Czech Technical University in Prague
Faculty of Electrical Engineering

České vysoké učení technické v Praze
Fakulta elektrotechnická

Ing. Tomáš Werner, Ph.D.

Linear Programming Approach to Discrete Energy Minimization

Minimalizace diskrétních energetických funkcí pomocí lineárního programování

Summary

Graphical models combine graph theory and probability theory into a general formalism to model interactions of multiple variables. The formalism has many applications in computer vision, machine learning, artificial intelligence, and other disciplines. A basic operation with a graphical model is inference, which requires computation of either the maximum or the marginals of the probability distribution defined by the model. The former problem is a combinatorial optimization one, often referred to as discrete energy minimization or valued/weighted constraint satisfaction.

In this document I first define the problem of discrete energy minimization and mention the relation of this interdisciplinary problem to different research disciplines. Then I summarize my contributions to one successful approach to this NP-hard problem, linear programming (LP) relaxation.

My first contribution was a revisit of an old and not widely known approach by Schlesinger et al., which was independently rediscovered recently and is a basis of the LP relaxation approach. My second contribution is an elegant generalization of this approach, which enables handling energy terms of arbitrary arity, constructing a hierarchy of progressively tighter relaxations, and incrementally tightening the relaxation in a cutting plane fashion. This comes with a very simple yet general algorithm, generalized min-sum diffusion, that computes a (coordinate-wise local) optimum of the dual LP relaxation for large instances. My third contribution is to show that trying to find a very efficient algorithm to solve the LP relaxation of discrete energy minimization is hopeless because the general linear programming problem can be reduced in linear time to this LP relaxation.

Souhrn

Grafové modely kombinují teorii grafů a teorii pravděpodobnosti do obecného formalismu vhodného k modelování interakcí množiny mnoha proměnných. Mají mnoho použití v počítačovém vidění, strojovém učení, umělé inteligenci, a jiných disciplínách. Základní operací v grafových modelech je inference, která vyžaduje výpočet buď maxima nebo marginálů rozdělení pravděpodobnosti definovaného modelem. První z těchto problémů je často nazýván minimalizace diskrétní energetické funkce nebo programování s váženými omezeními.

V tomto dokumentu nejdříve definuji minimalizaci diskrétní energetické funkce a zmíním vztahy tohoto multidisciplinárního problému k různým výzkumným oborům. Poté popíši mé výzkumné příspěvky k úspěšnému přístupu k tomuto NP-těžkému problému, založeném na relaxaci pomocí lineárního programování (LP).

Můj první příspěvek byla rekapitulace starého a nepříliš známého přístupu Schlesingera aj., který byl nedávno nezávisle znovu objeven and tvoří základ přístupu založeném na LP relaxaci. Druhý příspěvek je elegantní zobecnění tohoto přístupu, které umožňuje použít sčítance libovolné arity, sestavit hierarchii stále těsnějších relaxací problému, a inkrementálně zlepšovat relaxaci podobně jako v metodách řezných nadrovin. Částí příspěvku je velmi jednoduchý ale obecný algoritmus, zobecněná min-sum difúze, který počítá (po souřadnicích lokální) optimum duální LP relaxace. Třetí příspěvek ukazuje, že snaha najít opravdu efektivní algoritmus na řešení LP relaxace minimalizace disktrétní energie je odsouzena k nezdaru, neboť obecný problém lineárního programování se redukuje v lineárním čase na tuto LP relaxaci.

Keywords

Graphical model, Markov random field, MAP inference, discrete energy minimization, valued constraint satisfaction, combinatorial optimization, linear programming relaxation, local marginal polytope.

Klíčová slova

Grafový model, Markovské náhodné pole, MAP inference, minimalizace diskrétní energetické funkce, problém s váženými omezujícími podmínkami, kombinatorická optimalizace, LP relaxace, lokální marginální polytop.

Obsah

Kapitola 1

Discrete Energy Minimization

By *discrete energy minimization* we understand the following problem:

> *Given a set of discrete (i.e., with finite domains) variables and a set of functions each depending on a (usually small) subset of the variables, minimize the sum of the functions over the variables.*

Formally, let V be a finite set of variables, where each variable $i \in V$ attains states x_i from a finite domain X. Let $H \subseteq 2^V$ be a collection of variable subsets, i.e., a hypergraph over V. Let for each hyperedge $A \in H$ be given a function $f_A \colon X^A \to \mathbb{R} \cup \{\infty\}$. We want to minimize the 'energy function'

$$E(x_V) = \sum_{A \in H} f_A(x_A) \tag{1.1}$$

where $x_A \in X^A$ denotes an assignment to (labeling of) variables $A \subseteq V$. E.g., for $V = (1,2,3,4)$ and $H = \{(1,2),(1,3),(2,3,4)\}$ we have

$$E(x_1,x_2,x_3,x_4) = f_{12}(x_1,x_2) + f_{13}(x_1,x_3) + f_{234}(x_2,x_3,x_4).$$

The problem is very multidisciplinary, it has been studied in several disciplines under many different names:

- In *machine learning*, the problem has been studied in a wider framework of *graphical models*, which combine graph theory and probability theory into a general formalism to model interactions of a large set of variables (Lauritzen, 1996; Bishop, 2006; Wainwright and Jordan, 2008). They do so by modeling the joint distribution of the involved variables as

$$p(x_V) \propto \mathrm{e}^{-E(x_V)}. \tag{1.2}$$

Inference in a graphical model seeks to compute the states of a subset of variables V (*hidden* variables) from the states of the remaining (*observed* variables). Following the Bayesian decision theory, this is formalized as minimizing the expectation of a given loss function, which specifies the penalty for incorrect decisions. The two most common loss functions lead to the need to compute either the maximum (*maximum posterior* or *MAP* inference) or the marginals (*maximum posterior marginals* or *MPM* inference) of a distribution of the form (1.2). Obviously, MAP inference is isomorphic to energy minimization.

- In *computer vision*, it was observed that some image processing problems (typically low-level ones such as denoising, segmentation, or dense image matching) can be easily formalized as minimizing a function (1.1) where V is identified with the set of image pixels and H contains individual pixels and small subsets of pixels. Most often, H contains just pairs of neighboring pixels, so that

$$E(x_V) = \sum_{i \in V} f_i(x_i) + \sum_{\{i,j\} \in H} f_{ij}(x_i, x_j). \tag{1.3}$$

  In computer vision, the problem is often referred to as *discrete energy minimization* (Boykov et al., 2001; Szeliski et al., 2008; Kappes et al., 2013). Sometimes, this was understood more widely as MAP inference in Markov random fields (Li, 2009), which is just another name for one class of graphical models.

- *Pattern recognition* is a somewhat obsolete name for what has today differentiated to machine learning and computer vision. Here, our problem has been called a *two-dimensional grammar* or a *min-sum labeling problem* (Shlezinger, 1976; Schlesinger, 1989). It is a soft version of the *consistent labeling problem* (Haralick and Shapiro, 1979, 1992; Rosenfeld et al., 1976), which is the decision problem obtained for the 'crisp' local functions $f_A\colon X^A \to \{0, \infty\}$.

- *Constraint programming* (Rossi et al., 2006) is motivated by the idea that rather than to specify *how* to fulfill the task at hand (by giving a code) it is often easier to specify *what* should be fulfilled (by giving a set of constraints). The core of constraint programming is the *constraint satisfaction problem* (Mackworth, 1991; Freuder and Mackworth, 2006), equivalent to the consistent labeling problem. This (crisp) constraint satisfaction was later extended to various forms of *weighted* constraint satisfaction (Rossi et al., 2006, chapter 9). One of them is isomorphic to discrete energy minimization and has been called the *partial* (Koster et al., 1998), *weighted* (Meseguer et al., 2006) or *valued* (Živný, 2012) constraint satisfaction problem.

- *Statistical physics* (Mézard and Montanari, 2009) wants to understand how macroscopic properties of matter result from random behavior of locally interacting particles. It studies distributions of the form (1.2) (with a temperature parameter added) under the name *(Botzmann-)Gibbs distribution*, where the function (1.1) is the actual energy (Hamiltonian) of the system. A famous example is the explanation of ferromagnetism from interaction of spin orientations by the *Ising model.* The maxima of the Gibbs distribution (i.e., the minima of energy) are known as its *ground states.*

- It is interesting that despite its very natural formulation, discrete energy minimization in its full generality has not been studied in *combinatorial optimization.* However, it is nowadays changing (Živný, 2012; Kolmogorov et al., 2015).

Inference problems in graphical models are intractable. Namely, discrete energy minimization is APX-hard even to approximate and computing marginals is #P-complete. Therefore, one has to recourse to approximate algorithms. Non-existence of good approximate algorithms has long hindered applying the powerful formalism of graphical models in practice. However, in the last decade approximate inference algorithms have undergone a revolution within the fields of machine learning and computer vision. One class of these new successful methods is based on *linear programming (LP) relaxation* (Shlezinger, 1976; Kolmogorov, 2006) and, for more general inference problems, *variational inference* (Wainwright and Jordan, 2008). These approximate problems are solved by highly parallelizable (block-)coordinate ascent algorithms (Sontag et al., 2012), that can handle large instances of, e.g., image sizes.

In the following three chapters I will summarize my contributions to the LP relaxation approach to discrete energy minimization.

Kapitola 2

Review of the LP Relaxation Approach

My interest in energy minimization and graphical models began around the year 2000 when I attended lectures given by Mikhail I. Schlesinger from the Glushkov Institute of Cybernetics in Kiev, Ukraine, who was visiting our department. It turned out that he formulated the LP relaxation approach to discrete energy minimization as early as in 1970's (Shlezinger, 1976). The approach consisted in maximizing a lower bound on the true minimum of (1.3) by linear transformations of f that preserve the function (1.3). This leads to a linear program, which in fact was dual to the LP relaxation of the original problem. Schlesinger and his colleagues proposed algorithms to minimize the upper bound: a very simple algorithm nicknamed 'min-sum diffusion' (Kovalevsky and Koval, approx. 1975) and the algorithm (Koval and Schlesinger, 1976).

I became attracted by the approach and wanted to contribute to it. However, the approach was not known in my field of computer vision. In fact, active knowledge of optimization of a typical computer vision researcher that time rarely reached beyond the Levenberg-Marquardt algorithm. Having worked on multiple view geometry before, I was no exception and had a hard time to understand the approach. Therefore, before attempting any novel contribution, I decided to first write a review (Werner, 2005, 2007) which would summarize the approach in modern terms and relate it to the current state of the art. The topic turned out to be very multifaceted and I found lots of relevant works from quite different fields, using different formalisms and terminologies. Interestingly, it further turned out that a similar approach was independently discovered at about the same time by Wainwright et al. (2005) and Kolmogorov (2006).

I included in this review in particular:

- The formulation of the primal and dual LP relaxation of minimizing the pairwise energy (1.3).
- Its connection with the constraint satisfaction problem.
- The observation that min-sum diffusion not necessarily finds the global maximum of the LP dual, plus counter-example.

- A re-implementation of min-sum diffusion (which was easy) and the algorithm (Koval and Schlesinger, 1976) (very painful).
- A proof that the LP relaxation (in fact, min-sum diffusion suffices) exactly minimizes submodular energy functions.

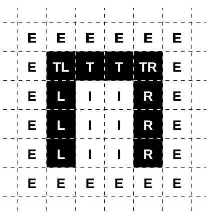
Let me remark that I reformulated these results to some extent, to facilitate possible extension to higher-arity functions, described in the next chapter.

Example As an example, let us model the class of images containing non-overlapping letters ‘Π’ of arbitrary width and height (the rightmost image below). Let V be the set of pixels and $H \subseteq \binom{V}{2}$ the set of pairs of neighboring pixels. Each variable takes states x_i from the set $X = \{\mathsf{E}, \mathsf{I}, \mathsf{T}, \mathsf{L}, \mathsf{R}, \mathsf{TL}, \mathsf{TR}\}$, encoding the syntactic parts of the letter. Let $f_i(y_i \,|\, x_i)$ be the negative log-probability that pixel i has color y_i given its state x_i. Let $f_{ij}(x_i, x_j)$ be 0 if syntactic parts x_i and x_j are ever incident (in the left picture below) and ∞ otherwise. Then (1.2) with

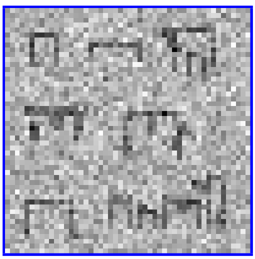
$$E(y_V \,|\, x_V) = \sum_{i \in V} f_i(y_i \,|\, x_i) + \sum_{\{i,j\} \in H} f_{ij}(x_i, x_j) \tag{2.1}$$

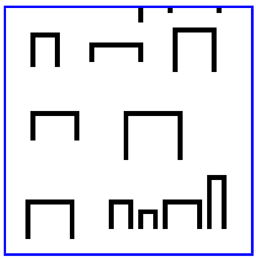
is the probability of an input image y_V (observed variables) given a labeling x_V formed by the syntactic parts (hidden variables). Minimizing (2.1) over x_V yields the most likely image from the class, given the input image y_V.

A local maximum of the dual LP relaxation is found by min-sum diffusion, which iterates a simple local operation (Shlezinger, 1976; Werner, 2007). Despite the problem belongs to none of the known tractable subclasses, min-sum diffusion often finds an exact solution. Moreover, handles very large instances (millions of variables) in reasonable time.



syntactic parts of letter ‘Π’


input image y_V


output image x_V

Kapitola 3

Energy Terms of Arbitrary Arity

The *arity* of an energy term f_A is the number $|A|$ of variables on which it depends. For quite some time, the researchers in energy minimization focused on functions of arity at most two (often called 'binary', 'pairwise', or 'second-order'), i.e., the form (1.3) (Szeliski et al., 2008). Seemingly, this is without any loss of generality because any non-binary function f_A can be represented by combining binary functions. However, representing some high-arity functions needs an exponential number of binary functions. Moreover, this translation destroys symmetry. Therefore, I wanted to handle energy terms of arbitrary arity *natively*, without translating them to binary terms.

First I derived the LP relaxation approach for a very general family of distributions known as the (discrete) exponential family (Wainwright and Jordan, 2008), which contains distributions of type (1.2) but also other distributions. This is described in (Werner, 2009). Then I took a less general view, considering only functions (1.1). This resulted in a very elegant framework with several desirable properties:

- It is applicable to energy terms f_A of arbitrary arity but it is at the same time very simple. The algorithm is a generalization of min-sum diffusion as I formulated it in (Werner, 2007).
- It straightforwardly allows to construct a partially ordered hierarchy of progressively tighter LP relaxations. This is much simpler than the existing method by Wainwright et al. (2005) to construct such a hierarchy by combining (hyper-)trees. The hierarchy is much finer that the well-known hierarchy by Sherali and Adams (1990).
- It is easy to tighten the relaxation *incrementally* during min-sum diffusion, resulting in fact in a (dual) cutting plane algorithm.
- It can be easily proved that the algorithm exactly solves problems with submodular functions of any arity.

The framework is described in (Werner, 2008, 2010)[1]. The hierarchy of LP relaxations, generalized min-sum diffusion, and the cutting plane algorithm are revisited in the book chapter (Franc et al., 2012).

[1]This work has been evaluated by the Czech government as an *excellent result* in the II. pillar of RIV in 2015.

Kapitola 4

Hardness of the LP Relaxation

The LP relaxation of discrete energy minimization is exact for a large class of min-sum instances (e.g., all tractable languages with finite costs (Thapper and Živný, 2012) and instances with bounded treewidth) and it is a basis for designing good approximate algorithms in general (Kappes et al., 2013). It would be therefore of great practical interest to have efficient algorithms to solve the LP relaxation. Indeed, many authors invested considerable effort to develop new efficient algorithms to solve the LP relaxation.

As linear programming is in the PTIME complexity class, one might think that solving the LP relaxation is easy. However, this is not so for large instances that occur, e.g., in computer vision. For energy functions with binary functions f_A and $|X| = 2$ labels, the LP relaxation can be solved efficiently by reduction to max-flow (Boros and Hammer, 2002; Rother et al., 2007). For more general problems, no really efficient algorithm is known. In particular, the well-known simplex and interior point methods are not applicable, if only due to their quadratic space complexity. Coordinate ascent algorithms, such as min-sum diffusion or TRW-S (Kolmogorov, 2006), do apply to large-scale instances but they find only a local optimum of the dual LP relaxation.

My colleague and me have shown (Průša and Werner, 2013, 2015b; Živný et al., 2014) that the quest for very efficient algorithms to solve the LP relaxation is futile, because this task is not easier than solving any linear program. Precisely, every linear program reduces in linear time to the LP relaxation of discrete energy minimization (allowing infinite costs) with $|X| = 3$ labels. From the polyhedral point of view, every polytope is a coordinate-erasing projection of a face of the feasible set of the LP relaxation (known as the *a local marginal polytope*) with 3 labels, whose description can be computed in linear time.

Recently, we proved a similar (somewhat weaker) result for a subclass of the discrete energy minimization problem, the attractive Potts problem (also known as the *uniform metric labeling problem* (Kleinberg and Tardos, 2002; Chekuri et al., 2005)). This is described in (Průša and Werner, 2015a).

Literatura

Christopher M. Bishop. *Pattern Recognition and Machine Learning.* Springer Science+Bussiness Media, New York, NY, 2006.

Endre Boros and Peter L. Hammer. Pseudo-Boolean optimization. *Discrete Applied Mathematics*, 123(1-3):155–225, 2002. ISSN 0166-218X.

Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

Chandra Chekuri, Sanjeev Khanna, Joseph Naor, and Leonid Zosin. A linear programming formulation and approximation algorithms for the metric labeling problem. *SIAM Journal on Discrete Mathematics*, 18(3):608–625, 2005.

V. Franc, S. Sonnenburg, and T. Werner. Cutting plane methods in machine learning. In S. Sra, S. Nowozin, and S. J. Wright, editors, *Optimization for Machine Learning.* MIT Press, 2012.

Eugene Freuder and Alan K. Mackworth. *Constraint satisfaction: An emerging paradigm.* In Rossi et al. (2006), 2006.

R. M. Haralick and L. G. Shapiro. The consistent labeling problem. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1(2):173–184, 1979.

Robert M. Haralick and Linda G. Shapiro. *Computer and Robot Vision.* Addison-Wesley Longman Publishing, 1992.

Jörg H. Kappes, Bjoern Andres, Fred A. Hamprecht, Christoph Schnörr, Sebastian Nowozin, Dhruv Batra, Sungwoong Kim, Bernhard X. Kausler, Jan Lellmann, Nikos Komodakis, and Carsten Rother. A comparative study of modern inference techniques for discrete energy minimization problem. In *Conf. Computer Vision and Pattern Recognition*, 2013.

Jon Kleinberg and Eva Tardos. Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. *J. ACM*, 49(5):616–639, 2002. ISSN 0004-5411.

Vladimir Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(10):1568–1583, 2006.

Vladimir Kolmogorov, Johan Thapper, and Stanislav Živný. The power of linear programming for general-valued CSPs. *SIAM Journal on Computing*, 44(1):1–36, 2015.

Arie Koster, C. P. M. van Hoesel, and A. W. J. Kolen. The partial constraint satisfaction problem: Facets and lifting theorems. *Operations Research Letters*, 23(3-5):89–97, 1998.

V. K. Koval and M. I. Schlesinger. Dvumernoe programmirovanie v zadachakh analiza izobrazheniy (Two-dimensional programming in image analysis problems). *Automatics and Telemechanics*, 8:149–168, 1976. In Russian.

V. A. Kovalevsky and V. K. Koval. A diffusion algorithm for decreasing the energy of the max-sum labeling problem. Glushkov Institute of Cybernetics, Kiev, USSR. Unpublished, approx. 1975.

S. Lauritzen. *Graphical Models.* Oxford University Press, 1996.

Stan Z. Li. *Markov Random Field Modeling in Image Analysis.* Springer, 2009. ISBN 9781848002784.

A. Mackworth. Constraint satisfaction. In *Encyclopaedia of Artificial Intelligence*, pages 285–292. Wiley, 1991.

Pedro Meseguer, Francesca Rossi, and Thomas Schiex. *Soft Constraints.* In Rossi et al. (2006), 2006.

M. Mézard and A. Montanari. *Information, Physics, and Computation.* Oxford University Press, Inc., New York, NY, USA, 2009.

Daniel Průša and Tomáš Werner. Universality of the local marginal polytope. In *Conf. on Computer Vision and Pattern Recognition*, pages 1738–1743. IEEE Computer Society, 2013.

Daniel Průša and Tomáš Werner. How hard is the LP relaxation of the Potts min-sum labeling problem? In *Conf. on Energy Minimization Methods in Computer Vision and Pattern Recognition, Hongkong*, pages 57–70. Springer, 2015a.

Daniel Průša and Tomáš Werner. Universality of the local marginal polytope. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 37(4):898–904, April 2015b.

A. Rosenfeld, R. A. Hummel, and S. W. Zucker. Scene labeling by relaxation operations. *IEEE Trans. on Systems, Man, and Cybernetics*, 6(6):420–433, June 1976.

Francesca Rossi, Peter van Beek, and Toby Walsh, editors. *Handbook of Constraint Programming.* Elsevier, 2006.

Carsten Rother, Vladimir Kolmogorov, Victor S. Lempitsky, and Martin Szummer. Optimizing binary MRFs via extended roof duality. In *Conf. Computer Vision and Pattern Recognition, Minneapolis, USA*, 2007.

Michail I. Schlesinger. *Matematicheskie sredstva obrabotki izobrazheniy (Mathematical Tools of Image Processing).* Naukova Dumka, Kiev, 1989. In Russian.

H. D. Sherali and W. P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal of Discrete Mathematics*, 3(3):411–430, 1990.

M. I. Shlezinger. Syntactic analysis of two-dimensional visual signals in noisy conditions. *Cybernetics and Systems Analysis*, 12(4):612–628, 1976. Translation from Russian.

David Sontag, Amir Globerson, and Tommi Jaakkola. Introduction to dual decomposition for inference. In Suvrit Sra, Sebastian Nowozin, and Stephen J. Wright, editors, *Optimization for Machine Learning.* MIT Press, 2012.

Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen, and Carsten Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(6):1068–1080, 2008.

Johan Thapper and Stanislav Živný. The power of linear programming for valued CSPs. In *Symp. Foundations of Computer Science*, pages 669–678. IEEE, 2012.

Stanislav Živný. *The Complexity of Valued Constraint Satisfaction Problems.* Cognitive Technologies. Springer, 2012.

M. Wainwright, T. Jaakkola, and A. Willsky. MAP estimation via agreement on (hyper)trees: message passing and linear programming approaches. *IEEE Trans. Information Theory*, 51(11):3697–3717, 2005.

Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.

Tomáš Werner. A linear programming approach to max-sum problem: A review. Technical Report CTU-CMP-2005-25, Center for Machine Perception, Czech Technical University, December 2005.

Tomáš Werner. A linear programming approach to max-sum problem: A review. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(7): 1165–1179, July 2007.

Tomáš Werner. High-arity interactions, polyhedral relaxations, and cutting plane algorithm for soft constraint optimisation (MAP-MRF). In *Computer Vision and Pattern Recognition Conf., Anchorage, USA*, June 2008.

Tomáš Werner. Revisiting the decomposition approach to inference in exponential families and graphical models. Research Report CTU-CMP-2009-06, Center for Machine Perception, K13133 FEE Czech Technical University, Prague, Czech Republic, May 2009.

Tomáš Werner. Revisiting the linear programming relaxation approach to Gibbs energy minimization and weighted constraint satisfaction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 32(8):1474–1488, August 2010.

Stanislav Živný, Tomáš Werner, and Daniel Průša. The power of LP relaxation for MAP inference. In S. Nowozin, P. V. Gehler, J. Jancsary, and C. Lampert, editors, *Advanced Structured Prediction.* MIT Press, 2014.

Ing. Tomáš Werner, Ph.D.

History Tomáš Werner was born in the Czech Republic in 1969. In 1999, he received his Ph.D. degree at the Center for Machine Perception, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, and started to work there as a researcher. In 2001–2002 he worked as a post-doctoral researcher at the Visual Geometry Group at the Oxford University, U.K., then he returned back. Today, he is an academic researcher at the Center for Machine Perception.

Research In the past, his main interest was multiple view geometry and three-dimensional reconstruction in computer vision. Since 2004, his interest is machine learning and optimization, in particular graphical models. He is a (co-)author of more than 70 publications, with 355 citations and H-index 8 in WoS.

Reviewing He regularly reviews for journals IEEE PAMI, IJCV, and JMLR. He has reviewed for conferences CVPR 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016; ICCV 2005, 2007, 2009, 2011, 2015 (area chair); ECCV 2004, 2006, 2008, 2010, 2012; NIPS 2011, 2012, 2013, 2014; ICML 2013, 2015; BMVC 2005; ACCV 2006; EMMCVPR 2009, 2011; SofT 2010.

Teaching He established and teaches Optimization (A4B33OPT) at the Faculty of Electrical Engineering, CTU.

Participation in projects

2001: Marie Curie Individual Fellowship HPMF-CT-2000-00975 (VGG, Univ. of Oxford)

2005-2007: INTAS 04-77-7347 PRINCESS

2004-2006: FP6 IST-004176 COSPAL

2007-2010: FP7 ICT-215078 DIPLECS (leader of CTU part)

2011-2015: FP7 ICT-270138 DARWIN

2010-2014: GACR P103/10/0783 Structure and Its Impact for Recognition)

2012-2015: GACR P202/12/2071 Structured Statistical Models for Image Understanding (applicant)

2016-2018: GACR 16-05872S Probabilistic Graphical Models and Deep Learning (applicant)

Awards

1995: Best Paper Award at *Austrian Pattern Recognition Conference* for paper [2]

1998: Best Paper Award at the same conference for paper [4]

2003: Best Paper Honourable Mention at *Computer Vision and Pattern Recognition Conference* for paper [8]

2007: Outstanding Reviewer award at *Computer Vision and Pattern Recognition Conference*

2008: Outstanding Reviewer award at *European Conference on Computer Vision*

2009: Outstanding Reviewer award at *International Conference on Computer Vision*

2015: Article [15] was evaluated by the Czech government as an excellent result in the II. pillar of RIV.

Selected Publications

1. T Werner, R D Hersch, and V Hlavac. Rendering real-world objects using view interpolation. *Intl. Conf. on Computer Vision*, Boston, USA, 1995.
2. T Werner, R D Hersch, V Hlavac. Rendering Real-World Objects Using View Interpolation. *16th meeting of the Austrian Assoc. for Pattern Recognition*, Maribor, Slovenia, 1995.
3. T Werner, T Pajdla, V Hlavac. Efficient rendering of projective model for image-based visualization. *Intl. Conf. on Pattern Recognition*, Brisbane, Australia, 1998.
4. T Werner, T Pajdla, V Hlavac. Oriented Projective Reconstruction. *Pattern Recognition and Medical Comp. Vision: 22-nd Workshop of the Austrian Assoc. for Pattern Recognition*, Illmitz, Austria, 1998.
5. T Werner, T Pajdla. Oriented Matching Constraints. *British Machine Vision Conf.*, Manchester, UK, 2001.
6. T Werner, A Zisserman. New Techniques for Automated Architectural Reconstruction from Photographs. *European Conf. on Computer Vision*, Kopenhangen, 2002,
7. O Chum, T Werner, T Pajdla. Joint Orientation of Epipoles. *British Machine Vision Conf.*, Norwich, UK, 2003.
8. T Werner. Constraint on Five Points in Two Images. *Computer Vision and Pattern Recognition Conf.*, Madison, USA, 2003.
9. T Werner. Combinatorial Constraints on Multiple Projections of a Set of Points. *Intl. Conf. on Computer Vision*, Nice, France, 2003.

10. O Chum, T Werner, J Matas. Two View Geometry Estimation Unaffected by Dominant Plane. *Computer Vision and Pattern Recognition Conf.*, San Diego, USA, 2005.
11. T Werner. A Linear Programming Approach to Max-sum Problem: A review. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 29(7), 2007.
12. T Werner. What Is Decreased by the Max-sum Arc Consistency Algorithm? *Intl. Conf. on Machine Learning*, Corvalis, USA, 2007.
13. T Werner. High-arity Interactions, Polyhedral Relaxations, and Cutting Plane Algorithm for Soft Constraint Optimisation (MAP-MRF). *Computer Vision and Pattern Recognition Conf.*, Anchorage, USA, 2008.
14. T Werner. Marginal Consistency: Unifying Constraint Propagation on Commutative Semirings. *Intl. Workshop on Preferences and Soft Constraints* (collocated with *Intl. Conf. on Principles and Practice of Constraint Programming*), Sydney, Australia, 2008.
15. T Werner. Revisiting the Linear Programming Relaxation Approach to Gibbs Energy Minimization and Weighted Constraint Satisfaction. *IEEE Trans. on Pattern Recognition and Machine Intelligence* 32(8), 2010.
16. T Werner. Primal View on Belief Propagation. *Conf. on Uncertainty in Artificial Intelligence*, Catalina Island, USA, 2010.
17. V Franc, S Sonnenburg, T Werner. Cutting Plane Methods in Machine Learning. A chapter in *Optimization for Machine Learning*, MIT Press, 2012.
18. D Prusa, T Werner. Universality of the Local Marginal Polytope. *Computer Vision and Pattern Recognition Conf.*, Portland, USA, 2013.
19. S Zivny, T Werner, D Prusa. The Power of LP Relaxation for MAP Inference. A chapter in *Advanced Structured Prediction*, MIT Press, 2014.
20. D Prusa, T Werner. Universality of the Local Marginal Polytope. *IEEE Trans. on Pattern Recognition and Machine Intelligence* 37(4), April 2015.
21. T Werner. Marginal Consistency: Upper-Bounding Partition Functions over Commutative Semirings. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 37(7), July 2015.
22. D Prusa, T Werner. How hard is the LP relaxation of the Potts min-sum labeling problem? *Conf. on Energy Minimization Methods in Computer Vision and Pattern Recognition*, Hongkong. Springer, 2015.