Maximum Persistency via Iterative Relaxed Inference with Graphical Models

Alexander Shekhovtsov
ICG, TU Graz, Austria
shekhovtsov@icg.tugraz.at

Paul Swoboda
HCI, Heidelberg University, Germany
swoboda@math.uni-heidelberg.de

Bogdan Savchynskyy
TU Dresden, Germany
bogdan.savchynskyy@tu-dresden.de

Introduction

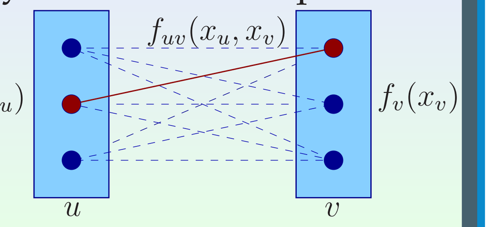
Abstract

- We consider: energy minimization for graphical models
- We obtain: a part of a globally optimal solution (persistent assignment)
- Properties:
  - scalable algorithm
  - maximizes the number of persistent variables
  - provably outperforms most of existing techniques

Energy Minimization

Given a graph $(\mathcal{V},\mathcal{E})$, associated variables $x_v \in \mathcal{X}_v$, $v \in \mathcal{V}$, and potentials $f_{\mathrm{C}}(x_{\mathrm{C}}) \in \mathbb{R}$, $\mathrm{C} \in \mathcal{V}\cup\mathcal{E}$, we consider the energy minimization problem:

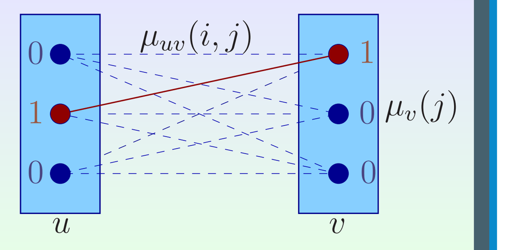
$$\min_{x\in\mathcal{X}} E_f(x)$$
$$= \min_{x\in\mathcal{X}} \Big\{ \textstyle\sum_{v\in\mathcal{V}} f_v(x_v) + \sum_{uv\in\mathcal{E}} f_{uv}(x_u,x_v) \Big\}$$
$$= \min_{x\in\mathcal{X}} \langle f, \delta(x)\rangle$$



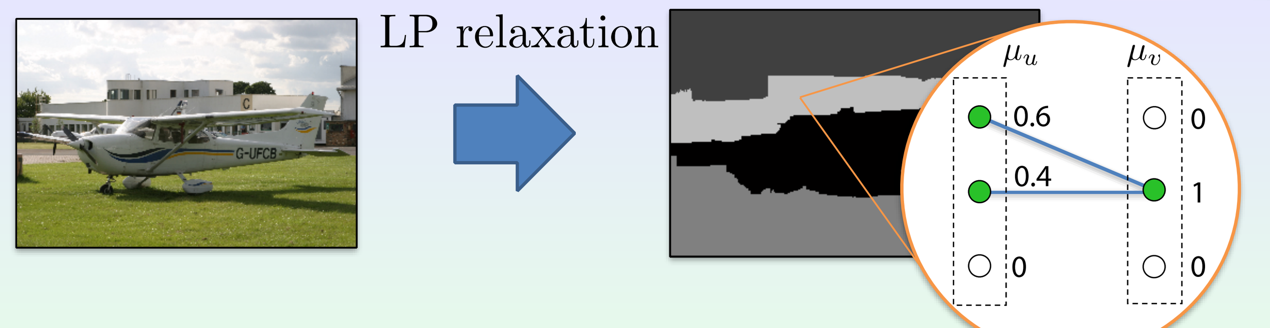
ILP / LP Relaxation

$$= \min_{\mu\in\mathrm{conv}(\delta(\mathcal{X}))} \langle f,\mu\rangle \geq \min_{\mu} \langle f,\mu\rangle$$

$$\Lambda \begin{cases} \sum_i \mu_u(i) = 1 \\ \sum_j \mu_{uv}(i,j) = \mu_u(i) \\ \mu \geq 0 \end{cases}$$

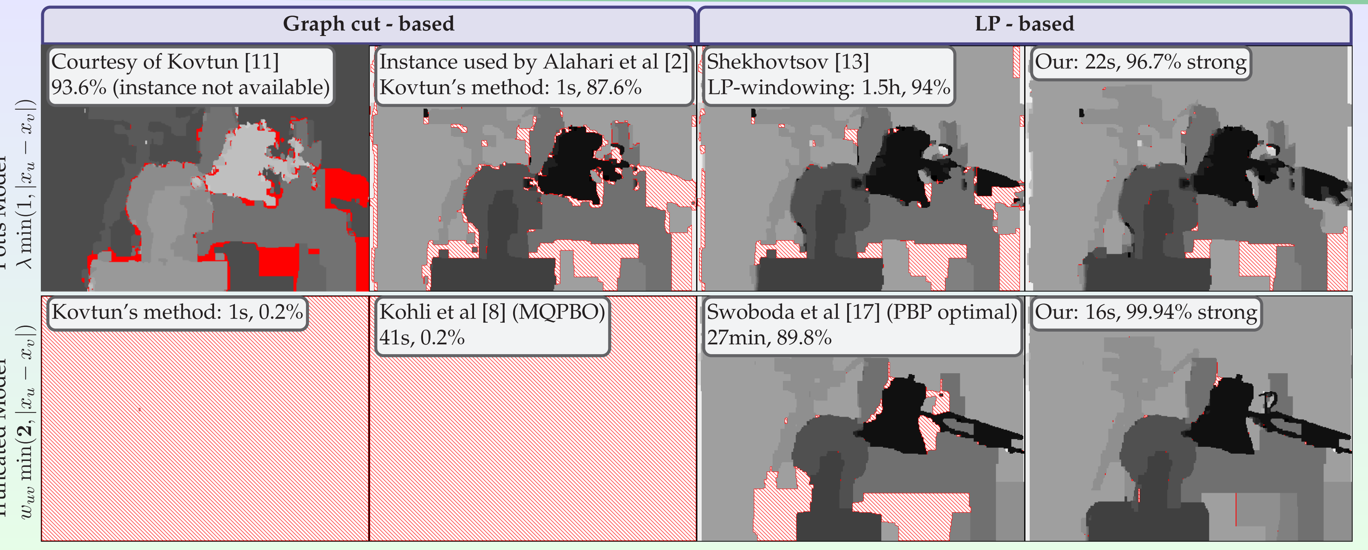


Persistency



- Is the integer part of the relaxed solution optimal?
- Can we eliminate labels that are not in the support set of relaxed solutions?

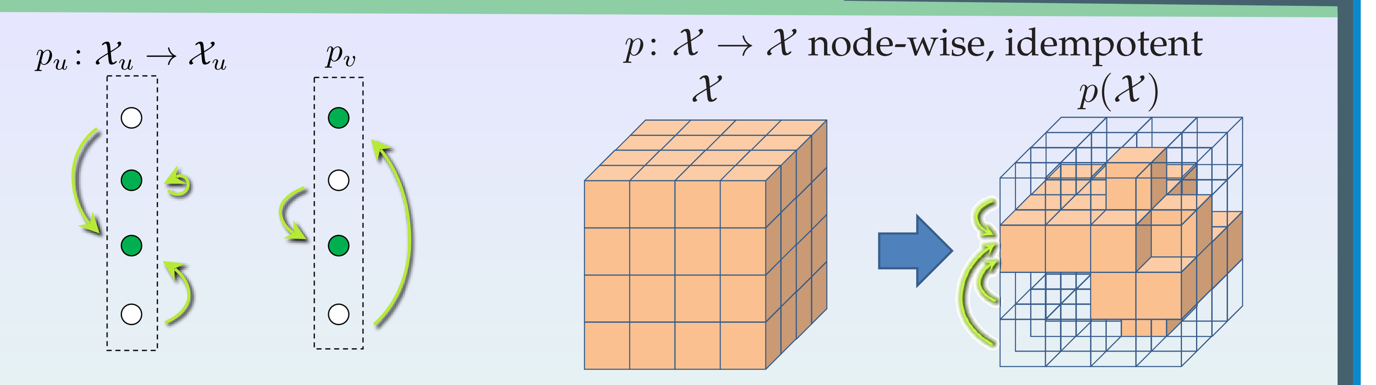
Progress in Partial Optimality Methods



• Model 1 [11, 2]: Potts, strong unaries; • Model 2 [18]: per-pixel unaries.

Relaxed-Improving Mapping

Improving Mapping (Substitution of labels)



Definition: Mapping $p\colon \mathcal{X}\to\mathcal{X}$ is *improving* if $\forall x\ E_f(p(x)) \leq E_f(x)$

- Equivalent to: $\min_{x\in\mathcal{X}}(\underbrace{E_f(x) - E_f(p(x))}) \geq 0$

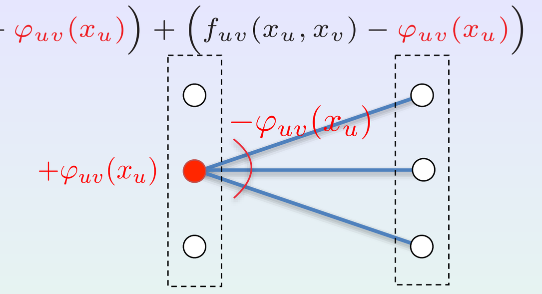
The difference energy $E_g(x)$, $\quad g = f - P^{\mathsf{T}} f$

(What is P?: $(P^{\mathsf{T}} f)_{\mathrm{C}}(x_{\mathrm{C}}) = f_{\mathrm{C}}(p_{\mathrm{C}}(x_{\mathrm{C}}))$, or, in primal: $P\delta(x) = \delta(p(x))$)

Definition: Mapping p is *relaxed-improving* if $\min_{\mu\in\Lambda}\langle (I-P^{\mathsf{T}})f, \mu\rangle \geq 0$ **(P)**

Dually, Through Equivalent Transformations

Equivalent Transformations:
$(\forall x\in\mathcal{X})\ E_f(x) = E_{f^\varphi}(x)$
$f^\varphi_u(x_u) := f_u(x_u) + \sum_{v\in\mathcal{N}(u)} \varphi_{uv}(x_u)$
$f^\varphi_{uv}(x_u,x_v) := f_{uv}(x_u,x_v) - \varphi_{uv}(x_u) - \varphi_{vu}(x_v)$



- Consider locally improving condition: $f_{\mathrm{C}}(p_{\mathrm{C}}(x_{\mathrm{C}})) \leq f_{\mathrm{C}}(x_{\mathrm{C}})$, $\forall x_{\mathrm{C}}$
- \+ equiv. transformations: $\exists\varphi\ \forall \mathrm{C}\in\mathcal{V}\cup\mathcal{E}, \forall x_{\mathrm{C}}\ f^\varphi_{\mathrm{C}}(p_{\mathrm{C}}(x_{\mathrm{C}})) \leq f^\varphi_{\mathrm{C}}(x_{\mathrm{C}})$ **(D)**

Theorem: The primal **(P)** and dual **(D)** definitions are equivalent.

[1] Adams WP, Lassiter JB, Sherali HD (1998) Persistency in 0-1 polynomial programming. Mathematics of Operations Research 23(2)
[2] Alahari K, Kohli P, Torr PHS (2008) Reduce, reuse & recycle: Efficiently solving multi-label MRFs. In: CVPR
[3] Fix A, Gruber A, Boros E, Zabih R (2011) A graph cut algorithm for higher-order Markov random fields. In: ICCV, pp 1020–1027
[4] Goldstein RF (1994) Efficient rotamer elimination applied to protein side-chains and related spin glasses. Biophysical Journal 66(5)
[5] Hammer P, Hansen P, Simeone B (1984) Roof duality, complementation and persistency in quadratic 0-1 optimization. Mathematical Programming pp 121–155
[6] Ishikawa H (2011) Transformation of general binary MRF minimization to the first-order case. PAMI 33(6):1234–1249
[7] Kahl F, Strandmark P (2011) Generalized roof duality for pseudo-Boolean optimization. In: ICCV, pp 255–262
[8] Kohli P, Shekhovtsov A, Rother C, Kolmogorov V, Torr P (2008) On partial optimality in multi-label MRFs. In: ICML, pp 480–487
[9] Kolmogorov V (2006) Convergent tree-reweighted message passing for energy minimization. PAMI 28(10):1568–1583
[10] Kolmogorov V (2010) Generalized roof duality and bisubmodular functions. In: NIPS, pp 1144–1152
[11] Kovtun I (2003) Partial optimal labeling search for a NP-hard subclass of (max, +) problems. In: DAGM-Symposium, pp 402–409
[12] Kovtun I (2011) Sufficient condition for partial optimality for (max, +) labeling problems and its usage. Control Systems and Computers (2)
[13] Shekhovtsov A (2014) Maximum persistency in energy minimization. In: CVPR, pp 1162–1169
[14] Shekhovtsov A (2015) Higher order maximum persistency and comparison theorems. CVIU SI on Inference and Learning of Graphical Models DOI 10.1016/j.cviu.2015.05.002, in press
[15] Sherali HD, Adams WP (1990) A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. SIAM J Discrete Math 3(3):411–430
[16] Swoboda P, Savchynskyy B, Kappes J, Schnörr C (2014) Partial optimality by pruning for MAP-inference with general graphical models. In: CVPR, pp 1170–1177, DOI 10.1109/CVPR.2014.153
[17] Swoboda P, Shekhovtsov A, Kappes JH, Schnörr C, Savchynskyy B (2014) Partial optimality by pruning for MAP-inference with general graphical models. Computing Research Repository abs/1410.6641, 1410.6641
[18] Szeliski R, Zabih R, Scharstein D, Veksler O, Kolmogorov V, Agarwala A, Tappen M, Rother C (2008) A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. PAMI
[19] Thapper J, Živný S (2013) The complexity of finite-valued CSPs. In: Symposium on the Theory of Computing (STOC), pp 695–704
[20] Werner T (2007) A linear programming approach to max-sum problem: A review. PAMI 29(7):1165–1179

The Problem

Maximum Persistency

- Given that the verification problem is solvable, which method is better?

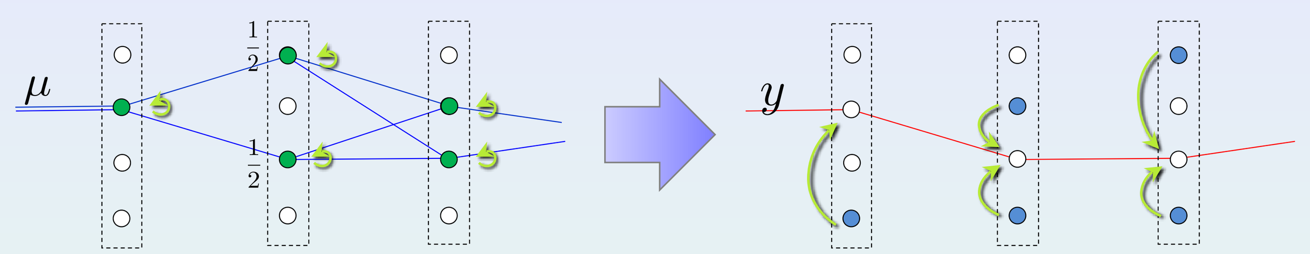
Proposition

Pose "the best partial optimality" as optimization problem. Find the mapping $p\colon\mathcal{X}\to\mathcal{X}$ that delivers the maximum problem reduction:

$$\min_{p\in\mathcal{P}} \sum_{u\in\mathcal{V}} |p_u(\mathcal{X}_u)| \text{ s.t. } p \text{ is relaxed-improving; } \mathcal{P} \text{ - class of mappings.}$$

Subset-to-one Class of Mappings

Theorem: Let μ be a solution to LP-relaxation: $\mu \in \mathrm{argmin}_{\mu\in\Lambda}\langle f,\mu\rangle$ and $p\colon\mathcal{X}\to\mathcal{X}$ be (strictly) relaxed-improving. Then $P\mu = \mu$.



- Fix a *test* labeling y from μ and try substitute other labels with it.
- Mapping p_u selects a subset of labels in u to be substituted with y_u, there are $2^{|\mathcal{X}_u|-1}$ choices.
- Covers all methods marked in the table below

Theorem: Maximum Persistency problem over subset-to-one class of mappings is solvable in polynomial time [13, 14].

- This work: new efficient algorithm, connecting [13] and Pruning-Based-Persistency [16] (CVPR'14).

Generality of Sufficient Conditions

Relaxed-improving condition with natural (local) relaxations are satisfied for all of the following [13, 14]:

	Method	
multilabel pairwise	Simple DEE [4]	✓
	MQPBO [8]	✓
	[11] one-agains-all	✓
	[12] iterative	✓
	Swoboda et al [16]*	✓
pseudo-Boolean higher order	Roof dual / QPBO [5]	✓
	Reductions: HOCR [6], [3]	FLP
	Bisubmodular relaxations [10]**	BLP
	Generalized Roof Dualilty [7]	FLP
	Persistency by Adams et al [1]	FLP

– all-to-one class of mappings

BLP = Basic LP Relaxation [20, 19];

FLP = Full Local LP Relaxation, equivalent to [15];

*[16] is higher order but the comparison proof is for pairwise case.

**Result holds for sum of bisubmodular functions over the same hypergraph as the BLP relaxation.

Algorithm

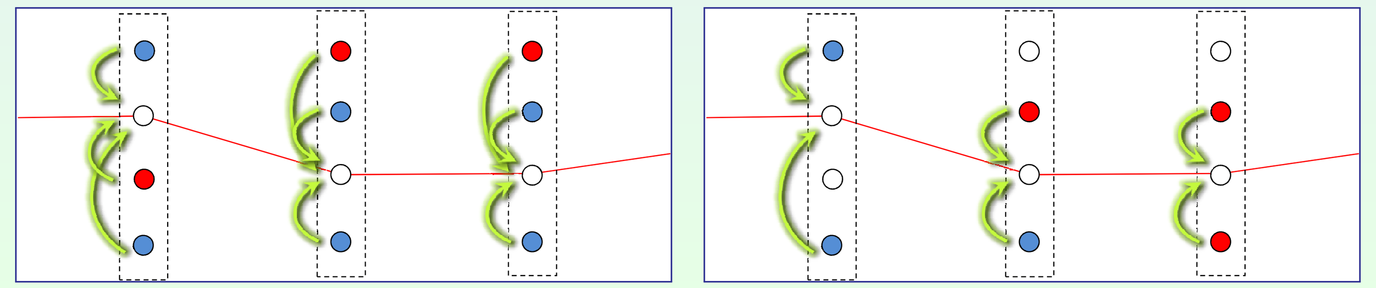
Discrete Cutting Plane

- Start with a mapping p that substitutes *everything* with y
- Construct auxiliary 'difference' problem $g = (I - P^{\mathsf{T}})f$
- Test persistency conditions (relaxed inference for g):

Primal	Dual
$\min\langle g,\mu\rangle \geq 0$? $\mu\in\Lambda$	$\max_{\varphi,\psi} \sum_{u\in\mathcal{V}} \psi_u \geq 0$? $g^\varphi_u(x_u) \geq \psi_u$ $g^\varphi_{uv}(x_u,x_v) \geq 0$

- If not satisfied, force p to identity on the following labels x_u:

in the support set of the minimizer: $\mu_u(x_u) > 0$	corresponding to active constraints: $g^\varphi_u(x_u) = \psi_u$



Correctness and Optimality

- Runs in polynomial time;
- Solves the maximum persistency problem exactly when the test relaxation is solved exactly and the solution is a strict relative interior optimal (e.g. interior point method);
- Returns an improving mapping even when the dual is solved sub-optimally – can use fast dual solvers, we used TRW-S [9].

Efficiency

Challenges:
- solving relaxed inference approximately even once is slow
- TRW-S is not finitely converging

How can we iterate such relaxed inference?

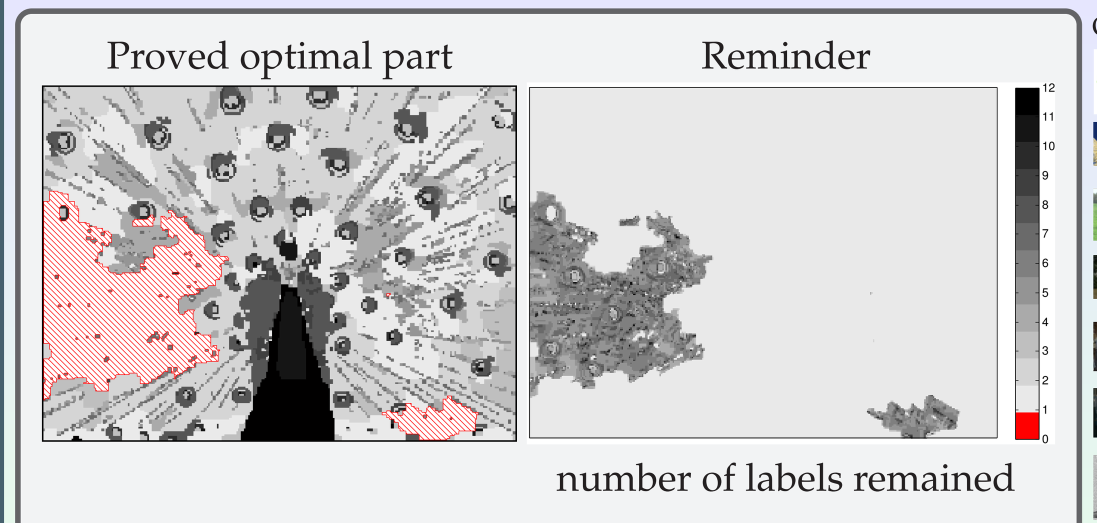
Fast implementation with TRW-S
- Warm-start: reuse reparametrizations φ in outer iterations
- Guaranteed to prune something even after 1 iteration of TRW-S
- An optimal **pruning** is often possible before the dual is solved
- Problem **reductions** preserving the sufficient condition
- **Fast message** passing for $(I-P^{\mathsf{T}})f$ with reductions

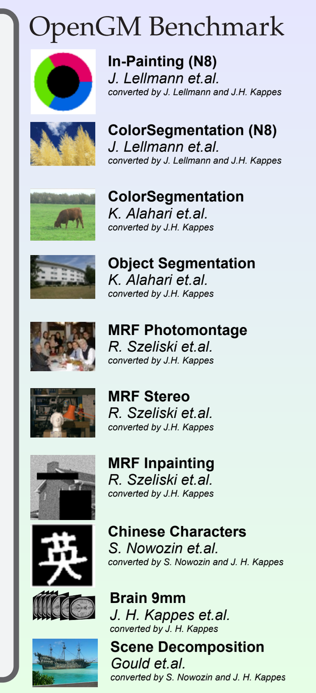
Combined Effect of Speedups

Instance	Initialization (1000 it.)	Extra time for persistency				
		no speedups	+reduction	+node pruning	+labeling pruning	+fast msgs
Protein folding `1CKK`	8.5s	268s (26.53%)	168s (26.53%)	2.0s (26.53%)	2.0s (26.53%)	2.0s (26.53%)
colorseg-n4 `pfau-small`	9.3s	439s (88.59%)	230s (93.41%)	85s (93.41%)	76s (93.41%)	19s (93.41%)

Experiments

Algorithm Demo

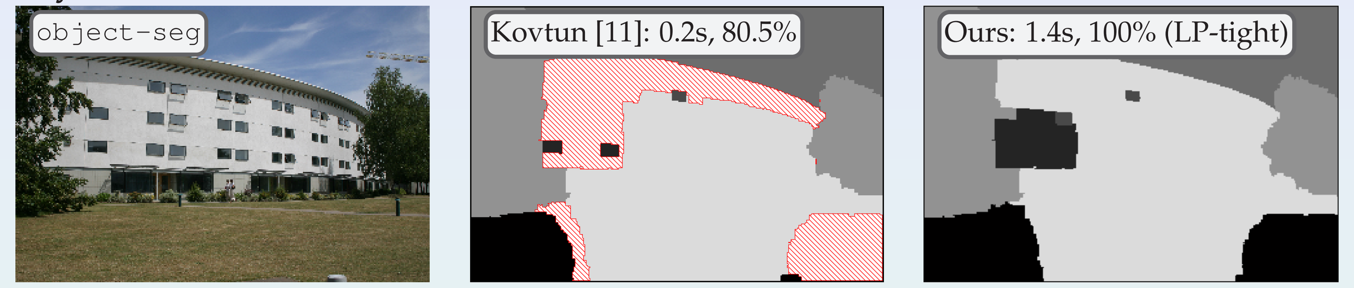




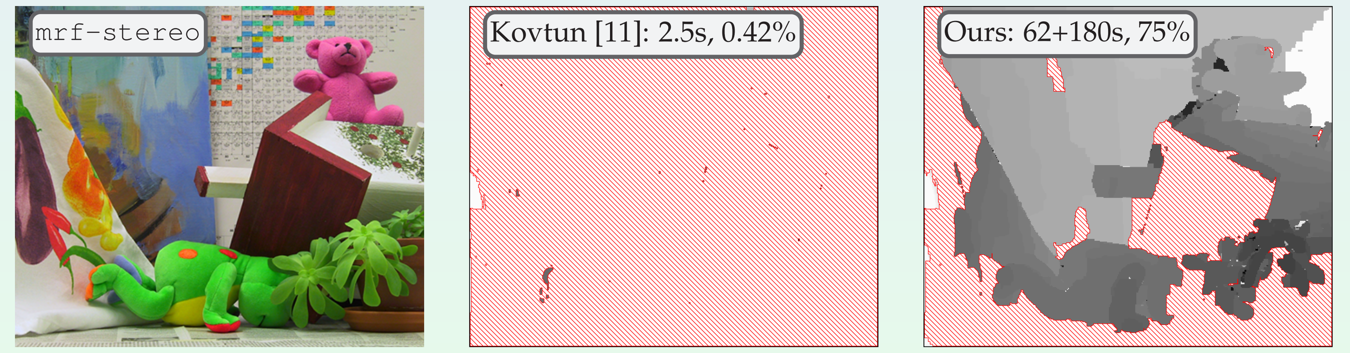
OpenGM Benchmark

Problem family	#I	#L	#V	MQPBO		MQPBO-10		Kovtun		[16]-TRWS		Our-TRWS	
`mrf-stereo`	3	16-60	>100000	†		†		†		2.5h	13%	117s	**73.56%**
`mrf-photomontage`	2	5-7	≤ 514080	93s	22%	866s	16%	†		3.7h	16%	483s	**41.98%**
`color-seg`	3	3-4	≤ 424720	22s	11%	87s	16%	**0.3s**	98%	1.3h	>99%	61.8s	**99.95%**
`color-seg-n4`	9	3-12	≤ 86400	22s	8%	398s	14%	**0.2s**	67%	321s	90%	4.9s	**99.26%**
`ProteinFolding`	21	≤ 483	≤ 1972	685s	2%	2705s	2%	†		48s	18%	9.2s	**55.70%**
`object-seg`	5	4-8	68160	3.2s	0.01%	†		**0.1s**	93.86%	138s	98.19%	2.2s	**100%**

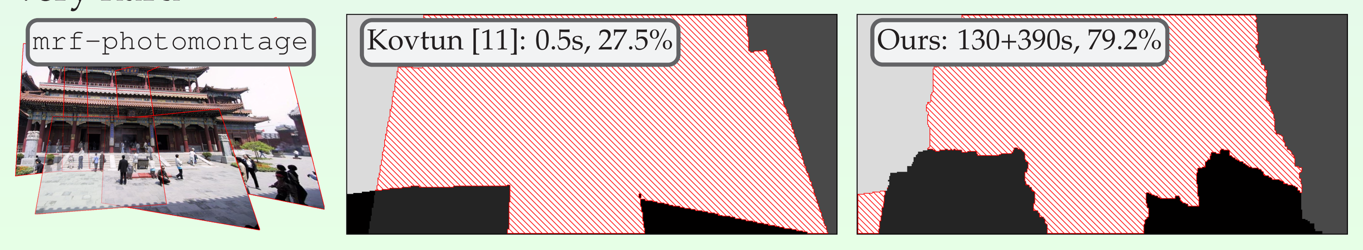
Easy



Hard



Very hard



Implementation

C++/Matlab
`http://icg.tugraz.at/Members/shekhovtsov/persistency`