscribes over 200 scientific journals. A scientific journal, Acta Cybernetica is published since 1969 in English by the institute. The journal is available in about 150 university departments worldwide.

The main activity of the institute is the education of modern informatics and computer science knowledge. The institute offers B.S. and M.S. degrees in the following subjects:

- Computer Science (B.S.)
  - It is a 3-years educational program.
  - The aim of this education is to instruct informatics experts who are in possession of up-to-date and high level informatics knowledge and the related foundations of mathematics.
- Computer Science (M.S.)
  - This educational program takes 5 years.
  - The aim of this education is to educate informatics experts who have up-to-date and high level computer science knowledge which are based on deep foundations of mathematics and computer science.
- Computer Science/Economics (M.S.)
  - This program is 5 years long.
  - The aim of this education is to teach informatics experts who are possession of up-to-date and high level informatics and economics knowledge which are based on deep foundations of mathematics.
- Technical Computer Science (M.S.)
  - This program is 5 years long.
  - The aim of this education is to teach informatics experts who possess up-to-date and high level informatics and technology knowledge which are based on deep foundations of mathematics.
- Education in Informatics
  - It is a 5-years program.
  - The aim of this education is to instruct the teachers for informatics who have up-to-date informatics and education knowledge and foundations of mathematics.

A more detailed description of the educational programs can be found at the web site of

`http://www.inf.u-szeged.hu/informatika/oktatase.html`

The curricula of the Departments of Informatics consist mainly of mandatory courses. The curricula have already been adjusted to the standards of leading universities, embracing most of the topics dealing with modern informatics and computer science program, e.g. programming languages, compilers, operating systems, databases, networks, architectures, computer graphics, etc. There are many optional special courses available on the most up-to-date topics in computer science and informatics, e.g. parallel and distributed computing, expert systems, image processing, advanced operating systems, etc. We offer a number of basic mandatory courses, which place the emphasis on the theoretical aspects of computer science.

In addition to these programs, a doctoral program in Computer Science is available since 1993. The aim of this program is to support postgraduate computer science studies at the university, leading to the degree of Ph.D. in computer science, with an emphasis on theoretical aspects. The possible research topics include mostly those parts of computer science and related areas, which are being investigated at the Institute of Informatics. The institute provide also introductory courses in informatics, programming, data processing and networks. These courses are available for all the students of the university (i.e. not only for computer science students).

These programs and research tasks are supported by the scientific and educational cooperation (e.g. CEEPUS, SOCRATES / ERASMUS) with the following higher education institutes:

Boston University, USA,
Columbia University, USA
Dresden University of Technology, Germany,
Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany,
Kyoto Sangyo University, Japan,
Stevens Institute of Technology Hoboken, USA,
Technische Universität Graz, Austria,
Technische Universität Ilmenau, Germany,
Universidad de Almería, Spain,
Universität Bern, Switzerland,
Universität Hamburg, Germany,
Universität Karlsruhe, Germany,
Universität Stuttgart, Germany,
University of Nis, Yugoslavia,
University of Novi Sad, Yugoslavia,
University of Rome, Italy, and
University of Turku, Finland.

Among others, the following research areas are represented at the Departments of Informatics:

- Analysis of algorithms

- Artificial intelligence
- Compositions of automata
- Global optimization
- Image processing
- Operations research
- Software engineering
- Term rewriting systems
- Tree transducers and tree automata

These research fields and the most important recent (1998–2001) results of these topics are described shortly in the following. Further information is available at the web page of the institute: `http://www.inf.u-szeged.hu`.

CURRENT RESEARCH & DEVELOPMENT IN SOFTWARE ENGINEERING

Árpád Beszédes, Rudolf Ferenc, Tibor Gyimóthy, and János Csirik

I. TOPICS AND PEOPLE

The Software Engineering activity of the group involves several research areas as well as related industrial projects that have produced significant results. Most of the work is related to the *Reverse Engineering* of large software systems, particularly C++ programs. The reverse engineering activity deals more with the front-end parts of the process, such as the parser front-end and the higher-level understanding of object oriented systems (recognition of Design Patterns, like that below). Apart from these significant research is also being made in the field of *Program Slicing*, which is described below.

Finally, a new piece of research deals with the exploitation of relationships between XML documents and attribute grammars. This makes it possible to define XML attributes via *semantic rules*, making this novel approach suitable for compressing and understanding XML documents [9].

The software engineering activity involves several people having a variety of interests, including researchers and highly skilled programmers. This has resulted in the industry quality implementations of several research topics. In addition, several PhD and MSc students are also involved in a number of topics.

II. REVERSE ENGINEERING

Most of the reverse engineering (RE) activity revolves around a reverse engineering tool developed under an industrial project with the Nokia Research Center in Finland. During this three-year project, a general RE framework was developed. In this work several research topics were studied that were related to reverse engineering: top-down parser implementation, graph drawing, abstracting higher-level models of programs, recognition of design patterns, etc.

II.1. C++ REVERSE ENGINEERING

The developed RE framework is able to analyze large C/C++ projects and to extract their UML class model, as well as conventional call graphs. The main motivation for developing the system has been to create a general framework for combining a number of RE tasks and to provide a common interface for them. Hence it is a framework tool which supports project handling, data extraction, data representation, data storage, filtering, and visualization. All these basic tasks of the RE process for the specific needs have been accomplished by using the appropriate modules (plug-ins) of the system. Some of these plug-ins are present as basic parts of the tool, but the system can be extended to serve other reverse engineering requirements as well [5].

The parser of the tool is a command-line application that allows its integration into the user's makefiles and other configuration files. One of the greatest assets of the parser is probably the handling of templates and their instantiation at source level.

The C++ language processed by the analyzer meets the ISO/IEC standard from 1998. Furthermore, its grammar has been extended with the extensions used in Microsoft Visual C++ and Borland C++ Builder. The parser itself is a top-down parser generated by a parser generator tool and is based on LL(2) strategy augmented with ambiguity resolution mechanisms. The parser builds an internal representation of the program, which is then filtered and exported to various output formats such as a relational database, HTML report and an exchange schema as described below.

II.2. C++ SCHEMA

One of the most important results of the reverse engineering project is that a *schema* has been elaborated to represent C++ programs. This schema is basically a model that describes how a C++ software system should be modeled using a reverse engineering tool. It also provides a common representation for the implementation, storage and exchange of a reverse-engineered software system. The C++ schema is given using UML class digrams, which has several advantages. Firstly it uses a standard notation, secondly it is very close to implementation level, and thirdly it can be used as a basis for a programming interface for a tool that employs this schema. Externally the model of a software system using this schema can be stored in various physical formats, including XML-based graph representation. The schema has been successfully implemented and is used by other tools as well. In two recent papers [6, 3] we present our schema and provide guidelines for future work.

II.3. RECOGNIZING DESIGN PATTERNS

The model of a C++ program that can be extracted via the reverse engineering tool can be used to recognize higher level relationships from the software (to address the recovery of design decisions). This goal is the aim of work, which tries to recognize standard Design Patterns which exist in the software, but which are not so explicitly expressed. The recognition is done by a separate tool that uses the output of the front-end.

The method consists of two separate phases; analysis and the reverse engineering of the C++ code, and architectural pattern matching over the reverse-engineered intermediate code representation. The pattern recognition effect is successfully realized by integrating two specialized software tools; our RE system and the architectural metrics analyzer *Maisa* (developed at the University of Helsinki). The method with the integrated power of the tool-set and initial results illustrated with small experiments are presented in [4].

III. DYNAMIC SLICING

Program slicing is a powerful means the for debugging, maintenance, reverse engineering and testing of programs. A slice consists of all statements and predicates that might affect a set of variables at a certain program point. A slice may be an executable program or a subset of the program code. More specifically, *dynamic* slicing methods compute those statements which influence the value of a variable occurrence for a specific program input. This means that dynamic slicing methods take advantage of dynamically available information during a program's execution.

An efficient algorithm has been elaborated for dynamic slicing, whose main advantage over prior methods is that it can be applied to real-size C programs because its memory requirements are proportional to the number of different memory locations used by the program (rather than the number of executed steps, which may be unbounded).

The presented algorithm is a forward global method for computing backward dynamic slices of C programs. In parallel to the program execution the algorithm determines the dynamic slices for any program instruction. The basic algorithm first appeared in [8], while in [1] we provide further experimental results and apply the algorithm to real C programs (this paper earned the title of the best paper of the conference). Another publication [2] deals with implementation details regarding the handling of unstructured statements by the algorithm.

IV. R&D Projects

1. NOKIA Researh Center, Helsinki: Code optimization methods, 2000-2002.
2. NOKIA Research Center, Helsinki: C++ Analyzer, 1998-2000.
3. OTKA T23307, New methods and algorithms for program analysis,(subcontractor)1997-2000.

V. CONFERENCES ORGANIZED

The seventh "Symposium on Programming Languages and Software Tools" (SPLST) [7] was organized by our Software Engineering group in Szeged in June 2001.
`http://rgai.inf.u-szeged.hu/~splst`

The scientific part of the next (sixth) "European Conference on Software Maintenance and Reengineering" (CSMR) will be organized by our Software Engineering group in March 2002.
`http://rgai.inf.u-szeged.hu/CSMR2002`

REFERENCES

[1] Árpád Beszédes, Tamás Gergely, Zsolt Mihály Szabó, János Csirik, and Tibor Gyimóthy. Dynamic slicing method for maintenance of large C programs. In *Proceedings of the Fifth European Conference on Software Maintenance and Reengineering (CSMR 2001)*, pages 105–113. IEEE Computer Society, March 2001. Best paper of the conference.

[2] Csaba Faragó and Tamás Gergely. Handling the unstructured statements in the forward dynamic slice algorithm. In *Proceedings of the Seventh Symposium on Programming Languages and Software Tools (SPLST 2001)*, pages 71–83. University of Szeged, June 2001.

[3] Rudolf Ferenc and Árpád Beszédes. Columbus schema for C++. Submitted to the *Sixth European Conference on Software Maintenance and Reengineering (CSMR 2002)*.

[4] Rudolf Ferenc, Juha Gustafsson, László Müller, and Jukka Paakki. Recognizing design patterns in C++ programs with the integration of Columbus and Maisa. In *Proceedings of the Seventh Symposium on Programming Languages and Software Tools (SPLST 2001)*, pages 58–70. University of Szeged, June 2001.

[5] Rudolf Ferenc, Ferenc Magyar, Árpád Beszédes, Ákos Kiss, and Mikko Tarkiainen. Columbus – tool for reverse engineering large object oriented software systems. In *Proceedings of the Seventh Symposium on Programming Languages and Software Tools (SPLST 2001)*, pages 16–27. University of Szeged, June 2001.

[6] Rudolf Ferenc, Susan Elliott Sim, Richard C Holt, Rainer Koschke, and Tibor Gyimóthy. Towards a standard schema for C/C++. Accepted for the *8th Working Conference on Reverse Engineering (WCRE 2001)*. To appear.

[7] Tibor Gyimóthy, editor. *Seventh Symposium on Programming Languages and Software Tools (SPLST 2001)*. University of Szeged, June 2001.

[8] Tibor Gyimóthy, Árpád Beszédes, and István Forgács. An efficient relevant slicing method for debugging. In *Proceedings of the Joint 7th European Software Engineering Conference and 7th ACM SIGSOFT International Symposium on the Foundations of Software Engineering (ESEC/FSE'99)*, number 1687 in Lecture Notes in Computer Science, pages 303–321. Springer-Verlag, September 1999.

[9] Ferenc Havasi. XML semantics extension. In *Proceedings of the Seventh Symposium on Programming Languages and Software Tools (SPLST 2001)*, pages 198–211. University of Szeged, June 2001.

NONLINEAR AND GLOBAL OPTIMIZATION

Tibor Csendes and Péter Gábor Szabó

This summary covers results achieved during 1998-2001 in the fields of developing nonlinear and global optimization procedures and in the investigation of their convergence properties. The majority of the studied algorithms are guaranteed reliability methods. This feature — not to be obtained for procedures based only on real arithmetic operations — was made possible by the use of interval arithmetic.

The basic problem is usually given in the form of

$$\min_{x \in X} f(x), \tag{1}$$

where $f : I\!R^n \to I\!R$ is an (often two times continuously differentiable) real function, and $X \subseteq I\!R^n$ is the set of feasible points. The set X is sometimes a simple n-dimensional interval, while in other cases it is determined by the constraints $g_i(x) \leq 0$ and $h_j(x) = 0$. Global optimization deals with such solutions of problem (1) which are not only locally optimal, but where the related objective function has a globally minimal value.

The above problem class is very general, closely all mathematical problems can be reformulated as such nonlinear optimization problems. This is the reason why we cannot expect to have an algorithm that is capable to solve the whole problem class. Even for simple quadratic problems no polynomial solution algorithm exists. For general nonlinear problems the situation is even worse: such problems are not solvable at the costs of a finite number of function evaluations. The aim of the research is presently to widen the set of solvable problems with new algorithms, and on the other hand, to improve the reliability and the efficiency of the related methods.

The nonlinear and global optimization algorithms have a wide application area, let us just mention the system identification and parameter estimation problems often used in natural sciences. The completed procedures can be downloaded by anonymous ftp from

`ftp.jate.u-szeged.hu`

from the directory of `/pub/math/optimization`. The clustering global optimization code available from here is downloaded 2-3 times daily, and in two independent comparison studies it proved to be the most efficient ("Of the programs tested, the Derivative-Free Boender-Timmer-Rinnooy Kan Algorithm by Tibor Csendes is the clear winner." is written in `http://www.mat.univie.ac.at/~neum/glopt.html`, and see also the paper of Mongeau et al.[1]).

I. INTERVAL METHODS FOR OPTIMIZATION

The largest set of results belongs to interval methods for global optimization. These techniques are based on a branch-and-bound algorithm, and the lower and upper bounds for the subproblems are calculated with either the interval arithmetic, or by more sophisticated inclusion functions based on it.

The present investigations aim to improve the efficiency of the guaranteed reliability algorithms, and in the same time enlarge the set of solvable problems. The first such task was to clear the role of multisection, i.e. the technique when not two (as in bisection), but a few more subintervals are produced in each iteration. The theoretical convergence results and the numerical testing was summarized in [6, 21].

In [16] we have investigated with an extensive a posteriory analysis which one-step-look-ahead decision rule would be the best to find out the multisection grade s. The answer is that the rule must produce the maximal possible difference in the inclusion function values for the resulting subintervals: the upper for one of the intervals must be as close to the lower bound of the other as possible.

Also the earlier mentioned multisection technique was improved by the newly found indicator, the so-called *Rejectindex*

$$pf(f^*, X) = \frac{f^* - \underline{F}(X)}{\overline{F}(X) - \underline{F}(X)}$$

first published by L.G. Casado and coworkers [2]. Its application ways to obtain better efficiency for other parts of the B&B method were discussed in [1, 10]. The latter paper is important, since it is the first successful approach since over 27 years to improve the interval selection step of the algorithm.

András Erik Csallner investigated the storage handling techniques of interval optimization methods in the paper [3], and the consequences of Lipschitz continuity for the stopping rules in [5]. He has successfully defended his PhD thesis [4], and has obtained the PhD degree with "Summa cum laude" in the year 2000.

II. CIRCLES PACKING IN THE UNIT SQUARE

The circles packing in the unit square problem is an old difficult field of computational geometry. It is also highlighted by the fact, that the ten circles case was first solved in 1990, and that no proven solutions exist above 27 circles (with the exception of $n = 36$). Our first results, improving the earlier packings were presented 2000 in Mátraháza, Hungary at the Nonlinear Optimization Winter School, and will appear soon in [2, 25]. Further results are published by Péter Gábor Szabó in [23, 24], and the Japanese versions are covered together with other similar problems in [24]. The paper [20] by Mihály Csaba Markót discusses a verified algorithm for the same problem class. The comment [26] corrects an earlier paper on circles packing.

The improved packing for 47 circles is demonstrated on Figure 1.

[1] Mongeau, M., H. Karsenty, V. Rouzé, J.-B. Hiriart-Urruty: Comparison of public-domain software for black-box global optimization. Optimization Methods & Software 13(2000) 203-226

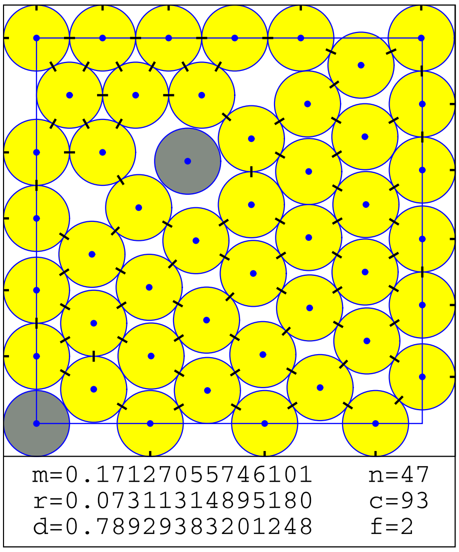


Figure 1. The found improved packing of 47 circles in the unit square. Here r is the radius of the circles, d is the density of the packing, c stands for the number of connections between circles or between a circle and a side, and f denotes the number of free circles.

III. OTHER OPTIMIZATION RELATED RESEARCH

János Balogh studied stochastic global optimization techniques applied successfully to a chemical phase stability problem in [1]. The same author has published a paper on a parallelized version of a random search method for global optimization with coauthors [22]. Boglárka Tóth and Tamás Vinkó tested a sophisticated interval arithmetic based algorithm on standard test problems. The results are published in [27].

The KÉSZ Ltd. asked us to contribute to its production control system with a cutting algorithm that improves the cutting of iron rods used in buildings. The research ended in the suggestion of a compound algorithm containing packing heuristics and enumeration procedures [17, 18] allowing a few percentages of savings in a very high yearly raw material budget.

The possible application of stochastic and interval optimization techniques in chemical process network design problems were discussed in [7].

IV. CONFERENCES ORGANIZED

The authors organized four conferences in the interval 1998 – 2001. The first two are meant as the possibly first conference for many PhD students of computer science, it is the Conference for PhD Students in Computer Science (CS^2) held in Szeged, Hungary, 18-22 July 1998 and 20-23 July, 2000. There were well over 60 accepted talks from many countries, and around 70 participants each time. The related information together with the lists of accepted papers and coordinates of the participants can be found at

`http://www.inf.u-szeged.hu/~cscs`

The SCAN-xx conference series covers scientific computing, validated numerics and computer arithmetic related fields. The SCAN-98 was organized in Budapest, Hungary, 22-25 September, 1998. Since the main research interest of the authors belong to this discipline, it is a big honour that they were asked to organize the next meeting. Over 90 papers were submitted from 20 countries, among them exotic like Jordan, India or Brazil. There were about 70 talks held and over 100 participants. The papers arising from the talks appeared in a special issue of the main journal of the field, Reliable Computing [9], and in an edited volume published by Kluwer Academic Publisher [8]. See also

`http://www.inf.u-szeged.hu/~scan98`

In the case of the fourth conference, Tibor Csendes was a member of the Organizing Committee of the EURO XVII conference, the main meeting of the European researchers in operations research. It was organized in Budapest, Hungary, July 16-19, 2000. The details can be found at

`http://www.sztaki.hu/conferences/euro17/`

Tibor Csendes edited the special issue of the Central European Journal of Operations Research devoted to the papers emerging from the talks of the Hungarian Operations Research Conference held in Veszprém, 1999 [19] with Tamás Rapcsák.

ACKNOWLEDGEMENTS

The research tasks were supported by the following grants: MTA BGA-20/2000, OMFB D-7/97, D-30/00, MÖB-DAAD 11/01, OTKA T 016413, T 017241, F 025743, T 032118, and T 034350; FKFP 0449/99.

REFERENCES

[1] Balogh J., Csendes T., and R.P. Stateva: A fázis-stabilitás elemzése egy új célfüggvény és egy globális optimalizálási módszer fölhasználásával (In Hungarian), Magyar Kémiai Folyóirat 107(2001) 82-89

[2] Casado, L.G., I. García, and T. Csendes: A new multisection technique in interval methods for global optimization. Computing 65(2000) 263-269

[3] Casado, L.G., I. García, and T. Csendes: A heuristic rejection criterion in interval global optimization. BIT 41(2001), 683-692

[4] Casado, L.G., I. García, P.G. Szabó, and T. Csendes: Equal circles packing in a square II. New results for up to 100 circles using the TAMSASS-PECS algorithm. In: F. Giannessi, P. Pardalos, and T. Rapcsák (eds.): Optimization Theory — Recent Developments from Mátraháza. Kluwer, Dordrecht, 207-224, 2001.

[5] Csallner A.E., Improving Storage Handling of Interval Methods for Global Optimization, Acta Cybernetica, Vol. 13, No. 4, pp. 413-422, 1998.

[6] Csallner A.E.: Intervallum-felosztási eljárások a globális optimalizálásban (In Hungarian). PhD dissertation, Szeged, 1999

[7] Csallner A.E.: Lipschitz Continuity and the Termination of Interval Methods for Global Optimization. Computers and Mathematics with Application 42(2001) 1035-1042

[8] Csallner, A.E., T. Csendes, M.C. Markót: Multisection in Interval Methods for Global Optimization I. Theoretical Results, J. Global Optimization 16(2000) 371-392

[9] Csendes, T.: Optimization methods for process network synthesis — a case study. In Christer Carlsson and Inger Eriksson (eds.), Global multiple criteria optimization and information systems quality. Abo Academy, Turku, 1998., 113-132

[10] Csendes, T. (ed.): Developments in Reliable Computing, Kluwer, 1999

[11] Csendes, T. (ed.): Proceedings of the SCAN98, Special Issue of Reliable Computing, 6(3) 1999

[12] Csendes, T.: New subinterval selection criteria for interval global optimization. J. Global Optimization 19(2001) 307-327

[13] Csendes, T.: Interval Analysis: Subdivision Directions in Interval B&B Methods. Encyclopedia of Optimization, Kluwer, Dordrecht, in print, 2001

[14] Csendes, T., R. Klatte, and D. Ratz: A Posteriori Direction Selection Rules for Interval Optimization Methods, CEJOR 8(2000) 225-236

[15] Csendes, T. and M. Markót: On a cutting stock problem of the metal industry (in Hungarian), KÉSZ Ltd., Szeged, 2000.

[16] Csendes T. and M. Markót: On a cutting stock problem of the metal industry II (in Hungarian), KÉSZ Ltd., Szeged, 2000.

[17] Csendes, T. and T. Rapcsák (eds.): Special Issue: Proceedings of the XXIV. Hungarian Operations Research Conference, CEJOR 8(2000) 1-92.

[18] Vladik Kreinovich and Tibor Csendes: Theoretical Justification of a Heuristic Subbox Selection Criterion for Interval Global Optimization. CEJOR 9(2001) 255-265.

[19] Markót, M.C.: An interval method to validate optimal solutions of the "packing circles in a unit square" problems. CEJOR 8(2000) 63-77

[20] Markót, M.C., T. Csendes, and A.E. Csallner: Multisection in Interval Methods for Global Optimization II. Numerical Results, J. Global Optimization 16(2000) 219-228

[21] Ortigosa, P.M., J. Balogh and I. García: A parallelized random search global optimization algorithm, Acta Cybernetica 14(1999) 403-418.

[22] Szabó, P.G.: Some new structures for the "equal circles packing in a square" problem. CEJOR 8(2000) 79-91

[23] Szabó P.G.: Optimális körelhelyezések a négyzetben (In Hungarian), Polygon 10(2000) 48-64

[24] Szabó P.G.: Sangaku, matematikai fatáblák japán templomokban, Középiskolai Matematikai és Fizikai Lapok (In Hungarian), 7:386-388, 2001

[25] Szabó P.G., T. Csendes, L.G. Casado, and I. García: Equal circles packing in a square I. problem setting and bounds for optimal solutions. In: F. Giannessi, P. Pardalos, and T. Rapcsák (eds.): Optimization Theory — Recent Developments from Mátraháza. Kluwer, Dordrecht, 191-206, 2001

[26] Szabó P.G. and T. Csendes: Kiegészítés Filep László tanulmányához. (In Hungarian). Magyar Tudomány, 8:(2001) 984-985

[27] Tóth, B. and T. Vinkó: Egy hatékony számítógépes eszköz matematikai problémák megoldására (In Hungarian), Polygon, accepted for publication, 2001

ANALYSIS OF ALGORITHMS

János Csirik

I. THE CLASSICAL BIN PACKING PROBLEM

In the classical bin packing, we are given a list L of items $(a_1, a_2, ..., a_n)$ each item $a_i \in (0, 1]$ and the goal is to find a packing of these items into a minimum number of unit-capacity bins.

In [5] we present a sequence of new linear-time, bounded-space, on-line bin packing algorithms, the K -Bounded Best Fit algorithms (BBF_K). They are based on the $\Theta(nlogn)$ Best Fit algorithm in much the same way as the Next-K Fit algorithms are based on the $\Theta(nlogn)$ First Fit algorithm. Unlike the Next-K Fit algorithms, whose asymptotic worst-case ratios approach the limiting value of $\frac{17}{10}$ from above as $K \rightarrow \infty$ but never reach it, these new algorithms have worst-case ratio $\frac{17}{10}$ for all $K \geq 2$. They also have substantially better average performance than their bounded-space competition, as we have determined based on extensive experimental results summarized here for instances with item sizes drawn independently and uniformly from intervals of the form $(0, u], 0 < u \leq 1$. Indeed, for each $u < 1$, it appears that there exists a fixed memory bound $K(u)$ such that $BBF_{K(u)}$ obtains significantly better packings on average than does the First Fit algorithm, even though the latter requires unbounded storage and has a significantly greater running time. For $u = 1$, BBF_K can still outperform First Fit (and essentially equal Best Fit) if K is allowed to grow slowly. We provide both theoretical and experimental results concerning the growth rates required.

In [10] we study online bounded space bin packing in the resource augmentation model of competitive analysis. In this model, the online bounded space packing algorithm has to pack a list L of items in (0,1] into a small number of bins of size $b \geq 1$. Its performance is measured by comparing the produced packing against the optimal offline packing of the list L into bins of size 1. We present a complete solution to this problem: For every bin size $b \geq 1$, we design online bounded space bin packing algorithms whose worst case ratio in this model comes arbitrarily close to a certain bound $\rho(b)$. Moreover, we prove that no online bounded space algorithm can perform better than $\rho(b)$ in the worst case.

In [8] and [7] we present experimental and theoretical results of the deterministic on-line Sum-of-Squares algorithm. SS is applicable to any instance of bin packing in which the bin capacity B and the item sizes $s(a)$ are integer (or can be scaled to be so). It performs remarkably well from an average case point of view: for any discrete distribution in which the optimal expected waste is sublinear, SS also has sublinear expected waste. For any discrete distribution where the optimal expected waste is boubded, SS has expected waste at most $O(logn)$. In addition, we present a randomized $O(nBlogB)$ -time on-line algorithm SS^* based on SS, whose expected behavior is essentially optimal for all discrete distributions. Algorithm SS^* also depends on a new linear-programming based pseudopolynomial-time algorithm for solving the NP-hard problem of determining, given a discrete distribution F, just what is the growth rate for the optimal expected waste. An off-line randomized variant SS^{**} performs well in a worst-case sense: For any list L of integer-sized items to be packed into bins of a fixed size B, the expected number of bins used by SS^{**} is at most $OPT(L) + \sqrt{OPT(L)}$.

II. BIN COVERING

Bin covering takes as input a list of item sizes and places them into bins of unit demand so as to maximize the number of bins whose demand is satisfied. This is in a sense a dual problem to the classical one-dimensional bin packing problem, but has for many years lagged behind the latter as far as the quality of the best approximation algorithms. In [6] we design algorithms to close the gap, both in terms of worst-case and average-case results. We present (1) the first approximation scheme for the offline version, (2) algorithms with bounded worst-case behavior whose expected behavior is asymptotically optimal for all discrete "perfect-packing distributions" (ones for which optimal packing has sublinear expected waste), and (3) a learning algorithm that has asymptotically optimal expected behavior for all discrete distributions. The algorithms of (2) and (3) are based on the recently-developed Sum-of-Squares algorithm [7] for bin packing, whose adaptation to the case of bin covering requires new ideas and analytical techniques.

The paper [1] deals with vector covering problems in d-dimensional space. The input to a vector covering problem is consists of a set X of d-dimensional vectors in $[0, 1]^d$. The goal is to partition X into a maximum number of parts, subject to the constraint that in every part of the sum of all vectors is at least one in every coordinate. This problem is known to be NP-complete, and we are mainly interested in its on-line and off-line approximability. For the on-line version, we construct approximation algorithms with worst case guarantee arbitrarily close to $1/(2d)$ in $d \geq 2$ dimensions. We prove that for $d \geq 2$ no on-line algorithm can have worst-case ratio better than 2/5. For the off-line version, we derive polynomial time approximation algorithms with worst case guarantee $\Omega(1/logd)$. For $d = 2$ we present a very fast and very simple off-line approximation algorithm that has a worst case ratio 1/2. Moreover, we show that a method from the area of compact vector summation can be used to construct off-line approximation algorithms with worst case ratio $1/d$ for every $d \geq 2$.

In [3] we define two simple algorithms for the bin covering problem and gave their asymptotic performance.

In[9] an overview of on-line bin packing and covering problems is given.

III. SCHEDULING

In [2] we consider the problem of scheduling a sequence of tasks in a multi-processor system with conflicts. Conflicting processors cannot process tasks at the same time. At certain times new tasks arrive in the system, where each task specifies the amount of work (processing time) added to each processor's workload. Each processor stores this workload in its *input buffer*. Our objective is to schedule task execution, obeying the conflict constraints, and minimizing the maximum buffer size of all processors. In the off-line case, we prove that, unless $P = NP$, the problem does not have a polynomial-time algorithm with a polynomial approximation ratio. In the on-line case, we provide the following results: (i) a competitive algorithm for general graphs, (ii) tight bounds on the competitive ratios for cliques and complete k-partite graphs, and (iii) a $(\Delta/2 + 1)$-competitive algorithm for trees, where Δ is the diameter. We also provide some results for small graphs with up to 4 vertices.

IV. PAGING

In [4] we consider a variant of the online paging problem where the online algorithm may buy additional cache slots at a certain cost. The overall cost incurred equals the total cost for the cache plus the number of page faults. This problem and our results are a generalization of both, the classical paging problem and the ski rental problem. We derive the following three tight results: (1) For the case where the cache cost depends linearly on the cache size, we give a λ-competitive online algorithm where $\lambda \approx 3.14619$ is a solution of $\lambda = 2 + ln\lambda$. This competitive ratio λ is best possible. (2) For the case where the cache cost grows like a polynomial of degree d in the cache size, we give an online algorithm whose competitive ratio behaves like $d/lnd + o(d/lnd)$. No online algorithm can reach a competitive ratio better than d/lnd. (3) We exactly characterize the class of cache cost functions for which there exist online algorithms with finite competitive ratios.

REFERENCES

[1] N. Alon, Y. Azar, J. Csirik, L. Epstein, S. V. Sevastianov, A. P. A. Vestjens, and G. J. Woeginger. On-line and off-line approximation algorithms for vector covering problems. *Algorithmica*, 21:104–118, 1998.

[2] M. Chrobak, J. Csirik, Cs. Imreh, J. Noga, J. Sgall, and G. J. Woeginger. The buffer minimization problem for multiprocessor scheduling with conflicts. In *Automata, Languages and Programming, Lecture Notes in Computer Science, Springer, Vol. 2076*, pages 862–874, 2001.

[3] J. Csirik, J. B. G. Frenk, M. Labbe, and S. Zhang. Two simple algorithms for bin covering. *Acta Cybernetica*, 14:13–25, 1999.

[4] J. Csirik, Cs. Imreh, J. Noga, S. S. Seiden, and G. J. Woeginger. Buying a constant competitive ratio for paging. In *9th Annual European Symposium, Lecture Notes in Computer Science, Springer, Vol. 2161*, pages 98–108, 2001.

[5] J. Csirik and D. S. Johnson. Bounded space on-line bin packing: Best is better than first. *Algorithmica*, 31:115–138, 2001.

[6] J. Csirik, D. S. Johnson, and C. Kenyon. Better approximation algorithms for bin covering. In *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*, pages 557–566. ACM Press, 2001.

[7] J. Csirik, D. S. Johnson, C. Kenyon, J. B. Orlin, P. W. Shor, and R. R. Weber. On the sum-of-squares algorithm for bin packing. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 208–217. ACM Press, 2000.

[8] J. Csirik, D. S. Johnson, C. Kenyon, P. W. Shor, and R. R. Weber. A self organizing bin packing heuristic. In *Algorithm Engineering and Experimentation, ALENEX '99, Lecture Notes in Computer Science, Vol. 1619, Springer,* , pages 246–265, 1999.

[9] J. Csirik and G. Woeginger. On-line bin packing and covering problems. In *In On-Line Algorithms—The State of the Art, A. Fiat and G. Woeginger, Eds., Lecture Notes in Computer Science, Springer, Vol. 1442*, pages 147–177, 1998.

[10] J. Csirik and G. Woeginger. Resource agmentation for online bounded space bin packing. In *Automata, Languages and Programming, Lecture Notes in Computer Science, Vol. 1853, Springer*, pages 296–304, 2000.

FUZZY OPERATORS AND MULTI-CRITERIA DECISIONS

József Dombi

In the last few years both the computers and the softwares have changed a lot. Nowadays there is an information overload. We can say that the amount of information stored in computers doubles in every year. So there is a great demand for programs reducing these huge data masses to usable and understandable quantities, helping in discovering relations not only for experts, but everyday people too.

Until now only artificial intelligence has been dealing with realizing human tasks for computer. In the 90s appeared the so-called computational intelligence, which aimed at mathematical modeling of human operations between perception and logical reasoning. They gave new algorithms and methods on the area of multi-criteria decision making, genetic algorithms and fuzzy sets.

I. MEMBERSHIP FUNCTION

Fuzzy sets realize the formal description of the human non-exact concepts. According to the earlier interpretations the membership function is the modeling of the uncertainty. But the successful practical applications do not use uncertainty. We showed that using elementary preference function (which is the approximation of the characterization function) instead of the membership function can solve this problem.

The proposed function basically represents the intensity of preference (for example smaller, better, more favourable soft estimation). This is a logical function, which has got the following parameters: expectation level, decision threshold and the sharpness of the preference.

We showed, that the membership functions used in the fuzzy theory can always be expressed as composition of two functions. The first is the mapping of the perception to a scale, the second is the earlier mentioned elementary preference function. So, the subjective factor, which is not the subject of mathematical disciplines (sociology, psychology), can be separated from the fuzzy theory, which makes possible to build a formal and theoretically consistent system.

II. CONNECTIVE OPERATORS AND THE NEGATION

Another critical point of the fuzzy theory is the choice of continuous logic operators. The number of the introduced “and” and “or” operators have been growing in the last few years. More and more new operators had been introduced beside the first “min” and “max” operators. General description of these operators is given by the t-norms and t-conorms. T-norms is the family of monotonous operators fulfilling the associativity.

Its main classes are the following: strictly monotonous operators (if there is no idempotent element inside the domain of the operator), nilpotent operators, min-max operators and drastic operators. If the associativity is not assumed, the number of operators is much greater. Their full description is given by solutions of associative functional equation and representations of ordered semigroups.

One aim of the researchers is finding these kinds of operators. Another task is to find reductions of the mentioned family of operators, to find those representations which are the best in the aspect of practical application. Further restriction is given by the consistency of the system. In case of strictly monotonous operators the necessary and sufficient conditions of DeMorgan classes are also the result of the latest research. Applying this theory we can also get the representation of the negations. Until now in the class of strictly monotonous connective operators and the negation were not in any relation. The negation is not unique. Its parameter is the neural value (the fix point of the negation), which can be interpreted as a decision level.

III. UNARY OPERATORS

In the fuzzy theory beside the negation there were also a lack of the mathematical basis of the other unary operators. We still use the modifiers (hedges) defined by Zadeh in a heuristic way. As the applied hedges are independent from the conjunctive and disjunctive operators, the transformation is always the same.

The consequence is that the fuzzy theory became more and more sophisticated. Unary operators could be defined to be consistent to the other operators. The negation – which has connection to the other logic operators by the DeMorgan law – helps us defining unary operators as composition of two negations with different neutral values. These unary operators are unique up to isomorphic. Functions of the isomorphic mappings are the same as the generator functions of the operators that come from the solutions of the associative functional equation.

Using the logic operators we have to deal with decision levels. Using unary operators the different decision levels can be transformed to the "same bases", they could be regarded as idempotent values. In this way an operator can be derived which is a special case of general mean-values when all the weights are the same.

IV. MULTI-CRITERIA DECISIONS AND OPERATORS

At multi-criteria decisions we have to deal with the importance of the certain features through the evaluation. We showed, that the operators can be generalized further keeping their earlier features. The weighting method is an operator dependent transformation and it can be linearized on the weights, which increase the practical applicability of such multi-criteria decisions.

The earlier studied aggregation can be connected with the logical operators, because it comes from the limes of weighted logical operators. A nice feature of these operators is that we can get back the conjunction, disjunction, aggregation, min max operator, if the parameter of the operator is $-\infty, -1, 0, +1, +\infty$, so by the change of just one parameter we can get the whole spectrum of the operators. By the help of aggregation another unary operator can be made, which transform the sharpness. Preference operator can be made from aggregation by the help of negation. Decisions based on preference and utility theory are equivalent if we apply the preference coming from aggregation operator.

V. FLEXIBLE (PLIANT) SYSTEM

There are several approaches to treat databases, like

- classical analysis: probability theory, statistical analysis (too technical, so the everyday people cannot use it)
- new algorithms: neural networks, genetic algorithms (more effective and a bit more user-friendly, but the success depends on the settings of the technical parameters made by the user)

As the classical mathematical logic is different from the everyday used logic, there is a need of introducing a new concept. Pliant (flexible) system can handle these concepts in a unified way, and by applying it reasoning can be made.

Pliant system contains logical elements and decision elements, like aggregation (which is a special summary of data features and it is very important in decision making) and preference modeling. It can integrate the often-used fuzzy systems, neural networks and decision trees, i.e. this logic gives a common base of paradigms developed in the last few years. It can also be applied in clustering methods and essentially improves the technologies based on fuzzy control. A brand new class of learning algorithms can be derived from the system.

We can get pliant system if we use Dombi formula instead of general operator. This operator system is isomorphic with the general one. We can see that in pliant system the preference operator and the interpretation of the membership function by sigmoid function is consistent. The nilpotent operator class can be built up similar like the earlier mentioned monotonous operator class. We can differ two kinds of applications: soft computing methods and decision support methods. Applying this system we can make operations on numbers and the result should be composed in natural language ensuring the easy understanding.

VI. GENETIC ALGORITHMS

Genetic algorithm is a stochastic searching method based on the modeling of evolution. We have introduced concepts which were present in the literature on an intuitive way. Starting from the theoretical questions, introducing the concept of species, using a new method of localization of competition and reproduction we made GAS suitable for discovering the local optimum of functions. The developed program called GAS realized the applicability of the theory.

The aim of the further researches was to choose the suitable diameter of species. In spite of the earlier theories this value was provided by a function while the algorithm was running, so the value was continuously decreasing. The theoretical base of this concept can be considered as an important step on this research area.

REFERENCES

[1] Dombi, J.: Comparability measure, 22nd Linz Seminar on Fuzzy Set Theory: Valued Relations and Capacities in Decision Theory, Linz, 2001. Febr. 6-10.

[2] Dombi, J.: Visual Query System for Datamining, Proc. of the Int. Symposium of Hungarian Researchers Computational Intelligence, Budapest, 2000. Nov. 2., pp. 111-124

[3] Koch, Gy., Dombi, J.: New clustering procedures in respect of practical realization, Conference of PhD students in Computer Science, 2000. July 20-23., pp. 62.

[4] Dombi, J.: Comparability measure, EUROXVII. European Conference on Operational Research, Budapest, 2000. July 16-19., pp. 107

[5] Dombi, J.: Inference models, Mathematical Aspects of Non-classical Logics and Fuzzy Inference, Linz, 2000. June 13-17.

[6] Dombi, J.: DeMorgan class and negations, Proceeedings of EUROFUSE SIC99 The fourth meeting of the EURO Working Group on Fuzzy Sets and The Second International Conference on Soft and Intelligent Computing, Budapest 335-344., 1999.

[7] Jelasity M., Dombi, J.: GAS, a concept on modeling species in genetic algorithms, Artificial Intelligence, 99. 1-19. 1998.

[8] Kocsor, A., Bálint, I., Dombi, J.: How to solve matrix-eigenvalue problems by optimization of various functional inequalities? SCAN 98, IMACS/GAMM International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics, Budapest 1998

[9] Dombi, J.: Cognitive aspects of datamining. DataScope: a visualization tool and visual query system. KESDA98 Knowledge Extraction from Statistical Data 27-28 April 1998 Luxembourg

RESEARCH IN THEORETICAL COMPUTER SCIENCE

Zoltán Ésik

I. ITERATION THEORIES AND THE AXIOMATIZATION OF THE FIXED POINT OPERATION IN COMPUTER SCIENCE

The purpose of the work carried out from 1998 to 2001 was to obtain complete, but simple, descriptions of the properties of the fixed point or iteration operation in computation. Previous work by Bloom, Ésik and others has resulted in a complete description of the equational laws satisfied by the iteration operation. This description takes the form of the axioms for *Iteration Theories* [6, 7], certain (enriched) Lawvere algebraic theories, which capture important features of many classes of structures of interest in the theory of computation involving ordered and metric structures, functors and 2-categories, trees and synchronization trees, and others. In [36], we have proved that functor theories over algebraically complete categories, where iteration is defined by initial algebras, result in iteration theories. The importance of algebraically complete categories for theoretical computer science is that these are exactly those categories in which one can define abstract data types in a canonical way. In [10], we have introduced iteration 2-theories that are a 2-categorical generalization of iteration theories. We have described the structure of the free iteration 2-theories.

One of the axioms of iteration theories, the commutative identity, was quite complex and dificult to verify in particular cases. In [21], we have replaced this axiom by the *group-identities* associated with the finite (simple) groups. The completeness of the group-identities confirms a classic conjecture of Conway in an extremely general setting. Refinements and further analysis of this result were obtained in [20, 23]. In [9], we have shown that iteration theories cannot be axiomatized by a finite number of equation schemes. Moreover, in [24], we have described all iteration theories of boolean functions. In the paper [5], we have described the structure of the free Conway theories.

II. APPLICATIONS OF THE EQUATIONAL LOGIC OF FIXED POINTS

As major beneficial corollaries of the general theory of fixed points, we have obtained relatively simple sets of equational and implicational axioms for the binary supremum operation on continuous functions [25], the concurrent behavior of processes with respect to several behavioral equivalences [25, 29], regular tree languages [17]. For an application to Lindenmayerian power series, see [33]. For a survey, see [28].

III. AXIOMATIZING REGULAR LANGUAGES AND RATIONAL POWER SERIES

In [13], we solved an open problem of A. Salomma by showing that the equational theory of commutative regular languages (commutative Kleene algebra) is the same as the theory of one letter regular languages. Thus, since the theory of commutative regular languages is nonfinitely based, so is the theory of one letter regular languages. In [15], we proved that the equational theory of regular languages with concatenation and Kleene star is nonfinitely based. This solves a problem of Bredikhin. In [34], we extended Kozen's axiomatization of the equational theory of regular languages to rational power series with coefficients in any finite commutative ordered semiring. In [14], we proved that the equational theory of Kleene algebras of binary relations with relational inverse is not finitely based.

IV. TROPICAL SEMIRINGS

In a series of papers, we studied the equational theory of the tropical semirings. We established several axiomatizability and decidability results and described the structure of the free algebras in the corresponding varieties. See [4, 1, 2, 3].

V. THE SHUFFLE OPERATION AND POMSET MODELS OF CONCURRENCY

In a series of papers, we extended the study of the equational theory of the regular operations on languages to include the shuffle operation which, in the interleaving model of concurrency, is used to model parallel execution of processes. We have shown that the equational theory of the shuffle operation is intimately related to the theory of pomsets equipped with variants of the series and parallel composition operations. See [38, 32, 30]. In [39], we solved an open problem of J.-E. Pin.

VI. COMPOSITIONS OF AUTOMATA

In [22], we gave a simple proof of the Krohn-Rhodes decomposition theorem for finite automata. In [16], we have shown that there is a Letichevsky automaton not complete with respect to homomorphic representation by the ν_2-product. (On the other hand, it is known that every Letichevsky automaton is complete with respect to homomorphic representation by the ν_3-product.)

VII. GENERALIZATIONS OF AUTOMATA

In [19], we defined the syntactic theory of a regular tree language as the minimal theory that recognizes the tree language. Using this concept, we gave an extension of Eilenberg's Variety Theorem to regular tree languages. This variety theorem establishes a bijective correspondence between classes of regular tree languages closed with respect to the Boolean operations, inverse morphic images and quotients, and suitably defined varieties of algebraic theories. In [37], we studied automata on higher dimensional words and extended several basic results of the theory of automata to this setting.

VIII. ALGEBRAS FOR ASYNCHRONOUS CIRCUITS

We have proposed a family of algebras to study the hazards of asynchronous circuits, cf. [11, 12]. These algebras are commutative De Morgan bisemigroups and are a generalization ternary algebras. In [27], we gave a geometric description of the free (commutative) De Morgan bisemigroups and De Morgan bisemilattices. In [18], we showed that every ternary algebra can be faithfully represented by ternary algebras of functions.

IX. CONFERENCES AND EDITORIAL WORK

In addition to a large number of other conference presentations, Zoltán Ésik was invited speaker at the following conferences: Words, Languages and Combinatorics (Kyoto, 2000), Category Theory (Como, 2000), Developments in Language Theory (Vienna, 2001), Fixed Points in Computer Science (Florence, 2001), Workshop on Max-Plus Algebras (Prague, 2001), Expressiveness in Concurrency (Aalborg, 2001), 6th Int. Conf. Discrete Mathematics and Applications (Banska, 2001). He served on the steering committee of the Fundamentals of Computer Science conferences and on the program committee of the conferences Universal Machines and Computations (Metz, 1998), Foundations of Software Science and Computation Structures (Lisbon, 1998), Fixed Points in Computer Science (Brno, 1998), Automata and Formal Languages (Vasszéchény, 1999), Fixed Points in Computer Science (Paris, 2000), Fixed Points in Computer Science (Florence, 2001). Zoltán Ésik edited the proceedings of FICS 98 and served as an editor of the journals Theoretical Computer Science, Theoretical Informatics and Applications, Discrete Mathematics and Theoretical Computer Science, Acta Cybernetica, Acta Sci. Math. He held visiting positions in Aizu, Waterloo and Aalborg.

ACKNOWLEDGMENTS

The above research has been supported by various grants from the National Foundation of Hungary for Scientific Research, the Ministry of Education of Hungary, the US-Hungarian Joint Fund, the Japan Society for the Promotion of Science, the Austrian-Hungarian Action Foundation, the French-Hungarian Bilateral Fund BALATON, and the Austrian-Hungarian Research and Development Program.

REFERENCES

[1] L. Aceto, Z. Ésik and A. Ingólfsdóttir, On the two-variable segment of the equational theory of the max-sum algebra of natural numbers, in: proc. *STACS 2000, Lille*, LNCS 1770, Springer-Verlag, 2000, 267-278

[2] L. Aceto, Z. Ésik and A. Ingólfsdóttir, Axiomatizing tropical semirings, in: *Foundations of Software Science and Computation Structures,* Genova, 2001, F. Honsell and M. Miculan editors), Lecture Notes in Computer Science 2030, Springer-Verlag, April 2001, 42-56.

[3] L. Aceto, Z. Ésik and A. Ingólfsdóttir, Non-finitely Based Tropical Semirings. Proceedings of the WORKSHOP ON MAX-PLUS ALGEBRAS and Their Applications to Discrete-event Systems, Theoretical Computer Science, and Optimization, 27-29 August, 2001, Prague, (S. Gaubert and J. J. Loiseau editors), IFAC (International Federation of Automatic Control) Publications, Elsevier Science, 2001, pp. 29-34.

[4] L. Aceto, Z. Ésik and A. Ingólfsdóttir, The max-sum algebra of natural numbers has no finite equational basis, *Theoretical Computer Science,* to appear.

[5] L. Bernátsky and Z. Ésik, Semantics of owchart programs and the free Conway theories, *Theoretical Informatics and Applications,* RAIRO, 32(1998), 35-78.

[6] S.L. Bloom and Z. Ésik, Iteration Theories: The Equational Logic of Iterative Processes, EATCS Monographs on Theoretical Computer Science, Springer-Verlag, 1993.

[7] S.L. Bloom and Z. Ésik, The equational logic of fixed points, *Theoretical Computer Science,* 179(1997), 1-60.

[8] S.L. Bloom and Z. Ésik, Shuffle binoids, *Theoretical Informatics and Applications,* 32(1998).

[9] S.L. Bloom and Z. Ésik, There is no finite axiomatization of iteration theories, in: proc. *LATIN 2000, Punta del Este,* Uruguay, LNCS 1776, Springer-Verlag, 2000, 367-376.

[10] S.L. Bloom, Z. Ésik, A. Labella and E.G. Manes, Iteration 2-theories, *Applied Categorical Structures,* 9(2001), 173-216.

[11] J. Brzozowski and Z. Ésik, Hazard algebras, in: *Half Century of Automata Theory,* World Scientific, to appear in 2001.

[12] J. Brzozowski, Z. Ésik and Y. Iland, Algebras for hazard detection, in: *31st IEEE International Symposium on Multiple-Valued Logic,* 22-24 May 2001, Warsaw, Poland, IEEE Computer Society, Los Alamitos, California, 2001, 3-12.

[13] S. Crvenković, I. Dolinka and Z. Ésik, A note on equations for commutative regular languages, *Inf. Proc. Letters,* 70(1999), 265-267.

[14] S. Crvenković, I. Dolinka and Z. Ésik, The variety of Kleene algebras with conversion is not finitely based, *Theoretical Computer Science,* 230(2000), 235-245.

[15] S. Crvenković, I. Dolinka and Z. Ésik, On equations for union-free regular languages, *Information and Computation,* 164(2001), 152-172.

[16] P. Dömösi and Z. Ésik, A note on completeness of the ν_3-product, *Publ. Math.,* to appear.

[17] Z. Ésik, Axiomatizing the equational theory of regular tree languages, in: proc. *STACS '98, Paris,* LNCS 1373, Springer-Verlag, 1998, 455-465.

[18] Z. Ésik, A Cayley theorem for ternary algebras, *Int. J. Algebra and Computation,* 8(1998), 311-316.

[19] Z. Ésik, A variety theorem for trees and theories, *Publicationes Mathematicae,* 54(1999), 711-762.

[20] Z. Ésik, Axiomatizing iteration categories, *Acta Cybernetica,* 14(1999), 65-82.

[21] Z. Ésik, Group axioms for iteration, Information and Computation, 148(1999), 131-180.

[22] Z. Ésik, A new proof of the Krohn-Rhodes decomposition theorem, *Theoretical Computer Science,* 234(2000), 287-300.

[23] Z. Ésik, The power of the group identities for iteration, *Int. J. Algebra and Computation,* 10(2000), 349-373.

[24] Z. Ésik, Iteration theories of boolean functions, in: proc. *Math. Found. of Computer Science, Bratislava, 2000,* LNCS 1893, Springer-Verlag, 2000, 343-352.

[25] Z. Ésik, Axiomatizing the least fixed point operation and binary supremum, in: proc. *Computer Science Logic, Munich, 2000,* LNCS 1862, Springer-Verlag, 302-316.

[26] Z. Ésik, Free algebras for generalized automata and language theory, in: proc. *Algebraic Systems, Formal Languages and Computation, RIMS Kokyuroku 1166,* Kyoto University, 2000, 52-58.

[27] Z. Ésik, Free De Morgan bisemigroups and bisemilattices. *Algebra Collquium,* to appear.

[28] Z. Ésik, The equational theory of fixed points with applications to generalized language theory, *Developments in Language Thoery,* Wien, July 2001, W. Kuich. ed., 25-44.

[29] Z. Ésik, Continuous additive algebras and injective simulations of synchronization trees, *J. Logic. Comput.,* to appear.

[30] Z. Ésik, Axiomatizing the subsumption and subword preorders on finite and infinite partial words, *Theoretical Computer Science,* to appear.

[31] Z. Ésik and M. Bertol, Nonfinite axiomatizability of the equational theory of shuffle, *Acta Informatica,* 35(1998), 505-539.

[32] Z. Ésik, M. Katsura, M. Ito, The equational theory of reversal, in: *Algebraic Engineering, proc. of the International Workshop on Formal Languages and Computer Systems, Kyoto, March 18 - 21 1997 and the First International Conference on Semigroups and Algebraic Engineering, Aizu, 24 - 28 March 1997,* Word Scientific, 1999, 502-521.

[33] Z. Ésik and W. Kuich, A Kleene theorem for Lindenmayerian algebraic power series, *J. of Automata, Languages and Combinatorics,* 5(2000), 109-122.

[34] Z. Ésik and W. Kuich, A generalization of Kozen's axiomatization of the equational theory of the regular sets, in: *Words, Semigroups, and Transductions,* World Scientific, 2001, 99-114.

[35] Z. Ésik and W. Kuich, Inductive *-semirings, *Theoret. Comput. Sci,* to appear.

[36] Z. Ésik and A. Labella, Equational properties of iteration in algebraically complete categories, *Theoretical Computer Science,* 195(1998), 61-89.

[37] Z. Ésik and Z. Németh, Automata on series-parallel biposets, in: *Developments in Language Thoery,* Wien, July 2001, W. Kuich, ed., 218-229.

[38] Z. Ésik and S. Okawa, Serial and parallel operations on pomsets, In: proc. *Foundations of Software Technology and Theoretical Computer Science, Chennai, India, 1999,* LNCS 1738, Springer-Verlag, 1999, 316-328.

[39] Z. Ésik and I. Simon, Modeling literal morphisms by shufle, *Semigroup Forum,* 56(1998), 225-227.

ANALYSIS OF MEDICAL IMAGES AND NETWORK COMMUNICATION

Márta Fidrich

I. ANALYSIS OF MEDICAL IMAGES

I started to study medical images in scale space during my PhD at INRIA in France. I applied this knowledge to analyse facial laser scans while I was a research fellow in Leeds, and this led to a successful collaboration between the Department of Informatics at the University of Szeged and the Centre of Medical Imaging Research at Leeds.

I.1. Scale-space theory

The scale-spce theory aims to apply the model of *human vision* in image processing, so that the analysis of images become automatic. The homogenious *linear scale space* is the most studied and understood example for scale spaces. An image in linear scale space is simplified as scale increases: noise and small-scale features disappear, only the main structures remain. Unfortunately, images become blurred, even at places where it is not requested (at edges). On the other hand, this scale space has a vital role to *extract image features correctly*.

We analysed features (such as high-curvature corner points, ridges), defined by differential invariants, in 2D and 3D at multiple extraction scales; see Fig. 1. In particular, we analysed the stability of corner points in this scale space [2], we assessed their significance based on scale-related properties and we compared the usefulness of different stability criteria for registration.

It is well-known, that the position of image features may change as scale increases. Thus *features have to be* detected at a coarse scale then *followed* to the original fine scale. This is rather difficult in practice, since both spatial position and scale are discretized. I developed a new method to follow features efficiently across scale, based on the concept of intersecting isosurfaces. This methodology was applied to medical image processing such as automatic labelling of organs, matching and craniofacial reconstruction. For a description, see [3].

I.2. Analysis of facial growth

Our primary interest was to study growing faces of individuals (versus cross-sectional data) in order to compare the development of control subjects with subjects having problems with asymmetry (hemifacial microsomia) or orthodontics. To measure facial dimensions, the optical *laser scanner* is used. I worked on the statistical description of the data, visual perception of age and modelling shape change during growth. First results can be found in our study report [4], which was followed by specialized statistical analysis of facial growth, see [5, 6].

I.3. Robust surface matching

A necessary tool to detect changes due only (mostly) to growth is a suitable *registration algorithm*. There is a vast literature on registration of grayscale datasets. However, these algorithms could not be applied to the alignement of facial scans, which needed specialized methodologies.

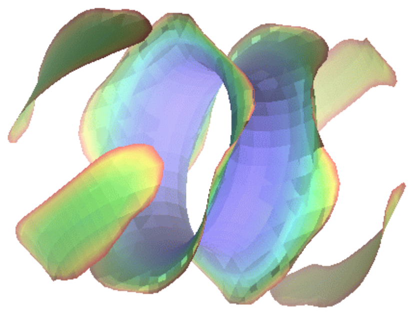


Figure 1. Change of parabolic lines on a saddle; shades indicate different scales.

That is why we developed a technique based on iteratively matching surfaces using information about surface curvature. What makes our algorithm accurate and robust is a new way to find correspondence between point sets: CSM. A robust Corresponding point is calculated by determining the Sensitivity of a correspondence to Movement of the point being matched. If the correspondence is reliable, a perturbation in the position of this point should not result in a large movement. A measure of reliability is also calculated. The correspondence calculation method is independent of the (iterative) registration algorithm, their combination provides very good results. We have several papers describing our method [7, 8, 9].

II. NETWORK COMMUNICATION

Research related to communication and network technologies has started only in 1999 at the Departments of Informatics under my direction. Our topics focus on the changeover of the networking protocol of the Internet that should take place soon. We have a public web page,
http://nokia5.inf.u-szeged.hu/pubpage/pubhp.html
containing information on our projects.

II.1. Comparison of IPv4 and IPv6

The Internet Protocol version 4 (IPv4) has been available since 1981. Although its success is indisputable, the rapid growth of the Internet has created a number of problems for the administration and operation of the global network. The new version 6 of the Internet Protocol (IPv6) was specified in 1995 and it has been finalized after years of debate. Presently the transition is underway, which would take still some years and will possibly raise new questions to be replied and new problems to be solved.

We wrote a comprehensive document [10] about the two network layer protocols: IPv4 and IPv6. We have pointed to the limits of IPv4 and compared the two protocols from several aspects (base functionality, addressing & routing, neighbor discovery & autoconfiguration, security, mobility, realtime support & quality of service). We also payed special attention to the transition from IPv4 to IPv6 and to the problems caused by the future co-existance of these incompatible protocols (possibility of interwork, routing aspects & practice, implementation issues). This handbook is composed of four chapters and an appendix, and suggested to be used for teaching purposes.

II.2. Inter-operation mechanisms

To facilitate the long process of changeover, several inter-working and transition mechanisms were designed between IPv4 and IPv, falling into two categories. The first case is when we want to run an IPv6 network using the existing infrastructure i.e. IPv4 network. This could be automatic / configured *tunneling*, *6over4* or *6to4*. The second, more difficult case is when IPv4-only and IPv6-only hosts want to communicate with each other, an example is NAT-PT employing SIIT *translator*).

Before buckling to the overall change of the installed protocols, it is worth analyzing and comparing these mechanisms – and that motivated our next study. We tested and compared configured tunneling, 6to4 and SIIT, against plain IPv4 and IPv6 sending and receiving. We had two aims concerning the tests: *algorithm-test* shows the overhead caused by the additional packet processing steps of the selected techniques against plain IPv4 and IPv6 mechanisms; while *router-test* indicates how fast the machine works as an Internet router and how big its packet throughput is. We presented our measurement results with detailed explication. The interested reader is referred to [11].

II.3. New communication environment

The 3G (Third Generation) Mobile Networks use SIP (Session Initiation Protocol) as their call control protocol and IPv6 as network layer protocol for wireless communication. When IPv6 was chosen for 3G networks, an urgent demand appeared in connecting the existing IPv4-based networks to the new (IPv6) one.

The goal of our following work was to create a demonstration system established on a novel way of connecting 3G mobile networks based on IPv6 and legacy Internet phones based on IPv4. The main idea was to connect a mobile SIP User Agent, based on IPv6, and another SIP User Agent, based on IPv4, via two SIP proxies (IPv6 and IPv4) and NAPT-PT (Network Address Port Translator - Protocol Translator). This communication environment (see Figure 2) requires the development of three entities making a fairly complex system, thus the application of novel solutions and expert techniques. See [12].

ACKNOWLEDGEMENT

The networking part was financed by Nokia Hungary. Participants of the team: Bátri, Dénes; Bohus, Mihály; Dikán, Gábor; Fidrich, Márta; Fóris, Gábor; Hendlein, Péter; Kiss, Gergő; Martonossy, László; Notaisz, Krisztián; Sebő, Marianna; Sógor, László; Somlai, Gábor; Tarjányi, Tamás.

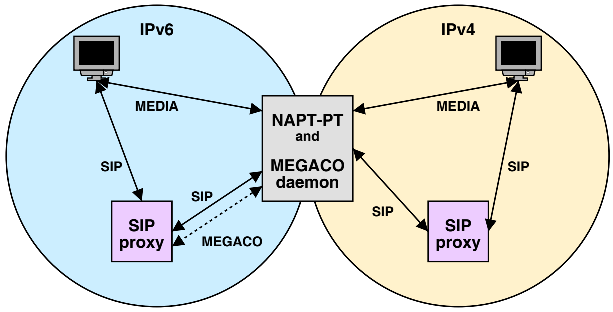


Figure 2. Test environment for multimedia call initiation between IPv4 and IPv6

REFERENCES

[1] M. Fidrich: *Multiscale analysis in medical image processing.* Invited lecture at the Chemometrics'99 Conference, Oct 1999.

[2] M. Fidrich J-P. Thirion: *Stability of Corner Points in Scale Space: The Effects of Small Non-Rigid Deformations.* Computer Vision and Image Understanding, vol. 72, no. 1, pp. 72-83, 1998.

[3] M. Fidrich: *Iso-surface Extraction in nD applied to Tracking Feature Curves across Scale.* Image and Vision Computing, vol. 16, pp. 545-556, 1998.

[4] M. Fidrich, K.V. Mardia, J.T. Kent, R.G. Aykroyd and A. Linney: *Analysis of facial change due to growth.* EPSRC Res. Report, Leeds, 1998.

[5] R.J. Morris, J.T. Kent, K.V. Mardia, M. Fidrich, R.G. Aykroyd and A. Linney: *Analysing Growth in Faces.* Proc. CISST, pp. 404-410, Las Vegas, 1999.

[6] R.J. Morris, J.T. Kent, K.V. Mardia, M. Fidrich, R.G. Aykroyd and A. Linney: *Exploratory Analysis of Facial Growth.* Proc. Conf. on Spatial Temporal Modelling and its Applications, pp. 39-42, Leeds, UK, 1999.

[7] E. Guest, M. Fidrich, S. Kelly and E. Berry: *Robust Surface Matching for Registration.* Proc. 3DIM, pp. 169-177, Ottawa, 1999.

[8] E. Guest, M. Fidrich, S. Kelly and E. Berry: *Evaluation of a Novel Algorithm for Surface Registration and Matching..* Proc. MIUA, pp. 109-112, Oxford, 1999.

[9] E. Guest, E. Berry, R.A. Baldock, M. Fidrich and M.A. Smith: *Robust Point Correspondance Applied to 2- and 3-Dimensional Image Registration* IEEE PAMI, vol. 23, pp. 165-179, 2001.

[10] M. Bohus, G. Dikán, M. Fidrich, L. Martonossy and G. Somlai: *Comparison of IPv4 and IPv6.* Lect. Notes on Network Comm., Nokia, 1999.

[11] L. Sógor, M. Fidrich, L. Martonossy, P. Hendlein and G. Somlai: *Comparison of inter-operation mechanisms between IPv4 and IPv6.* Proc. 9th IFIP Conf., 2001, pp 402-413.

[12] L. Sógor, P. Hendlein, K. Notaisz, M. Fidrich, G. Fóris, and G. Kiss: *Development of a Communication Environment between IPv4 and IPv6.* Proc. 7th SPLST Symp., 2001, pp. 231-243.

RESEARCH RESULTS ON TREE PROCESSING DEVICES

Zoltán Fülöp

I. TREE TRANSDUCERS

In [1] we made a comprehensive study of top-down tree transducers, attributed tree transducers, and macro tree transducers. The previous results are summarized and discussed in a unified terminology. Moreover, some new results are presented mainly concerning the comparison of the computation power of the different tree transducer models.

It is known for a long time that macro tree transducers are more powerful than attributed tree transducers. In [5] we give a characterization of the class of tree transformations which are computed by (non circular) attributed tree transducers, in terms of so called attributed-like macro tree transducers. Moreover, we prove that it is decidable whether a macro tree transducer is attributed-like.

In [6] we prove that the domains of partial attributed tree transducers are the tree languages recognized by tree walking automata in universal acceptance mode. Moreover, we prove that the domains of partial attributed tree transducers having monadic output are the tree languages recognized by deterministic tree walking automata.

In [8] we consider attributed tree transducers of type simple multi-visit, multi-sweep, multi-alternating pass, and multi-pass following the definition of the same subclasses of attribute grammars. We characterize the tree transformation classes induced by the considered four subclasses of attributed tree transducers in terms of macro tree transducers. Namely, we define simple multi-visit, multi-sweep, multi-alternating pass and multi-pass macro tree transducers and prove formally that the attributed tree transducers and the macro tree transducers of the same type induce the same tree transformation class. Also we give an inclusion diagram of the tree transformation classes induced by the above and some further fundamental types of attributed and of macro tree transducers.

In [10] we consider the closure $UCI(Rel)$ of the class of relabeling tree transformations, under U=union, C=composition and I=iteration. We give a characterization of $UCI(Rel)$ in terms of a short expression built up from Rel with composition and iteration. We also give a characterization of $UCI(Rel)$ in terms of one-step rewrite relations of very simple term rewrite systems. We give a similar characterization of $UC(FRel_+)$ where $FRel_+$ is the class consisting of the transitive closures of all functional relabeling tree transformations. Finally we show that $UCI(Rel) = UCI(FRel)$.

In [11] we generalize bottom-up tree transducers and top-down tree transducers to the concept of *bottom-up tree series transducer* and *top-down tree series transducer*, respectively, by allowing formal tree series as output rather than trees, where a formal tree series is a mapping from output trees to some semiring. We associate two semantics with a tree series transducer: a mapping which transforms trees into tree series (for short: tree to tree series transformation or t-ts transformation), and a mapping which transforms tree series into tree series (for short: tree series transformation or ts-ts transformation).

We show that the standard case of tree transducers is reobtained by choosing the boolean semiring under the t-ts semantics. Moreover, we show that certain fundamental constructions and results concerning bottom-up and top-down tree transducers can be generalized for the corresponding tree series transducers. Among others, we prove that polynomial bottom-up t-ts transformations can be characterized by the composition of finite state relabeling t-ts transformations and boolean homomorphism t-ts transformations. Moreover, we prove that every deterministic top-down t-ts transformation can be characterized by the composition of a boolean homomorphism t-ts transformation and a deterministic linear top-down t-ts transformation. We prove that deterministic top-down t-ts transformations are closed under right composition with nondeleting and linear deterministic top-down t-ts transformations and are closed under left composition with boolean and total deterministic top-down t-ts transformations. Finally we show that nondeleting linear bottom-up and nondeleting linear top-down tree series transducers generate the same t-ts transformation class.

II. COOPERATING DISTRIBUTED TREE AUTOMATA AND TRANSDUCERS

In [4] distributed tree processing devices, like regular tree grammars, top-down tree automata and transducers, and bottom-up tree automata and transducers are considered. The concept of distribution lies in that the set of rewriting rules are distributed among n sets, called components. The components work on deriving a sentential form such that they cooperate with each other concerning cooperation strategies. We develop a technical toolkit for studying distributed tree processing devices. We mainly consider the $\dashv$ strategy which means that a component should work on a sentential form as far as it can. We show that in general the above distributed tree processing devices with $\dashv$ cooperation strategy are more powerful than ordinary ones with respect to generating capacity.

In [2] we prove that the class $ETOLT$, the class of tree languages generated by ETOL tree systems, can be characterized in the following way: $ETOLT$ is equal to the class of tree languages generated by distributed regular tree grammars cooperating with termination strategy [4]. This result is a generalization of the corresponding one for ETOL systems and distributed context-free (string) grammars cooperating with terminal strategy.

In [9] we show that the component hierarchy of chain-free distributed regular tree grammars cooperating with terminal strategy [4] is infinite with respect to tree language generating capacity. More exactly, we prove that $cfCDRTG_{\dashv}(n) \subset cfCDRTG_{\dashv}(2(n-1)^2+3)$, where $n \geq 1$ and $cfCDRTG_{\dashv}(n)$ denotes the class of tree languages generated by chain-free distributed regular tree grammars of at most n components cooperating with terminal strategy.

III. TREE TRANSDUCERS VS TERM REWRITING

In [3] two restricted ways to apply a term rewriting system (TRS) to a tree are considered. When the *one-pass root-started* strategy is followed, rewriting starts from the root and continues stepwise towards the leaves without ever rewriting a part of the current tree produced in a previous rewrite step. *One-pass leaf-started rewriting* is defined similarly, but rewriting begins from the leaves. In the *sentential form inclusion problem* one asks whether all trees which can be obtained with a given TRS from the trees of some regular tree language T belong to another given regular tree language U, and in the *normal form inclusion problem* the same question is asked about the normal forms of T. We show that for a left-linear TRS these problems can be decided for both of our one-pass strategies. In all four cases the decision algorithm involves the construction of a suitable tree recognizer.

In [7] we consider restricted versions of ground tree transducers: total, deterministic, and symmetric subclasses and all other subclasses created by applying any combination of these restrictions. We present the inclusion diagram of the tree transformation classes induced by these restricted ground tree transducers.

We show that the following four classes of term relations are the same: (i) tree transformations induced by symmetric deterministic ground tree transducers, (ii) congruence relations on term algebras induced by reduced ground term rewriting systems, (iii) congruence relations on term algebras induced by convergent ground term rewriting systems, and (iv) finitely generated congruence relations on term algebras.

As a by-product of our results, we obtain a new ground completion algorithm.

Moreover, we show that the following three classes of term relations on term algebras with at least one non-nullary function symbol are also the same: (i) tree transformations induced by total symmetric deterministic ground tree transducers, (ii) congruence relations on term algebras of finite index, (iii) finitely generated congruence relations on term algebras of which the trunk is the whole set of terms.

ACKNOWLEDGEMENT

The above research has been supported by the grant FKFP 0095/1997 from the Ministry of Education of Hungary, and the grants OTKA T 015843 and OTKA 030084 from the National Foundation of Hungary for Scientific Research.

REFERENCES

[1] Z. Fülöp and H. Vogler, Syntax-Directed Semantics – Formal Models based on Tree Transducers, Monographs in Theoretical Computer Science, An EATCS Series, Springer-Verlag, 1998.

[2] Z. Fülöp and S. Maneth, A Characterization of ETOL Tree Languages by Cooperating Regular Tree Grammars, Grammatical Models of Multi-Agent Systems (G. Paun and A. Salomaa Eds.), Gordon and Breach, 1998, 182-192.

[3] Z. Fülöp, E. Jurvanen, M. Steinby, and S. Vágvölgyi, On One-Pass Term Rewriting, Proc. of MFCS 98, *Lecture Notes in Computer Science* (Szerk. L. Brim, J. Gruska, és J. Zlatusak), Vol. 1450, Springer-Verlag, 1999, 248-256. Journal version: *Acta Cybernetica* **14** (1999) 83-98.

[4] Z. Fülöp, Distributed Tree Processing Devices, *Computers and Artificial Intelligence*, **18** (1999) 37-71.

[5] Z. Fülöp and H. Vogler, A Characterization of Attributed Tree Transformations by a Subclass of Macro Tree Transducers, *Theory of Computing Systems*, **32** (1999) 649-676.

[6] Z. Fülöp and S. Maneth, Domains of Partial Attributed Tree Transducers (S. Manethval), *Inf. Process. Letters*, **73** (2000) 175-180.

[7] Z. Fülöp and S. Vágvölgyi, Restricted Ground Tree Transducers, *Theoret. Comput. Science*, **250** (2001) 219-233.

[8] Z. Fülöp, Characterizing Important Subclasses of Attributed Tree Transformations by Macro Tree Transducers, *Journal of Automata, Languages and Combinatorics* **6** (2001) 145-189.

[9] G. Dányi and Z. Fülöp, The Component Hierarchy of Chain-Free Cooperating Distributed Regular Tree Grammar Systems, *Theoret. Comput. Science*, **262** (2001) 229-240.

[10] Z. Fülöp and A. Terlutte, Iterated Relabeling Tree Transducers, *Theoret. Comput. Science*, 2001, accepted for publication.

[11] Z. Fülöp and H. Vogler, Bottom-up and Top-down Tree Series Transformations, *Journal of Automata, Languages and Combinatorics*, 2001, accepted for publication.

STRUCTURE THEORY OF AUTOMATA

Ferenc Gécseg and Balázs Imreh

Structure theory is a traditional area of automata theory that has been the source of many deep results on this field. Properties of the members of special classes of automata and nondeterministic automata render it possible to

(1) characterize the subdirect irreducible members of the class considered,

(2) help to describe the isomorphically and homomorphically complete systems for specific classes of automata under different compositions,

(3) describe the languages accepted by the members of the classes under consideration,

(4) study some questions connected with the directable automata.

The results concerning the structural investigations of the authors are summarized below.

Regarding the subdirectly irreducible automata, three special classes, the classes of definite, reverse-definite and generalized definite automata are studied in [3], where the subdirectly irreducible members of these classes are characterized. By using structural properties, originated from semigroup theory, a general description of the subdirectly irreducible members of the class of automata is presented in [1].

As far as the isomorphically complete systems are concerned, in [5] the class of the generalized definite automata is studied and the isomorphicaly complete systems of generalized definite automata with respect to the α_i-products are characterized. From this characterization it follows that there is no finite isomorphically complete system for this class with respect to the α_0-product, while there is a singleton isomorphically complete system for the class under consideration with respect to any α_i-product with $i \geq 1$. Another specific class, the class of the commutative asynchronous automata is studied in [14] where the isomorphically complete systems of the commutative asynchronous automata are described with respect to the Glushkov-type product. More general classes, the class of monoton tree automata, and the class of nondeterministic monotone tree automata are investigated in [6] where the isomorphically complete systems are characterized for these classes with respect to the α_0-products. From these descriptions it follows that there are singleton isomorphically complete systems for both classes with respect to the α_0-product. In [4] a structural characterization of three classes of tree automata is presented, namely, the classes of nilpotent, definite, and monotone tree automata are homomorphically represented by means of quasi-cascade-products of unary nilpotent and unary definite tree automata in the first two cases, and by means of products of simpler tree automata in the third case. The isomorphically complete systems for the class of all automata under different compositions are studied in the papers [10] and [8]. In [10] it is shown that the cube-product is equivalent to the Glushkov's product of automata, and thus, the isomorphically complete systems collapse for this two compositions. In [8] a general notion of composition is studied which is defined by a graph-class. A sufficient condition is presented for the graph-class which provides that there exists a finite isomorphically complete system for the composition defined by this graph-class.

Different special classes of languages are described in the papers [7, 9, 11, 12, 13]. In [11, 12] the classes of languages consisting of the languages containing the directing words of nondeterministic directable automata are described and compared. A similar investigation can be found in [16] for commutative directable nondeterministic automata. A small particular language class, the class of the languages accepted by commutative asynchronous automata is described in [13]. Another special class, the class of the languages recognizable by monotone automata is characterized in [7]. The tree languages recognized by deterministic root-to-frontier tree automata are characterized in [9].

In [15] the directable commutative asynchronous automata are studied and a tight upper bound is presented for the maximum of the shortest directing words of n-state directable commutative asynchronous automata. The notion of the directability of automata is extended to nondeterministic automata in [17] where lower and upper bounds of the maximum of the shortest directing words of n-state directable nondeterministic automata are investigated. This extension yielded further studies (see [11, 12, 16]).

RELATIONSHIP BETWEEN THIS RESEARCH TOPIC AND THE EDUCATION

Results concerning the structure studies of automata can constitute different PhD courses, furthermore, they provide numerous possibilities for further research work.

ACKNOWLEDGEMENT

Research sketched above was supported by the Hungarian National Foundation for Scientific Research, Grant T 030143, the Ministry of Culture and Education of Hungary, Grant FKFP 0704/1997, the Japanese Ministry of Education, Mombusho International Scientific Research Program, Joint Research 10044098, the Hungarian-Finish S & T Co-operation Programme for 1997-1999, Grant SF-10/97.

REFERENCES

[1] Bogdanović, S., M. Ćirić, B. Imreh, T. Petković, M. Steinby, Traps, Cores, Extensions and Subdirect Decomposition of Unary Algebras, *Fundamenta Informatica* **38**, (1999), 51-60.

[2] Bogdanović, S., M. Ćirić, B. Imreh, T. Petković, Directable automata and their generalizations: a survey, *Novi Sad J. Math.* **29**, (1999), 29-69.

[3] Ćirić, M., B. Imreh, M. Steinby, Subdirectly irreducible definite, reverse-definite, and generalized definite automata, *Univ. Beograd, Publ. Elektrotechn. Fak. Ser. Mat.* **10** (1999), 69-79.

[4] Gécseg, F., On Some Classes of Tree Automata and Tree Languages, *Annales Academiae Scientiarum Fennicae Mathematica* **25** (2000), 325-336.

[5] Gécseg, F., B. Imreh, On isomorphic representations of generalized definite automata, *Acta Cybernetica* **15** (2001), 33-44.

[6] Gécseg, F., B. Imreh, On isomorphic representation of monotone tree and non-deterministic tree automata, to appear.

[7] Gécseg, F., B. Imreh, On monotone automata and monotone languages, *Journal of Automata, Languages, and Combinatorics*, to appear.

[8] Gécseg, F., B. Imreh, A. Pluhár, On existence of finite isomorphically complete systems, *Journal of Automata, Languages, and Combinatorics* **3** (1998), 77-84.

[9] Gécseg, F. and Steinby, M.: Minimal recognizers and syntactic monoids of DR tree languages, to appear.

[10] Imreh, B., On the equivalence of the cube-product and the generalized product of automata, *Theoretical Computer Science* **245** (2000), 103-113.

[11] Imreh, B., M. Ito, On some special classes of regular languages, in: *Jewels are Forever* (eds.: J. Karhumaki, H. Maurer, G. Paun and G. Rozenberg,), Springer-Verlag, Berlin-Heidelberg-New York-Tokyo, 1999, 25-34.

[12] Imreh, B., M. Ito, Nondeterministic directable automata and related languages, *RIMS Kokyuroku* **1106** (1999), 81-87.

[13] Imreh, B., M. Ito, A. Pukler, A note on the languages recognized by commutative asynchronous automata, *RIMS Kokyuroku* **1166** (2000), 95-99.

[14] Imreh, B., M. Ito, A. Pukler, On commutative asynchronous nondeterministic automata, *Acta Cybernetica* **14** (2000), 607-617.

[15] Imreh, B., M. Ito, A. Pukler, On commutative asynchronous automata, in the Proceedings of the Third International Colloquium on Words, Languages, and Combinatorics, Kyoto, 2000, to appear.

[16] Imreh, B., M. Ito, M. Steinby, On commutative directable nondeterministic automata, to appear.

[17] Imreh, B., M. Steinby, Directable nondeterministic automata, *Acta Cybernetica* **14** (1999), 105-115.

DEPENDENCE ANALYSIS OF DECLARATIVE PROGRAMS AND THEIR APPLICATIONS IN MACHINE LEARNING PROBLEMS

Tibor Gyimóthy, Gyöngyi Szilágyi and Szilvia Zvada

Attribute Grammars (AG) are generalizations of the concept of Context-Free Grammars. The formalism of AGs has been widely used for the specification and implementation of programming languages. On the other hand there is a close relationship between AGs and another declarative approach, (Constraint) Logic Programming (LP) an emerging software technology with growing number of applications. Data analysis of (Constraint) Logic Programs is useful in finding the connected components of programs, plays an important role for example in debugging, testing and program maintenance.

Efficient techniques for learning attribute grammars can help in finding good definitions of AGs, which usually require a lot of effort.

In this short study we give a summary of results achieved during 1998-2001 in the fields of program analysis (slicing) of (Constraint) Logic Programs, and the learning of semantic functions of Attribute Grammars.

I. DATA FLOW ANALYSIS OF LOGIC PROGRAMS

Slicing is a program analysis technique originally developed for imperative languages. It can be applied in a number of software engineering tasks, is a natural tool for debugging, is useful in incremental testing, and can help to detect dead code or to find parallelism in programs. Intuitively, a program slice with respect to a specific variable at some program point contains all those parts of the program that may affect the value of the variable (backward slice) or may be affected by the value of the variable (forward slice).

Data flow in logic programs is not explicit, and for this reason the concept of a slice and the slicing techniques of imperative languages are not directly applicable. Moreover, implicit data flow makes the understanding of program behavior rather difficult. Thus program analysis tools explaining data flow to the user are of great practical importance. Papers [1, 2, 3] extends the scope and optimality of previous algorithmic debugging techniques of Prolog programs using slicing techniques. They provide dynamic slicing algorithms augmenting the data flow analysis with control flow dependence to help one locate the connected components of a program and the source of a bug included in the program. A tool has been developed for debugging Prolog programs which also handles the specific programming techniques (cut, if-then, or).

II. DATA FLOW ANALYSIS OF CONSTRAINT LOGIC PROGRAMS

Constraint Logic programming is a fusion of two declarative paradigms: constraint solving and logic programming. The framework extends classical logic programming by removing the restriction on programming within the Herbrand universe alone; unification is replaced by the more general notion of constraint satisfaction. In [4, 5, 6] we formulate declarative notions of a slice suitable for CLP. (The problem of finding minimal slices may be undecidable in general, since satisfiability may be undecidable.) Our slice definition provides a basis for defining slicing techniques based on variable sharing. The techniques are further extended by using groundness information. We also present a prototype slicing tool for CLP programs written in SICStus Prolog.

III. LEARNING SEMANTIC FUNCTIONS OF ATTRIBUTE GRAMMARS

In the framework of compilation oriented language implementation, Attribute Grammars (AG) are the most widely applied semantic formalism. Since the definition of an AG and its semantic functions may be complex it is very useful to have an efficient tool for inferring semantic rules of AGs from examples. Based on the correspondence of Attribute Grammars and Logic Programs some of the learning technique developed for Logic Programs (ILP) could be applied to AGs. Introducing an AG based description language in ILP implies the definition of an Attribute Grammar Learner. In [7] we present a parallel method for learning semantic functions of Attribute Grammars based on an ILP approach. The method given is adequate for S-attributed and for L-attributed grammars. The parallelism can help to reduce the large number of user queries posed during the interactive learning process.

In [8, 9] a learning method called LAG is presented, which infers semantic functions for simple classes of AGs by means of examples and background knowledge. LAG generates the training examples on its own via the effective use of background knowledge. The LAG method makes use of the C4.5 decision tree learner during the learning process. This method has been applied to the Part-of-Speech tagging of Hungarian sentences, as well.

ACKNOWLEDGEMENTS

The above mentioned research tasks were supported by the following grants: OTKA T25721, FKFP 1354/1997, IKTA 8/99, ESPRIT 20237 ILP, OMFB 96/97 67-1352 IQDEBUG.

REFERENCES

[1] L. Harmath, Gy. Szilágyi, T. Gyimóthy, J. Csirik: Dynamic Slicing of Logic Programs. Program Analysis and Verification, Fenno Ugric Symposium (FUSST'99), Estonia, 1999, pages 101-113.

[2] L. Harmath, Gy. Szilágyi, T. Gyimóthy: Debug Slicing of Logic Programs. Conference of PhD Students on Computer Sciences, Hungary, 2000, pages 43-44.

[3] Gy. Szilágyi, L. Harmath, T. Gyimóthy: Debug Slicing of Logic Programs. Acta Cybernetica, 2001 (accepted for publication).

[4] Gy. Szilágyi, J. Maluszynski, T. Gyimóthy: Slicing of Constraint Logic Programs. Linköping

University Electronic Press, 1998,
http://www.ep.liu.se/ea/cis/1998/002.

[5] Gy. Szilágyi, J. Maluszynski, T. Gyimóthy: Slicing of Constraint Logic Programs. In Proceedings of the Fourth International Workshop on Automated Debugging (AADEBUG 2000), München, Germany, 2000, pages 176-187.

[6] Gy. Szilágyi, J. Maluszynski, T. Gyimóthy: Static and Dynamic Slicing of Constraint Logic Programs. Journal of Automated Software Engineering. Kluwer Academic published, 2001 (accepted for publication).

[7] Gy. Szilágyi, A. M. Thanos: PAGELEARN: Learning Semantic Functions of Attribute Grammars in Parallel. Journal of Computing and Information Technology (C.I.T.), Vol. 8, No. 2, June 2000, pages 115-131.

[8] Z. Alexin, Sz. Zvada, T. Gyimóthy: Application of AGLEARN on Hungarian Part-of-Speech Tagging. Proceedings of the Second Workshop on Attribute Grammars and their Applications (WAGA'99), Amsterdam, The Netherlands, 1999, pages 133-152.

[9] Sz. Zvada, T. Gyimóthy: Using Decision Trees to Infer Semantic Functions of Attribute Grammars. Acta Cybernetica, 2001 (accepted for publication).

GENETIC AND EVOLUTIONARY ALGORITHMS

Márk Jelasity

Most of the research done in the interval 1999-2001 is included in my Ph.D. thesis. [5]. Research conducted in artificial speech recognition [9, 8] is covered elsewhere.

I. HEURISTIC GLOBAL OPTIMIZATION

This line of research falls under the field of evolutionary computing. [5, 6] describe global optimizers that also explore the structure of the local optima of an optimization problem. The first version is called GAS. This algorithm is able to make use of structural information adaptively using a niching technique based on a subpopulation approach simulating the formation of species. This means that the search space is divided to clusters (which we call species), and these cluster-structure is constantly modified according to the results of global search restricted to these clusters. The cluster structure reflects the distribution of local optima.

The bias of the algorithm can be controlled explicitly, as the structure of the space (the distance function of the space) can be implemented arbitrarily, and the automatic parameter setting of some important parameters is possible based on this distance function and other user-given parameters. The output of the algorithm contains the structural information as well as the final solution. The second version is called UEGO.

The difference from GAS is that the optimizer that works inside of a species can be any algorithm. This turns the approach into a general hybrid paradigm that can be applied along with any optimizer if the structure of the local optima is important or if stable solutions must be found. Another difference is that the structure of the algorithm was simplified to make parallel implementation easier. For results of the parallel implementation please consult [11].

II. EVOLUTIONARY COMPUTATION THEORY

[7] offers a way of visualizing the structure of spaces and also provides a measure of problem difficulty. In this case the visualization is done based on the evaluation function *and* the algorithm at hand so it is a tool to look at the structure defined by the evaluation function from the viewpoint of the search algorithm. In 2001 Boglárka Tóth and Tamás Vinkó won the first prise at the National Competition of Students (OTDK) in this subject area under my supervision.

[4] offers an approach that makes this automatic process possible. A framework is presented in which binary encodings of a problem domain can be learned using ideas from machine learning. The actual learning procedure can be arbitrary. Using such an approach, it may be possible in the future to solve the reverse engineering problem in a similar way evolution solved it for us evolving "devices" such as the eye and human language. The later — besides its other functions — allows us to describe the human domain, i.e. the aspects of the world that are important for us. This makes is possible to plan our actions and to predict future events efficiently. In abstract domains a similar approach of evolving such representations that allow efficient search, information compression and even communication is becoming computationally feasible.

III. OTHER GENETIC COMPUTATION

[1] it is suggested that the way of looking at evolution in the field of evolutionary computation is very similar to the so called *adaptational stance*. It is also shown that — beside the similarities — there are significant differences as well, and these are connected to the epistemological crisis of computational science. The point is that since biologists have a natural vocabulary for talking about features of things (wing, eye, hart, etc.) in computer science the situation is radically different. To imagine this try to talk about an elephant referring only to its DNA sequence; well, in evolutionary computation people have no other information. I also propose that the problem is deep and we have no grounds to think that it will ever be solved by humans. Human intelligence was evolved in a specific environment for specific tasks. Though the emergence of language made it possible to reach high levels of abstraction, to really describe such spaces and to make use of their structure the only way might be to do it automatically using artificial intelligence and machine learning where concepts that are completely un-natural for humans can be handled and evolved.

REFERENCES

[1] Márk Jelasity. The adaptationist stance and evolutionary computation. In Wolfgang Banzhaf, Jason Daida, Agoston E. Eiben, Max H. Garzon, Vasant Honavar, Mark Jakiela, and Robert E. Smith, editors, *GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1859–1864, San Francisco, CA, 1999. Morgan Kaufmann.

[2] Márk Jelasity. Genetikus algoritmusok. In Iván Futó, editor, *Mesterséges Intelligencia*, pages 549–568. Aula Kiadó, 1999.

[3] Márk Jelasity. Instance-based manifesto? *Behavioral and Brain Sciences*, 23(4):482–483, August 2000. Commentary.

[4] Márk Jelasity. Towards automatic domain knowledge extraction for evolutionary heuristics. In Marc Schoenauer, Kalyanmoy Deb, Günter Rudolph, Xin Yao, Evelyne Lutton, Juan Julian Merelo, and Hans-Paul Schwefel, editors, *Parallel Problem Solving from Nature - PPSN VI*, volume 1917 of *Lecture Notes in Computational Science*, pages 755–764. Springer-Verlag, 2000.

[5] Márk Jelasity. *The shape of evolutionary search: discovering and representing search space structure.* PhD thesis, Leiden University, January 2001. IPA Dissertation Series 2001-01.

[6] Márk Jelasity, Pilar Martínez Ortigosa, and Inmaculada García. UEGO, an abstract cluster-

ing technique for multimodal global optimization. *Journal of Heuristics*, 7(3):215–233, May 2001.

[7] Márk Jelasity, Boglárka Tóth, and Tamás Vinkó. Characterizations of trajectory structure of fitness landscapes based on pairwise transition probabilities of solutions. In *Proceedings of the 1999 Congress on Evolutionary Computation (CEC99)*, pages 623–630. IEEE, IEEE Press, 1999.

[8] András Kocsor, András Kuba, Jr, László Tóth, Márk Jelasity, László Felföldi, Tibor Gyimóthy, and János Csirik. A segment-based statistical speech recognition system for isolated/continuous number recognition. In Jaan Penjam, editor, *Software Technology (FUSST'99) Proceedings*, pages 201–211, Tallinn, Estonia, 1999.

[9] András Kocsor, László Tóth, András Kuba, Jr, Kornél Kovács, Márk Jelasity, Tibor Gyimóthy, and János Csirik. A comparative study of several feature transformation and learning methods for phoneme classification. *International Journal of Speech Technology*, 3(3/4):263–276, 2000.

[10] Pilar Martínez Ortigosa, Inmaculada García, and Márk Jelasity. Evaluating UEGO, an evolutionary global optimization algorithm. In *International Workshop on Global Optimization 99*, pages 97–100, Firenze, Italy, 1999.

[11] Pilar Martínez Ortigosa, Inmaculada García, and Márk Jelasity. Reliability and performance of UEGO, a clustering-based global optimizer. *Journal of Global Optimization*, 19(3):265–289, March 2001.

AUTOMATIC MAP INTERPRETATION

Endre Katona

To create a spatial database for some GIS (Geographic Information System) application, it is a big challenge to recognize all the simple and complex map objects automatically on scanned maps. This research field is termed as *map interpretation*. Related topics are OCR (Optical Character Recognition), *form analysis* and *engineering drawing interpretation*.

I. APPROACHES

Published map interpretation systems can be classified into three categories:

- Raster-based systems perform all recognition on scanned raster data.
- Vector-based solutions start with a raw vectorization, and all the recognition steps are performed on vector data.
- Hybrid solutions usually perform segmentation to separate text from graphics, and text is recognized with some raster-based OCR technique while graphics is analyzed after vectorization.

In our work we prefer the vector-based approach discussed below.

II. INTERPRETATION OF CADASTRAL MAPS

A new data model - called DG (Drawing Graph) - has been developed [5] to support vector-based recognitions. Our model involves four object types:

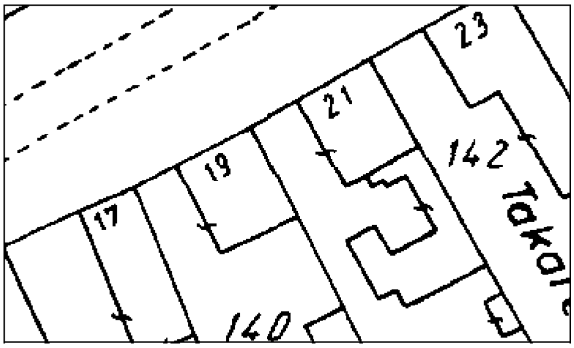
- NODE is a point with coordinates (x, y),
- EDGE is a straight line segment between two nodes,
- TEXT is an inscription on the map,
- PAT is an arbitrary set of other objects.

We show that these four types are sufficient to describe all complex structures required in map interpretation. The DG model can be discussed both on logical (abstract) level and on physical (machine data structure) level. Considering the latter one, it can be shown that initial topology generation takes $\mathcal{O}\ (N \cdot log(N))$ time for N vectors, and all update operations take constant time [3]. As a conclusion, DG is a simple, universal and efficient model and in this sense it is prior to other data structures applied in map interpretation.

The DG model has been implemented in the interpretation system MAPINT. Although the system is basically universal, it has special support for Hungarian cadastral maps [1]. Processing starts with an affine coordinate transform followed by raw vectorization, the result of which is converted into DG format. At this stage of processing DG contains only NODE and EDGE objects, while TEXT and PAT objects are generated during recognition. The final result can be exported in DXF format. Interpreted map objects are as follows (Fig. 1):

- dashed lines,
- inscriptions (house numbers and parcel numbers),
- connection signs (a special notation of Hungarian maps expressing the relationship between a building and a parcel),
- null-circles (denoting geodetical identified points),
- building and parcel polygons.

a,



b,

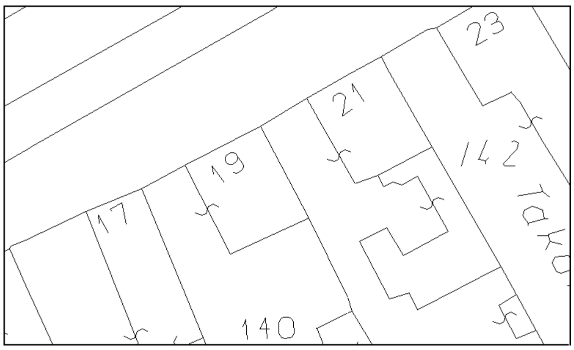


Figure 1. Automatic interpretation of Hungarian cadastral maps: a, scanned raster image, b, result of automatic recognition.

Character recognition is performed by a feed-forward back-propagation neural network ensuring learning abilities for the system. We show that practically all recognition operations can be performed in linear time when grid indexing is used for the DG [3].

MAPINT has been applied in the Phare HU 905.0203 Land Consolidation Project to support the creation of a Hungarian cadastral map database. In this project MAPINT has served for recognizing parcel numbers and connecting them with parcel records stored in an Oracle database [2].

III. TERRAIN INTERPRETATION

Thin plate spline interpolation is a widely used approach to generate a digital elevation model (DEM) from contour lines and elevation dots. In practice, contour maps are scanned and vectorized, and after resampling in the target grid resolution, interpolation is performed. In [3] we reveal the restricted accuracy of this process, and propose a high resolution processing method (without vectorization) ensuring maximum utilization of information involved in source data.

There are several ways to create a DEM from a scanned contour map. We prefer the thin plate spline interpolation with multigrid relaxation to speedup convergence. This method works well for scattered data points, but it gives incorrect result when applied for contour maps in scanning resolution. To overcome the problem, a *contour thinning reduction method* [4, 6] has been developed and applied in the multigrid process. This method gives correct result even if contour lines and elevation dots are processed at the same time (Fig. 2).

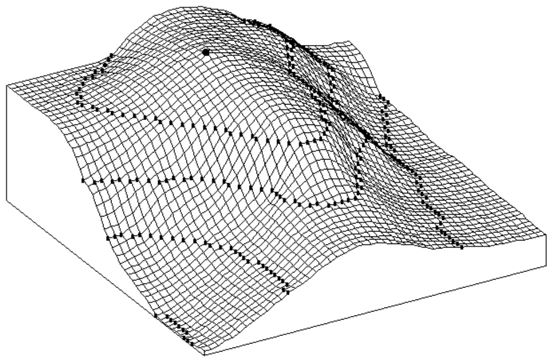


Figure 2. Result of multigrid relaxation with contour thinning reduction.

An earlier version of our technology has been applied in a Hungarian project producing the first DEM covering the whole country with 10 x 10 meter resolution. To overcome the problem of computation time of high resolution processing, a special parallel processor has been applied to speed up convolution.

REFERENCES

[1] E. Katona, Gy. Hudra: An Interpretation System for Cadastral Maps. Proceedings of 10th Internat. Conf. on Image Analysis and Processing, IEEE Computer Society, pp. 792-797, 1999.

[2] E. Katona, Gy. Hudra: Aquisiton of cadastral data by automatic map recognition. Térinformatika, 1999/8, pp. 12-13, 1999. (in Hungarian)

[3] E. Katona: Automatic map interpretation. Ph.D. Thesis, Department of Physical Geography, University of Szeged, 2001 (in Hungarian).

[4] E. Katona: Digital terrain modelling. Accepted to the journal Polygon, 2001 (in Hungarian).

[5] E. Katona: A data model for vector-based map interpretation. Submitted to Internat. Journal on Document Analysis and Recognition, Springer, 2001.

[6] E. Katona: Raster-based generation of digital elevation models using multigrid relaxation. Submitted to Internat. Journal of Geographical Information Science (Taylor and Francis, London), 2001.

CURRENT RESEARCH & DEVELOPMENT IN SPEECH RECOGNITION

András Kocsor, László Tóth, and János Csirik

I. MOTIVATION

When choosing the directions of our speech recognition research, we decided to focus on Hungarian with the hopes that we can address some special issues concerning the processing of our national language, and also that we can make use of our previous experience with NLP for Hungarian. Furthermore, we were looking for a flexible framework that allows experimentation with different preprocessing techniques, feature-space transformation methods and machine learning algorithms. These expectations led us to the stochastic segmental approach which, in a certain sense, can be viewed as an extension of hidden Markov modelling.

II. THE "OASIS" RECOGNITION SYSTEM

Our recognition system follows the stochastic segmental approach mentioned above and was named "OASIS" [5, 6, 8] (from Our Acoustics-based Speaker-Independent Speech recognizer). The system was designed to be as modular as possible, so we can easily conduct experiments with combining different techniques for the several subtasks of recognition. The first module is the usual frame-based preprocessing phase. After this an additional step, the modelling of phonetic segments was inserted. This allows the combination of different preprocessing techniques and also incorporation of phonetic knowledge. The output of this module are the so-called segmental features, which are used by phoneme classifiers for the classification of the segment. For this task, again, several learning algorithms can be applied. Finally, the segmental classification scores are combined by the matching engine, which performs an utterance-level search in the graph of the possible segmentations. Its goal is to find the best matching transcription from those provided by the language model. The system makes it possible to experiment with many search techniques and strategies.

III. DISCRIMINATIVE PHONEME CLASSIFICATION

The main advantage of the segmental approach from our point of view is that one can apply many traditional pattern recognition algorithms for the phoneme classification task. In a comparative study we found that discriminative classifiers like Artificial Neural Nets or Support Vector Machines can significantly outperform the usual HMM technique [7].

IV. NON-LINEAR FEATURE SPACE TRANSFORMATIONS

These classifiers can be made even more efficient with the aid of feature-space transformations. One of our main research interest currently is the application of non-linear transformations for phoneme classification. We wrote several studies investigating the efficiency of such transformations as Kernel-LDA, Kernel-PCA and Kernel-ICA [9, 10, 11, 12].

V. RUNNING PROJECTS

A very important issue in statistical speech recognition is the training database. Without a proper corpus the recognition system cannot be trained reliably. Unfortunately, there are no such corpora for Hungarian which would be large enough to train large vocabulary continuous speech recognizers. In the framework of the national grants "IKTA" we are currently working on such a large telephone-based speech corpus in co-operation with the Technical University of Budapest. We have already made a small corpus containing natural numbers, and our recognizer is trained and tested on this database. This corpus will be made available to any research institute in Hungary with the financial aid of governmental funds, within the framework of grant SZT-IS-10. This project also contains the creation of free source-code algorithms that aim to encourage speech recognition research and development in Hungary.

Besides, we have applied for grants in the field of speech impediment therapy (see below).

VI. SPEECH IMPEDIMENT THERAPY

A very interesting application of automatic speech recognition is the speech therapy of the hearing impaired, as this poses the most serious challenge for phoneme recognition. We study the applicability of our phoneme classification methods to this field [11]. Our cooperative partner in this project is the "Elementary School of Deaf Children".

VII. SOFTWARE

As regards to softwares, the most important result of our development is the "OASIS" modular system. It was developed in MS Visual C++ and lately was converted into a library. In this modular environment one can very easily conduct experiments with several feature extraction, phoneme classification and word recognition techniques. The number of functions contained in the library is dynamically extending as newer and newer routines are being added.

For the manual segmentation and labeling of our corpora we developed a software tool that we named "Segmentation Assistant". Besides the graphical environment, this software also lends segmentation algorithms to speed up manual segmentation. We intend to develop this program further by incorporating our recognizer into it, which is hoped to suggest even better segmentations.

Furthermore, we recently developed a software tool called "Learning Assistant", which gives a graphical interface to study the result of combining various feature space transformation and learning methods. The software already includes several linear and non-linear transformation algorithms and learning techniques, and is continuously being extended.

VIII. HUMAN RESOURCES

Currently we have 3 researchers, 3 Phd students and 2 programmers who work on the speech project. We enlist their names and expertise below. Besides them there are about 6-7 graduate students who take part in the software development.

Prof. János Csirik, head of the group: project management, pattern matching algorithms

Dr. Tibor Gyimóthy, deputy head: project management, machine learning algorithms

András Kocsor, research assistant: feature transformations, pattern classification

Lászl Tóth, research assistant: signal processing, feature extraction

Kornél Kovács, PhD student: feature transformations, machine learning

Gábor Gosztolya, research assistant: machine learning

László Felföldi, PhD student: development of the modular environment of "OASIS"

Dénes Paczolay, PhD student: machine learning in speech impediment therapy

András Bíró, programmer: development of the "Segmentation Assistant"

REFERENCES

[1] Kocsor, A., Tóth, L., Bálint I.,: On the Optimal Parameters of a Sinusoidal Representation of Signals, Conference of PhD students on Computer Sciences, Volume of Extended Abstracts, 12-18th July, Szeged, Hungary, pp. 61-63, 1998.

[2] Tóth, L.: Computer, Can You Hear Me?; Új Alaplap magazin, Vol. XVI. No. 6., p. 5-7. , June, 1998. (in Hungarian)

[3] Tóth, L.: Automatic Speech Recognition, (a chapter in the book "Artificial Intelligence", edited by: Futó Iván), Aula, pp. 815-872, 1999. (in Hungarian)

[4] Kocsor, A., Tóth, L., Bálint I.: On the Optimal Parameters of a Sinusoidal Representation of Signals, Acta Cybernetica 14, pp. 315-330, 1999.

[5] Kocsor, A., Kuba A. Jr., Tóth L., Jelasity M., Felföldi L., Gyimóthy T., Csirik J.,: A Segment-Based Statistical Speech Recognition System for Isolated/Continuous Number Recognition, Proceedings of the FUSST'99, Aug. 19-21, Sagadi,Estonia, pp. 201-211, 1999.

[6] Kocsor, A., Kuba A. Jr., Tóth L.: An Overview of the OASIS speech recognition project, Proceedings of the 4th International Conference on Applied Informatics, Aug. 30-Sept. 3, Eger-Noszvaj, Hungary, pp. 94-102, 1999.

[7] Kocsor, A., Tóth, L., Kuba, A. Jr., Kovács, K., Jelasity, M., Gyimothy, T., Csirik, J.: A Comparative Study of Several Feature Space Transformation and Learning Methods for Phoneme Classification, International Journal of Speech Technology, Vol. 3, Number 3/4, pp. 263-276, 2000.

[8] Tóth, L., Kocsor, A., Kovács, K.: A Discriminative Segmental Speech Model and its Application to Hungarian Number Recognition, Third International Workshop on Text, Speech and Dialogue, Brno, Czech Republic, 13-16 September, LNAI 1902, PP. 307-313, Springer-Verlag, 2000.

[9] Kocsor, A., Tóth, L., Kuba, A. Jr.: Phoneme Classification Using Kernel-PCA, accepted for publication in Periodica Polytechnica

[10] Kocsor, A., Tóth, L., Felföldi, L.: Application of Feature Transformation and Learning Methods in Phoneme Classification, in: L. Monostori, J. Váncza and M. Ali (Eds.): Proc. 14th Int. Conf. On Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, IEA/AIE 2001, LNAI 2070, pp. 502-512, Springer-Verlag, 2001.

[11] Kocsor, A., Tóth, L., Paczolay, D.: A Nonlinearized Discriminant Analysis and its Application to Speech Impediment Therapy, in: Proceedings of the 4th Int. Conf. On Text, Speech and Dialogue, LNAI 2166, pp. 249-257, Springer, 2001.

[12] Kocsor, A., Csirik, J.: Fast Independent Component Analysis in Kernel Feature Spaces, accepted for the SOFSEM 2001 Conference, LNCS/LNAI, 2001.

THINNING ALGORITHMS FOR 3D BINARY PICTURES

Kálmán Palágyi and Attila Kuba

Thinning is an iterative object reduction technique that provides an approximation to the skeleton in a topology preserving way. It means that the result of the thinning process is topologically equivalent to the original object. Our attention is focused on developing 3D thinning algorithm that are to be used in medical applications. This report describes the skeleton, the major skeletonization techniques in the case of discrete objects, the thinning methodologies, and our results.

I. SKELETON AS A SHAPE FEATURE

Shape is a fundamental concept in computer vision. It can be regarded as the basis for high–level image processing stages concentrating on scene analysis and interpretation. There are basically two different approaches for describing the shape of an object:

- using the boundary that surrounds it and
- using the occupied region.

Boundary–based techniques are widely used but there are some deficiencies which limit their usefulness in practical applications especially in 3D. Therefore, the importance of the region–based shape features show upward tendency. The local object symmetries represented by the skeleton certainly cannot replace boundary–based shape descriptors, but complement and support them.

The skeleton is a region–based shape feature that has been proposed by Blum as the result of the Medial Axis Transform. A very illustrative definition of the skeleton is given using the prairie–fire analogy: the object boundary is set on fire and the skeleton is formed by the loci where the fire fronts meet and quench each others. The formal definition of the skeleton has been stated by Calabi: the skeleton of an object is the locus of the centers of all the maximal inscribed hyperspheres. The continuous skeleton of a solid 3D box is illustrated in Fig. 1.

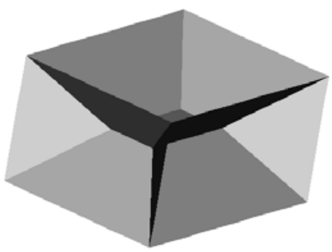


Figure 1. Example of 3D skeleton. The original object was a solid box.

II. SKELETONIZATION TECHNIQUES IN DISCRETE SPACES

During the last two decades skeletonization in the digital image raster has been an important research field. There are two major requirements to be complied with. The first one is geometrical. It means that the "skeleton" must be in the "middle" of the object and invariant under geometrical transformations. The second one is topological requiring that the "skeleton" must be topologically equivalent to the original object.

There are three major discrete skeletonization methods:

- based on distance transformation,
- thinning, and
- based on Voronoi–diagram.

We prefer thinning, since it:

- preserves topology,
- makes easy implementation possible (as a sequence of local Boolean operations),
- takes the least computational costs, and
- can be executed in parallel.

III. THINNING METHODOLOGIES

A 3D binary picture is a mapping that assigns value of 0 or 1 to each point with integer coordinates in the 3D digital space denoted by $\mathbb{Z}^3$. Points having the value of 1 are called black points, while 0's are called white ones. Black points form objects of the picture. White points form the background and the cavities of the picture. Both the input and the output of a picture operation are pictures. An operation is reduction if it can delete some black points (i.e., changes them to white) but white points remain the same. There is a fairly general agreement that a reduction operation is *not* topology preserving if any object in the input picture is split (into two or more ones) or completely deleted, if any cavity in the input picture is merged with the background or another hole, or if a cavity is created where there was none in the input picture. There is an additional concept called hole in 3D pictures. A hole (that doughnuts have) is formed by 0's, but it is not a cavity. Topology preservation implies that eliminating or creating any hole is not allowed.

Thinning must be a topology–preserving reduction. Existing 3D thinning algorithms can be classified from several points of view. One of them is the classification on the produced skeletons: some of the developed algorithms result in medial surfaces and others can produce medial lines (see Fig. 2).

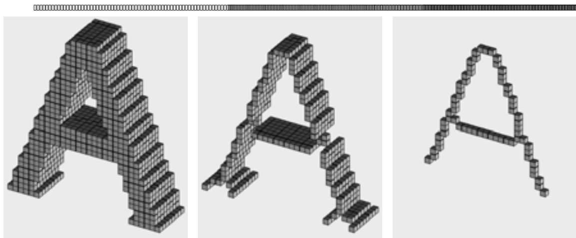


Figure 2. A 3D synthetic picture containing a character "A" (left); its medial surface (centre); its medial lines (right).

Three types of parallel thinning methodologies have been proposed.

The first type examine the $3 \times 3 \times 3$ neighbourhood of each border point (i.e., black point that has at least one white point in its neighbourhood). The iteration steps are divided into a number of subiterations.

Only border points having the prescribed $3 \times 3 \times 3$ neighbourhood can be deleted during a subiteration. It means that each prescribed neighbourhood gives a deletion condition. Prescribed neighbourhoods can be usually associated to a direction (e.g., up, down) depending on the position of the border points to be deleted. These algorithms use directional or border sequential strategy. Each subiteration is executed in parallel (i.e., all black points satisfying the deletion condition of the actual subiteration are simultaneously deleted). Most of existing parallel thinning algorithms are border sequential. Generally 6 subiterations are used.

A directional algorithm consisting of k subiterations follows the following process:

repeat
  for $i = 1$ **to** k **do**
    simultaneous deletion of the black points that satisfy the condition assigned to the i–th direction
until *no points are deleted*

We have developed three different border–sequential 3D thinning algorithms. They use 3, 6, 8, and 12 subiterations, respectively [1, 2, 3, 4].

The second type of algorithms does not need subiterations but — in order to preserve topology — it investigates larger $(5 \times 5 \times 5)$ neighbourhood. This approach can be sketched by the following program:

repeat
  simultaneous deletion of the black points that satisfy the global condition
until *no points are deleted*

The third approach is the subfield sequential method. The set of points $\mathbb{Z}^3$ is divided into more disjoint subsets which are alternatively activated. At a given iteration step, only black points of the active subfield are designated to be deleted.

A subfield based algorithm consisting of k subfields can be described as follows:

repeat
  for $i = 1$ **to** k **do**
    simultaneous deletion of the black points in the i–th subfield that satisfy the global condition assigned to each subfield
until *no points are deleted*

We have proposed a new class of thinning methods called hybrid algorithms. It applies both directional and subfield strategies [5].

IV. APPLICATIONS

Thinning provides "skeleton–like" shape feature that are extracted from binary image data. It is a common preprocessing operation in raster–to–vector conversion or in pattern recognition. Its goal is to reduce the volume of elongated objects. Some important applications have been appeared in medical image processing (see Fig. 3), too [6, 7, 8, 9].

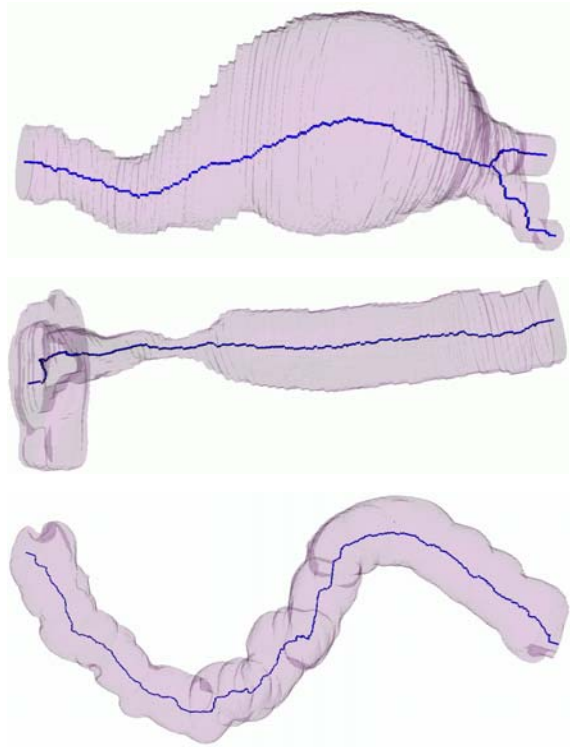


Figure 3. Central paths for a segmented part of an infrarenal aorta (top), a segmented upper respiratory tract (middle), and a segmented cadaveric phantom (bottom) produced by our curve-thinning algorithm.

REFERENCES

[1] K. Palágyi, A 3-subiteration 3D thinning algorithm for extracting medial surfaces, Pattern Recognition Letters, to appear

[2] K. Palágyi, A. Kuba, A 3D 6–subiteration thinning algorithm for extracting medial lines, Pattern Recognition Letters 19, 1998, 613-627.

[3] K. Palágyi, A. Kuba, A parallel 8–subiteration 3D thinning algorithm, in Proc. 8th Int. Conf. on Discrete Geometry for Computer Imagery, DGCI'99, Marne-la-Vallé, France, Lecture Notes in Computer Science 1568, Springer, 1999, 325-336.

[4] K. Palágyi, A. Kuba, A parallel 3D 12-subiteration thinning algorithm, Graphical Models and Image Processing 61, 1999, 199-221.

[5] K. Palágyi, A. Kuba, A hybrid thinning algorithm for 3D medical images, J. Computing and Information Technology 6, 1998, 149-164.

[6] K. Palágyi, E. Sorantin, Cs. Halmai, A. Kuba:, 3D thinning and its applications to medical image processing, TASK Quarterly 3, 1999, 397-408.

[7] K. Palágyi, E. Sorantin, E. Balogh, A. Kuba, Cs. Halmai, B. Erdőhelyi, K. Hausegger, A sequential 3D thinning algorithm and its medical applications, in Proc. 17th Int. Conf. Information Processing in Medical Imaging, IPMI 2001, Davis, USA, Lecture Notes in Computer Science 2082, Springer, 2001, 409-415.

[8] E. Sorantin, Cs. Halmai, B. Erdőhelyi, L. Martonossy, K. Palágyi, B. Geiger, New advances for imaging of laryngotracheal stenosis by post processing of spiral-CT data, in W. Hruby (ed.) Digital Revolution in Radiology, Springer, 2000, pp. 275-285.

[9] E. Sorantin, E. Balogh, A. Vilanova i Bartroli, K. PalPalágyi, L.G. Nyúl, Virtual dissection of the colon based on helical CT data - Can it be done? in Proc. 2nd Int. Symposium on Image and Signal Processing and Analysis, ISPA 2001, Pula, Croatia, 2001, 224-229.

MEDICAL IMAGE REGISTRATION

Attila Tanács, Kálmán Palágyi, Attila Kuba

Image registration is a fundamental task in digital image processing used to match two independently acquired images. These images are taken at different times, from different viewpoints or even from different imaging devices. To register images, the geometrical relationship between them is to be determined. Image registration is an important area of medical image processing. Matching all the geometric data available for a patient provides better diagnostic capability, better understanding of data, and improves surgical and therapy planning and evaluation. This report describes the registration problem, the phases of the process, the proposed methods, and our results.

I. REGISTRATION PROBLEM

There is an increasing number of medical applications that require accurate aligning of one medical image with another. The geometrical transformation is to be found that maps a *reslice image data set* in precise spatial correspondence with a *base image data set*. This process of alignment is known as *registration*, although other words, such as *co–registration*, *matching* and *fusion*, also are used (see Fig. 1).

One may distinguish between *multimodality* and *unimodality registration*. Multimodality registration is used to match 2D and 3D anatomical images with each other and/or with 3D functional images. Aligning the complementary information content of the images from different modalities means a great benefit during diagnosis and treatment planning. Besides multimodality registration, important applications exist in unimodality registration. It is useful to combine images in time–series or even to align images with different protocol settings (e.g., with and without contrast injection). It is commonly used to show changes between images over a period of time. The progress of a disease can be tracked or the efficiency of the applied treatments can be followed.

II. REGISTRATION TECHNIQUES

Registration techniques depend on image content; thus no general methodology has been proposed. A variety of registration methodologies have been developed. Registration methods can be viewed as different combinations of choices for the following four components:

- *search space*
  what kind of transform we have to consider, i.e., what is the class of transformations that is capable of aligning the images;
- *feature space*
  what to match, i.e., what are the features to be used in matching;
- *similarity measure*
  how to evaluate the match;
- *search strategy*
  how to match, i.e., the process for achieving a consistent match.

The registration process consists of the following four steps:

- *Feature extraction*
  Feature data sets F_1 and F_2 are to be extracted from the base image I_1 and the reslice image I_2, respectively;
- *Finding a consistent transformation*
  Transformation $\mathcal{T}$ is to be determined, where $\mathcal{S}(F_1, \mathcal{T}(F_2))$ is maximized or minimized in the sense of the selected similarity measure $\mathcal{S}$;
- *Applying the found geometrical transformation*
  Resampled image $\mathcal{T}(I_2)$ is to be calculated;
- *Using registered data sets*
  Registered image data sets I_1 and $\mathcal{T}(I_2)$ may be used in various ways (e.g., I_1 is displayed overlaid on top of $\mathcal{T}(I_2)$).

III. REGISTRATION BASED ON FUZZY OBJECTS

We have developed an automated registration method that is used effectively in Multiple Sclerosis projects of Medical Image Processing Group, Department of Radiology, University of Pennsylvania, Philadelphia. It works within the 3DVIEWNIX software system.

Our registration method is based on fuzzy objects used as features. Approaches to object information extraction from images should attempt to use the fact that images are by nature fuzzy. The notion of “hanging togetherness” of image elements specified by their fuzzy connectedness is to be considered in image segmentation research.

Our approach is described by giving the four components:

- *search space:* rigid body transformation parameterized by 6 degrees of freedom;
- *feature space:* fuzzy objects (or raw image data);
- *similarity measure:* correlation;
- *search strategy:* multi-resolution pyramid.

The major advantage of the method compared to direct correlation of image intensities is that it computes new images representing “objectness” in which regions to be matched are likely to have similar intensities, although this may not be true for the original images.

IV. REGISTRATION BASED ON POINT LANDMARKS

We have developed a method based on interactively identified anatomical landmark points for multimodality and unimodality registration [1].

The applied method can be described by the following components:

- *search space:* 4×4 general transformation matrix capable of giving arbitrary affine transformations;
- *feature space:* anatomical landmark points;
- *similarity measure:* sum of distances;
- *search strategy:* least square minimising of the distance function.

Our method is capable of finding the optimal global affine transformation that matches two sets of points in arbitrary dimensions. We have managed to give and prove a sufficient existence condition for a unique transformation [3, 4].

Two additional methods are also implemented for searching rigid body transform and nonlinear warping based on thin–plate spline interpolation.

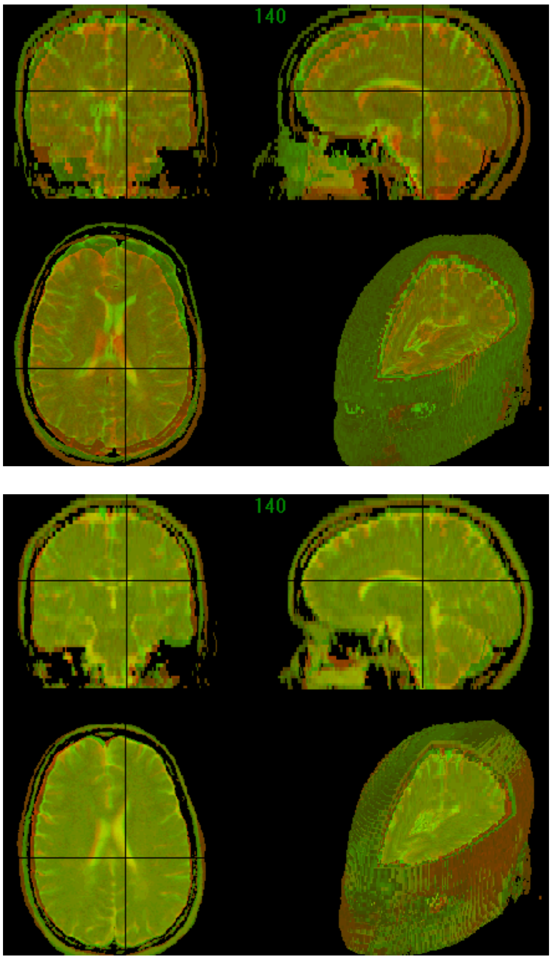


Figure 1. Example of 3D image fusion. Overlaid MR coronal, sagittal, transversal 2D slices and a rendered 3D view before registration (top) and after registration (bottom).

We have investigated the registration error for point-based registration [2]. Numerical simulations were performed for four kinds of registration methods assuming different search spaces. We showed that the (target) registration error depends not only on the number of homologous points, and on the accuracy of point localization, but on the volume spanned by the fiducials as well.

V. REGISTRATION BASED ON MUTUAL INFORMATION

Our fully automated registration method utilizes mutual information as similarity measure. The CT or PET image is resampled to the voxel size of the MR image. The intensities are scaled to [0,127] for both images, then a multiresolution pyramid is built from the 2D slices. A variant of Powell's multidimensional minimization technique is used to find the transformation. At lower resolution nearest neighbour, at the final level bilinear iterpolation between slices is used.

Our method has been validated within the Retrospective Registration Evaluation Project (National Institutes of Health, Project Number 1 R01 NS33926-01, Principal Investigator, J.M. Fitzpatrick, Vanderbilt University, Nashville, TN). Results can be found at

`www.vuse.vanderbilt.edu/~jayw/results.html`

More details are to be read at

`www.inf.u-szeged.hu/~tanacs/regist/regidx.html`

ACKNOWLEDGEMENTS

Our research was supported by National Science Foundation and the Hungarian Academy of Sciences Grant INT-9121281 and by OTKA Grant T023804.

REFERENCES

[1] A. Tanács, K. Palágyi, A. Kuba, Medical image registration based on interactively defined anatomical landmark points, Machine Graphics & Vision 7, 151–158, 1998.

[2] A. Tanács, K. Palágyi, A. Kuba, Target registration error of point-based methods assuming rigid-body and linear motions, in Proc. Int. Workshop on Biomedical Image Registration WBIR'99, Bled, Slovenia, 1999, 223-233.

[3] A. Tanács, G. Czédli, K. Palágyi, A. Kuba, Point-based registration assuming affine motion, in Proc. Int. Workshop Algebraic Frames for the Perception-Action Cycle, AFPAC 2000, Kiel, Germany, Lecture Notes in Computer Science 1888, Springer, 2000, 329-338.

[4] A. Tanács, G. Czédli,K. Palágyi, A. Kuba, Affine matching of two sets of points in arbitrary dimensions, Acta Cybernetica 15, 2001, 101-106.

TREE AUTOMATA AND TERM REWRITE SYSTEMS

Sándor Vágvölgyi

I. TERM REWRITE SYSTEMS EFFECTIVELY PRESERVING RECOGNIZABILITY

Tree automata and recognizable tree languages proved to be an efficient tool in the theory of term rewrite systems. Let Σ be a ranked alphabet, let R be a term rewrite system over Σ, and let L be a tree language over Σ. Then $R_\Sigma^*(L)$ is the set of descendants of trees in L. A term rewrite system R over Σ preserves Σ-recognizability, if for each recognizable tree language L over Σ, $R_\Sigma^*(L)$ is recognizable. The signature $sign(R)$ of a term rewrite system R is the ranked alphabet consisting of all symbols appearing in the rules of R.

Let R be a term rewrite system over a ranked alphabet Σ. We say that R effectively preserves Σ-recognizability if for a given tree automaton $\mathcal{B}$ over Σ, we can effectively construct a tree automaton $\mathcal{C}$ over Σ such that $L(\mathcal{C}) = R_\Sigma^*(L(\mathcal{B}))$. Let R be a term rewrite system over a ranked alphabet Δ. We say that R effectively preserves recognizability if for a given ranked alphabet Σ with $sign(R) \subseteq \Sigma$ and a given tree automaton $\mathcal{B}$ over Σ, we can effectively construct a tree automaton $\mathcal{C}$ over Σ such that $L(\mathcal{C}) = R_\Sigma^*(L(\mathcal{B}))$. Let R be a term rewrite system over $sign(R)$, and let $\Sigma = \{ f, \sharp \} \cup sign(R)$, where $f \in \Sigma_2 - sign(R)$ and $\sharp \in \Sigma_0 - sign(R)$. We [3] showed that R effectively preserves Σ-recognizability if and only if R effectively preserves recognizability.

A term rewrite system R is gsm (generalized semi-monadic) if there is no rule $l \to r$ in R with $l \in X$ and the following holds. For any rules $l_1 \to r_1$ and $l_2 \to r_2$ in R, for any occurrences $\alpha \in O(r_1)$ and $\beta \in O(l_2)$, and for any supertree $l_3 \in T_\Sigma(X)$ of l_2/β with $var(l_3) \cap var(l_1) = \emptyset$, if

(i) $\alpha = \lambda$ or $\beta = \lambda$,

(ii) r_1/α and l_3 are unifiable, and

(iii) σ is a most general unifier of r_1/α and l_3,

then

(a) $l_2/\beta \in X$ or

(b) for each $\gamma \in O(l_3)$, if $l_2/\beta\gamma \in X$, then $\sigma(l_3/\gamma) \in X \cup T_\Sigma$.

We [3] showed that a linear gsm (lgsm) term rewrite system R over Δ effectively preserves recognizability in the following way. Let L be a recognizable tree language over Σ with $\Delta \subseteq \Sigma$, and let $\mathcal{B} = (\Sigma, B, R_\mathcal{B}, B')$ be a tree automaton recognizing L. Similarly to the constructions of Salomaa and Coquidé et al., we constructed a sequence of bottom-up tree automata $\mathcal{C}_i = (\Sigma, C, R_i, B')$, $i \geq 0$ having the same ranked alphabet, state set, and final state set. The rule set R_0 contains $R_\mathcal{B}$. Moreover, R_0 contains rules which enable R_0 to recognize the right-hand sides of rules in R. For each $i \geq 0$, R_{i+1} contains R_i, and for each rule $l \to r$ in R, $\mathcal{C}_{i+1}$ simulates, on the right-hand side r, the computation of $\mathcal{C}_i$ on the left-hand side l. There is a least integer $M \geq 0$ such that $R_M = R_{M+1}$. Hence $\mathcal{C}_M = \mathcal{C}_{M+1}$. We [3] showed that $L(\mathcal{C}_M) = R_\Sigma^*(L)$.

Brainerd, Kozen, and Fülöp and Vágvölgyi showed that a tree language L is recognizable if and only if there exists a ground term rewrite system R such that L is the union of finitely many $\leftrightarrow_R^*$-classes. We [3] showed that a tree language L is recognizable if and only if there exists a term rewrite system R such that $R \cup R^{-1}$ is an lgsm term rewrite system and that L is the union of finitely many $\leftrightarrow_R^*$-classes.

We [3] showed the following decidability results.

(1) Let R_1, R_2 be term rewrite systems. Let R_1 effectively preserve recognizability. Then it is decidable if $\to_{R_2}^* \subseteq \to_{R_1}^*$.

(2) Let R_1 and R_2 be term rewrite systems effectively preserving recognizability. Then it is decidable which one of the following four mutually excluding conditions hold.

(i) $\to_{R_1}^* \subset \to_{R_2}^*$,

(ii) $\to_{R_2}^* \subset \to_{R_1}^*$,

(iii) $\to_{R_1}^* = \to_{R_2}^*$,

(iv) $\to_{R_1}^* \bowtie \to_{R_2}^*$,

where "$\bowtie$" stands for the incomparability relationship.

(3) For an lgsm term rewrite system R, it is decidable whether R is left-to-right minimal. (A term rewrite system R is left-to-right minimal if for each rule $l \to r$ in R, $\to_{R-\{ l \to r \}}^* \subset \to_R^*$.)

(4) Let R_1 and R_2 be term rewrite systems such that $R_1 \cup R_1^{-1}$ and $R_2 \cup R_2^{-1}$ are term rewrite systems and effectively preserve recognizability. Then it is decidable which one of the following four mutually excluding conditions holds.

(i) $\leftrightarrow_{R_1}^* \subset \leftrightarrow_{R_2}^*$,

(ii) $\leftrightarrow_{R_2}^* \subset \leftrightarrow_{R_1}^*$,

(iii) $\leftrightarrow_{R_1}^* = \leftrightarrow_{R_2}^*$,

(iv) $\leftrightarrow_{R_1}^* \bowtie \leftrightarrow_{R_2}^*$.

(5) Let R be a term rewrite system such that $R \cup R^{-1}$ is an lgsm term rewrite system. Then it is decidable whether R is two-way minimal. (A term rewrite system R is two-way minimal if for each rule $l \to r$ in R, $\leftrightarrow_{R-\{ l \to r \}}^* \subset \leftrightarrow_R^*$.)

(6) Let R_1, R_2 be term rewrite systems over a ranked alphabet Σ. Let R_1 effectively preserve recognizability. Let $g \in \Sigma - \Sigma_0$ be such that g does not occur on the left-hand side of any rule in R_1, and let $\sharp \in \Sigma_0$ be irreducible for R_1. Then it is decidable if $\to_{R_2}^* \cap (T_\Sigma \times T_\Sigma) \subseteq \to_{R_1}^* \cap (T_\Sigma \times T_\Sigma)$.

(7) Let R_1 and R_2 be term rewrite systems over Σ effectively preserving recognizability. Moreover, let $g_1, g_2 \in \Sigma - \Sigma_0$ be such that for each $i \in \{ 1, 2 \}$, g_i does not occur on the left-hand side of any rule in R_i. Let $\sharp_1, \sharp_2 \in \Sigma_0$ be such that for each $i \in \{ 1, 2 \}$, $\sharp_i$ is irreducible for R_i. Then it is decidable which one of the following four mutually excluding conditions hold.

(i) $\to_{R_1}^* \cap (T_\Sigma \times T_\Sigma) \subset \to_{R_2}^* \cap (T_\Sigma \times T_\Sigma)$,

(ii) $\to_{R_2}^* \cap (T_\Sigma \times T_\Sigma) \subset \to_{R_1}^* \cap (T_\Sigma \times T_\Sigma)$,

(iii) $\to_{R_1}^* \cap (T_\Sigma \times T_\Sigma) = \to_{R_2}^* \cap (T_\Sigma \times T_\Sigma)$,

(iv) $\to_{R_1}^* \cap (T_\Sigma \times T_\Sigma) \bowtie \to_{R_2}^* \cap (T_\Sigma \times T_\Sigma)$.

(8) Let R be an lgsm term rewrite system over Σ. Moreover, let $g \in \Sigma - \Sigma_0$ be such that g does not occur on the left-hand side of any rule in R, and let $\sharp \in \Sigma_0$ be irreducible for R. Then it is decidable whether R is left-to-right ground minimal. (A term rewrite system R over Σ is left-to-right ground minimal if for each rule $l \to r$ in R, $\to_{R-\{ l \to r \}}^* \cap (T_\Sigma \times T_\Sigma) \subset \to_R^* \cap (T_\Sigma \times T_\Sigma)$.)

(9) Let R be a term rewrite system over Σ such that $R \cup R^{-1}$ is an lgsm term rewrite system. Moreover, let

$g \in \Sigma - \Sigma_0$ be such that g does not occur in any rule of R, and let $\sharp \in \Sigma_0$ be irreducible for $R \cup R^{-1}$. Then it is decidable whether R is two-way ground minimal. (A term rewrite system R over Σ is two-way ground minimal if for each rule $l \to r$ in R, $\leftrightarrow^*_{R-\{\,l\to r\,\}} \cap (T_\Sigma \times T_\Sigma) \subset \leftrightarrow^*_R \cap (T_\Sigma \times T_\Sigma)$.)

(10) Let R be a term rewrite system over Σ effectively preserving recognizability, and let $p, q \in T_\Sigma(X)$. Then it is decidable if there exists a tree $r \in T_\Sigma(X)$ such that $p \to^*_R r$ and $q \to^*_R r$.

(11) Let R be a term rewrite system over Σ effectively preserving recognizability. Then it is decidable if R is locally confluent.

Let R be a term rewrite system over $sign(R)$, and let $\Sigma = \{\, f, \sharp \,\} \cup sign(R)$, where $f \in \Sigma_2 - sign(R)$ and $\sharp \in \Sigma_0 - sign(R)$. We [3] showed that R effectively preserves Σ-recognizability if and only if R effectively preserves recognizability. We [4] improved this result for left-linear term rewrite systems. We showed the following. Let R be a left-linear term rewrite system over $sign(R)$, and let $\Sigma = \{\, g, \sharp \,\} \cup sign(R)$, where $g \in \Sigma_1 - sign(R)$ and $\sharp \in \Sigma_0 - sign(R)$. Then R effectively preserves Σ-recognizability if and only if R effectively preserves recognizability. This result makes it easier to show that a given left-linear term rewrite system effectively preserves recognizability. Furthermore, R preserves Σ-recognizability if and only if R preserves recognizability.

We [6] introduced and studied the half-monadic term rewrite system, which is an extension of the right-ground term rewrite system and is a slight extension of the semi-monadic term rewrite system. Takai, Kaji, and Seki introduced the concept of the finite path overlapping term rewrite system, and showed that right-linear finite path overlapping term rewrite systems effectively preserve recognizability. A half-monadic term rewrite system is a finite path overlapping term rewrite system as well. Hence right-linear half-monadic term rewrite systems effectively preserve recognizability. We [6] showed that termination and convergence are decidable properties for right-linear half-monadic term rewrite systems.

II. GROUND TERM REWRITE SYSTEMS

In paper [5], we showed that a reduced ground term rewrite system over some ranked alphabet Σ has a good property: if we omit any rules, the omitted rules and the remaining rules generate congruence relations such that their intersection is the identity relation on T_Σ. In general, given two ground term rewrite systems A and B over some ranked alphabet Σ with $\leftrightarrow^*_A \subseteq \leftrightarrow^*_B$, one may want to eliminate all nontrivial pairs of $\leftrightarrow^*_A$ from $\leftrightarrow^*_B$. In other words, one may want to construct a ground term rewrite system (gtrs for short) C over Σ such that $\leftrightarrow^*_{A\cup C} = \leftrightarrow^*_B$ and that $\leftrightarrow^*_A \cap \leftrightarrow^*_C$ is the identity relation on T_Σ. Let A and B be gtrs's over some ranked alphabet Σ with $\leftrightarrow^*_A \subseteq \leftrightarrow^*_B$. We say that a gtrs C over Σ is a congruential complement of A for B, if $\leftrightarrow^*_{A\cup C} = \leftrightarrow^*_B$ and $\leftrightarrow^*_A \cap \leftrightarrow^*_C$ is the identity relation over T_Σ. We note that A may have more than one congruential complement for B, which are pairwise nonequivalent. Consider the following example. Let $\Sigma = \{\, a, b, f \,\}$, where a, b are of rank 0, and f is of rank 1. Consider the gtrs's

$$A = \{\, f(a) \to a,\ \ f(b) \to b \,\}\ ,$$

$$B = \{\, f(a) \to a,\ \ f(b) \to b,\ \ a \to b \,\}\ ,$$

$$C = \{\, a \to b \,\}\ ,$$

$$D = \{\, a \to f(b) \,\}\ ,$$

$$E = \{\, f(a) \to f^3(b) \,\}$$

over Σ. Then each one of C, D, and E is a congruential complement of A for B. Moreover, $\leftrightarrow^*_C \neq \leftrightarrow^*_D$, $\leftrightarrow^*_C \neq \leftrightarrow^*_E$, and $\leftrightarrow^*_D \neq \leftrightarrow^*_E$.

Our [5] main result is the following. Given gtrs's A, B, C over some ranked alphabet Σ with $\leftrightarrow^*_A \subseteq \leftrightarrow^*_B$, one can effectively decide if C is a congruential complement of A for B.

In paper [1] we studied that version of the ground tree transducer game where the same tree automaton appears as the first and second component of the associated ground tree transducer. We gave conditions which imply that Beta has a winning strategy. Furthermore, we showed the following decidability result. Given a ground tree transducer game where the underlying tree automaton A cannot evaluate some tree into a state or A is deterministic, we can decide which player has a winning strategy. Moreover, whatever player has a winning strategy, we can effectively construct a partial recursive winning strategy for him.

ACKNOWLEDGEMENTS

The mentioned research tasks were supported by the following grants: MKM 223/95, FKFP 0095/1997 of the Hungarian Cultural and Educational Ministry, OTKA F012852, OTKA T022994, OTKA T030084 of the Research Foundation of Hungary, and the Bolyai János Research Scholarship of the Hungarian Academy of Sciences.

REFERENCES

[1] J. Apró and S. Vágvölgyi, The ground tree transducer game with identical tree automata, *Fundamenta Informaticae*, **45** (2001) 157-172.

[2] Z. Fülöp and S. Vágvölgyi, Restricted Ground Tree Transducers, *Theoretical Computer Science*, **250** (2001) 219-233.

[3] P. Gyenizse and S. Vágvölgyi, Linear Generalized Semi-monadic Rewrite Systems Effectively Preserve Recognizability, *Theoretical Computer Science*, **194** (1998) 87-122.

[4] P. Gyenizse, S. Vágvölgyi, A property of left-linear rewrite systems preserving recognizability, *Theoretical Computer Science*, **242** (2000) 477-498.

[5] S. Vágvölgyi, Congruential Complements of Ground Term Rewrite Systems, *Theoretical Computer Science*, **238** (2000) 247-274.

[6] S. Vágvölgyi, Right-Linear Half-Monadic Term Rewrite Systems, manuscript.