Global MAP-Optimality by Shrinking the Combinatorial Search Area with Convex Relaxation

Bogdan Savchynskyy, Jörg Kappes, Paul Swoboda,
Christoph Schnörr

Heidelberg Collaboratory for Image Processing (HCI)
University of Heidelberg

Acknowledgement: *Thanks to A. Shekhovtsov, B. Flach, T. Werner, K. Antoniuk, V. Franc from CMP of TU Prague for the extreme patience and fruitful discussions*

MRF Energy Minimization

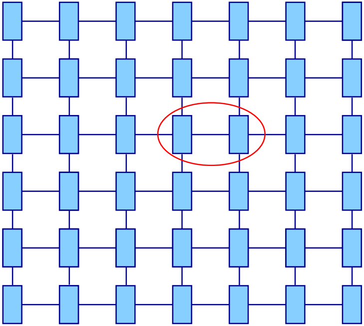
$$\min_{x \in \mathcal{X}} E(x) := \min_{x \in \mathcal{X}} \sum_{v \in \mathcal{V}} \theta_v(x_v) + \sum_{uv \in \mathcal{E}} \theta_{uv}(x_u, x_v)$$

- **Segmentation** [Rother et al. 2004], [Nowozin, Lampert 2010]
- **Multi-camera stereo** [Kolmogorov, Zabih 2002]
- **Stereo and Motion** [Kim et al. 2003]
- **Clustering** [Zabih, Kolmogorov. 2004]
- **Medical imaging** [Raj et al. 2007]
- **Pose Estimation** [Bergtholdt et al. 2010], [Bray et al. 2006]
- ...

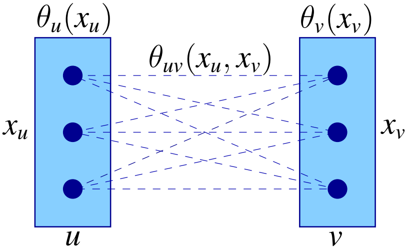
Computer Vision energy minimization benchmarks:
[Szeliski et al. 2008], [Kappes et al. CVPR, 2013]

MRF Energy Minimization

$$\min_{x \in \mathcal{X}} E(x) := \min_{x \in \mathcal{X}} \sum_{v \in \mathcal{V}} \theta_v(x_v) + \sum_{uv \in \mathcal{E}} \theta_{uv}(x_u, x_v)$$



graph $(\mathcal{V}, \mathcal{E})$



Integer LP Formulation

$$\min_{\mu \geqslant 0} \sum_{v\in\mathcal{V}} \sum_{x_v\in\mathcal{X}_v} \theta_v(x_v)\mu_v(x_v) + \sum_{uv\in\mathcal{E}} \sum_{x_u,x_v\in\mathcal{X}_{uv}} \theta_{uv}(x_u,x_v)\mu_{uv}(x_u,x_v)$$

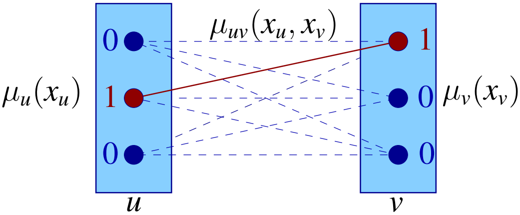
s.t.

$$\sum_{x_v\in\mathcal{V}} \mu_v(x_v) = 1,\ v\in\mathcal{V}$$

$$\sum_{x_v\in\mathcal{V}} \mu_{uv}(x_u,x_v) = \mu_u(x_u),\ x_u\in\mathcal{X}_u,\ uv\in\mathcal{E}$$

$$\sum_{x_u\in\mathcal{V}} \mu_{uv}(x_u,x_v) = \mu_v(x_v),\ x_v\in\mathcal{X}_v,\ uv\in\mathcal{E}\,.$$

$$\mu\in\{0,1\}^N$$



~~Integer~~ LP Formulation

$$\min_{\mu \geqslant 0} \sum_{v \in \mathcal{V}} \sum_{x_v \in \mathcal{X}_v} \theta_v(x_v)\mu_v(x_v) + \sum_{uv \in \mathcal{E}} \sum_{x_u, x_v \in \mathcal{X}_{uv}} \theta_{uv}(x_u, x_v)\mu_{uv}(x_u, x_v)$$

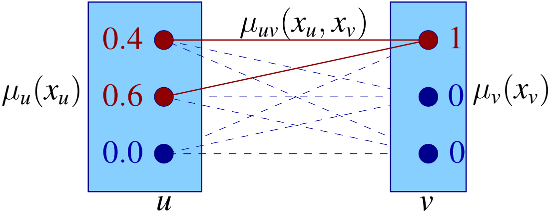
s.t.

$$\sum_{x_v \in \mathcal{V}} \mu_v(x_v) = 1,\ v \in \mathcal{V}$$

$$\sum_{x_v \in \mathcal{V}} \mu_{uv}(x_u, x_v) = \mu_u(x_u),\ x_u \in \mathcal{X}_u,\ uv \in \mathcal{E}$$

$$\sum_{x_u \in \mathcal{V}} \mu_{uv}(x_u, x_v) = \mu_v(x_v),\ x_v \in \mathcal{X}_v,\ uv \in \mathcal{E}\ .$$

$$\cancel{\mu \in \{0,1\}^N}\ \mu \in [0,1]^N$$



LP Relaxation: typical solution

color segmentation problem

integer and fractional labelings

- Is the integer part of the solution correct?
- In general - NO! In practice - mostly YES.
- How can it be exploited to find **an optimal** integer solution?

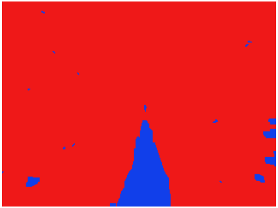
Related Approach: Partial Optimality

QPBO:[Hammer et al. 1984],[Boros, Hammer 2002],[Rother et al. 2007], [Kohli et al. 2008],[Windheuser et al. 2012], [Kahl,Strandmark 2012];
Submodular relaxation:[Kovtun 2003], [Kovtun PhD Thesis 2005], [Shekhovtsov,Hlavač 2011];
LP relaxation:[Swoboda et al. 2013, 2014],[Shekhovtsov 2014].

integer and fractional labeling

$\Rightarrow$

solved



partial solution

Our approach

[Kovtun 2003]

Related Approach: Partial Optimality

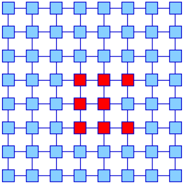
QPBO:[Hammer et al. 1984],[Boros, Hammer 2002],[Rother et al. 2007], [Kohli et al. 2008],[Windheuser et al. 2012], [Kahl,Strandmark 2012];
Submodular relaxation:[Kovtun 2003], [Kovtun PhD Thesis 2005], [Shekhovtsov,Hlavač 2011];
LP relaxation:[Swoboda et al. 2013, 2014],[Shekhovtsov 2014].

integer and fractional labeling $\Rightarrow$ NOT solved | partial solution

Our approach | [Kovtun 2003]

Algorithm Idea

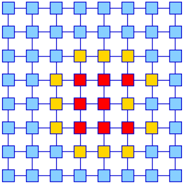
0) Initialize:



Identify LP and ILP parts.

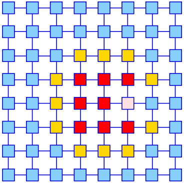
t) **Iterate** till agreement on the border:

→



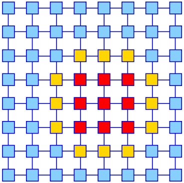
solve ILP (+) and LP (+) separately

→



check agreement on the border

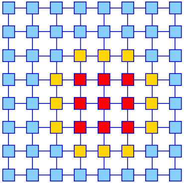
→



increase ILP subproblem + if disagree

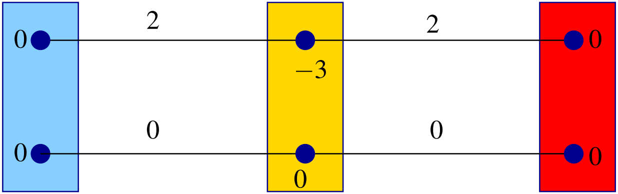
→

From Idea to Algorithm



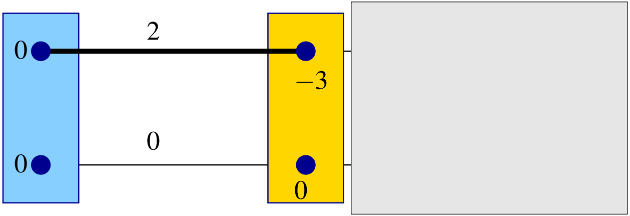
- Is agreement on the border sufficient for optimality?
- How to select the initial LP/ILP splitting?
- How to encourage agreement on the border?
- How to avoid re-solving the LP part?
- (Do we need to solve the LP relaxation to optimality?)

Is agreement on the border sufficient for optimality?



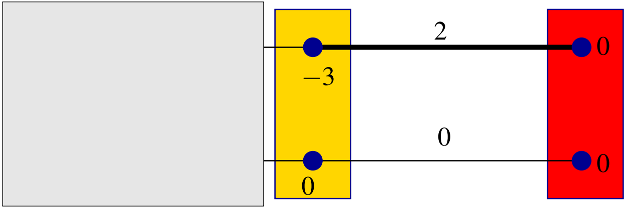
Counterexample due to A. Shekhovtsov

Is consistency on the border sufficient for optimality?



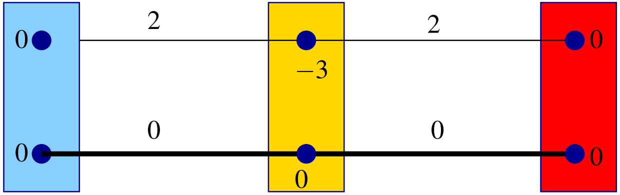
Counterexample due to A. Shekhovtsov

Is consistency on the border sufficient for optimality?



Counterexample due to A. Shekhovtsov

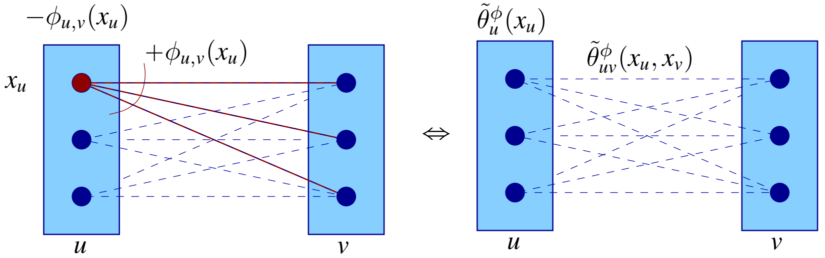
Is consistency on the border sufficient for optimality?



Counterexample due to A. Shekhovtsov

Background: Reparametrization (Equivalent transformations)

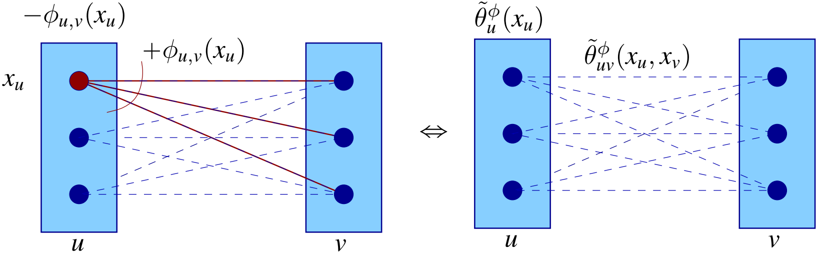
$$\sum_{v\in\mathcal{V}} \theta_v(x_v) + \sum_{uv\in\mathcal{E}} \theta_{uv}(x_u,x_v) \equiv \sum_{v\in\mathcal{V}} \tilde{\theta}_v^{\phi}(x_v) + \sum_{uv\in\mathcal{E}} \tilde{\theta}_{uv}^{\phi}(x_u,x_v)$$



Background: Reparametrization, Dual problem

Primal: $$E(x) = \min_{x} \sum_{v \in \mathcal{V}} \theta_v(x_v) + \sum_{uv \in \mathcal{E}} \theta_{uv}(x_u, x_v)$$

Dual: $$D(\phi) = \max_{\phi} \sum_{v \in \mathcal{V}} \min_{x_v} \tilde{\theta}^{\phi}_v(x_v) + \sum_{uv \in \mathcal{E}} \min_{x_{uv}} \tilde{\theta}^{\phi}_{uv}(x_{uv})$$

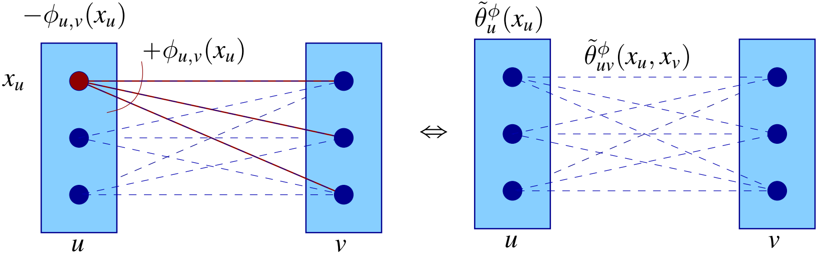


Background: Reparametrization, Dual problem

$$\text{Primal:} \quad E(x) = \min_x \sum_{v\in\mathcal{V}} \tilde{\theta}_v^{\phi}(x_v) + \sum_{uv\in\mathcal{E}} \tilde{\theta}_{uv}^{\phi}(x_u, x_v)$$

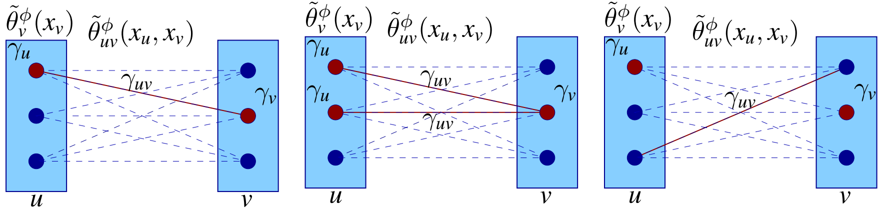
$$\text{Dual:} \quad D(\phi) = \max_{\phi} \sum_{v\in\mathcal{V}} \min_{x_v} \tilde{\theta}_v^{\phi}(x_v) + \sum_{uv\in\mathcal{E}} \min_{x_{uv}} \tilde{\theta}_{uv}^{\phi}(x_{uv})$$

$$D(\phi) \leqslant E(x)$$



Background: Arc Consistency

Dual: $$D(\phi) = \max_{\phi} \sum_{v \in \mathcal{V}} \underbrace{\min_{x_v} \tilde{\theta}^{\phi}_{v}(x_v)}_{\gamma_v} + \sum_{uv \in \mathcal{E}} \underbrace{\min_{x_{uv}} \tilde{\theta}^{\phi}_{uv}(x_{uv})}_{\gamma_{uv}}$$



strict arc consistency | ~~strict~~ arc consistency | ~~arc consistency~~

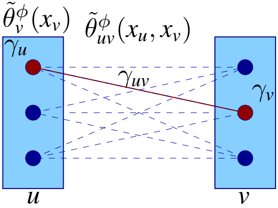
Background: Trivial Problem, Strict Arc Consistency

Theorem.

$$\begin{gathered}\text{Strict arc consistency in all nodes}\\ \Downarrow \\ \text{the non-relaxed problem is solved.}\end{gathered}$$

Proof. Strict arc consistency $\Rightarrow D(\phi) = E(x^*)$, x^* consists of $\gamma_v,\ \gamma_{uv}$.

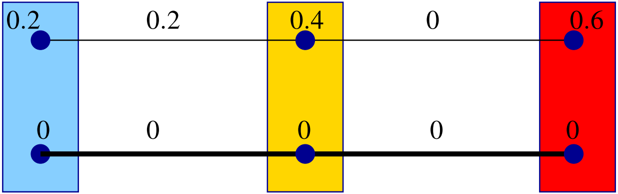
$D(\phi) \leqslant E(x) \Rightarrow x^*$ is the solution.



strict arc consistency

Consistency on the border sufficient for optimality.

Theorem. Let θ^{ϕ} be strictly arc consistent on ■+■. Then if LP (■+■) and ILP (■+■) solutions agree on the border (■) their concatenation is globally optimal.

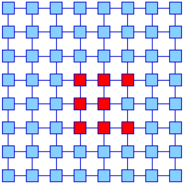


LP Relaxation: typical (approximate) solution

Blue - strictly arc consistent, red - otherwise.

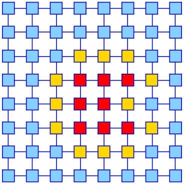
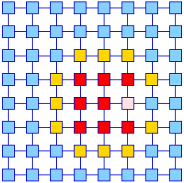
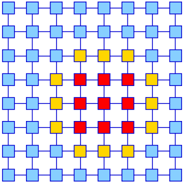
Algorithm

0) Initialize:



Solve LP relaxation and reparametrize: $\theta \rightarrow \tilde{\theta}^{\phi}$
'Blue' = the strictly arc consistent nodes.

t) Iterate till agreement on the border:

→  →  →  →

apply ILP solver to +

check agreement on the border

increase ILP subproblem + if disagree

Why reparametrize?

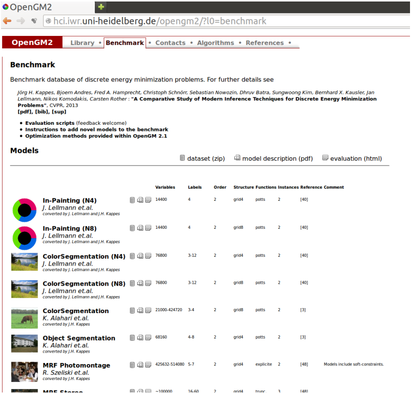
Reparametrization provides:

- optimality condition *(= consistency on border (■))*
- initial splitting criterion *(to ■ and ■)*
- encouraging of border consistency
  Optimal labels "vote" for themselves in both LP (■+■) and ILP (■+■) subproblems
- potential speed-up of combinatorial solvers
  Acts as LP pre-solving

Moreover...

- An LP solver needs to be executed only *once*
  Because due to strict consistency local decisions are optimal: nodes removal does not change the remaining part of the solution
- *Suboptimal* reparametrization can be used as well
  Because we did not employ optimality of the reparametrization

Experimental Evaluation: OpenGM Library and Benchmark



OpenGM2

hci.iwr.uni-heidelberg.de/opengm2/?l0=benchmark

OpenGM2 | Library • **Benchmark** • Contacts • Algorithms • References •

Benchmark

Benchmark database of discrete energy minimization problems. For further details see

Jörg H. Kappes, Bjoern Andres, Fred A. Hamprecht, Christoph Schnörr, Sebastian Nowozin, Dhruv Batra, Sungwoong Kim, Bernhard X. Kausler, Jan Lellmann, Nikos Komodakis, Carsten Rother : **"A Comparative Study of Modern Inference Techniques for Discrete Energy Minimization Problems"**, CVPR, 2013
[pdf], [bib], [sup]

- **Evaluation scripts** (feedback welcome)
- **Instructions to add novel models to the benchmark**
- **Optimization methods provided within OpenGM 2.1**

Models

dataset (zip) model description (pdf) evaluation (html)

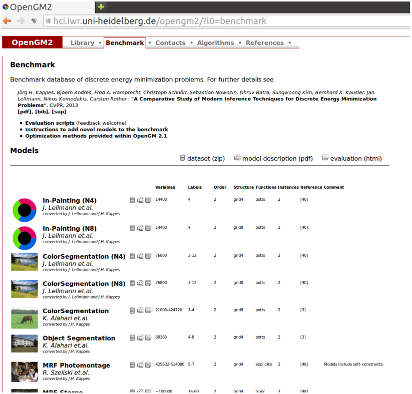
	Variables	Labels	Order	Structure	Functions	Instances	Reference	Comment
In-Painting (N4) *J. Lellmann et.al.* *converted by J. Lellmann and J.H. Kappes*	14400	4	2	grid4	potts	2	[40]	
In-Painting (N8) *J. Lellmann et.al.* *converted by J. Lellmann and J.H. Kappes*	14400	4	2	grid8	potts	2	[40]	
ColorSegmentation (N4) *J. Lellmann et.al.* *converted by J. Lellmann and J.H. Kappes*	76800	3-12	2	grid4	potts	2	[40]	
ColorSegmentation (N8) *J. Lellmann et.al.* *converted by J. Lellmann and J.H. Kappes*	76800	3-12	2	grid8	potts	2	[40]	
ColorSegmentation *K. Alahari et.al.* *converted by J.H. Kappes*	21000-424720	3-4	2	grid8	potts	2	[3]	
Object Segmentation *K. Alahari et.al.* *converted by J.H. Kappes*	68160	4-8	2	grid4	potts	2	[3]	
MRF Photomontage *R. Szeliski et.al.* *converted by J.H. Kappes*	425632-514080	5-7	2	grid4	explicite	2	[48]	Models include soft-constraints.
MRF Stereo	~100000	16-60	2	grid4	trunc.	3	[48]	

Open Library for Graphical Models:

- inference algorithms;
- benchmark data;
- includes Middlebury MRF benchmark
- comparison tables;

Just type `'OpenGM'` in Google...

Experimental Evaluation: OpenGM Library and Benchmark



OpenGM2

hci.iwr.uni-heidelberg.de/opengm2/?l0=benchmark

OpenGM2 | Library • **Benchmark** • Contacts • Algorithms • References •

Benchmark

Benchmark database of discrete energy minimization problems. For further details see

Jörg H. Kappes, Bjoern Andres, Fred A. Hamprecht, Christoph Schnörr, Sebastian Nowozin, Dhruv Batra, Sungwoong Kim, Bernhard X. Kausler, Jan Lellmann, Nikos Komodakis, Carsten Rother : **"A Comparative Study of Modern Inference Techniques for Discrete Energy Minimization Problems"**, CVPR, 2013
[pdf], [bib], [sup]

- **Evaluation scripts** (feedback welcome)
- **Instructions to add novel models to the benchmark**
- **Optimization methods provided within OpenGM 2.1**

Models

dataset (zip) model description (pdf) evaluation (html)

	Variables	Labels	Order	Structure	Functions	Instances	Reference	Comment
In-Painting (N4) *J. Lellmann et.al.* *converted by J. Lellmann and J.H. Kappes*	14400	4	2	grid4	potts	2	[40]	
In-Painting (N8) *J. Lellmann et.al.* *converted by J. Lellmann and J.H. Kappes*	14400	4	2	grid8	potts	2	[40]	
ColorSegmentation (N4) *J. Lellmann et.al.* *converted by J. Lellmann and J.H. Kappes*	76800	3-12	2	grid4	potts	2	[40]	
ColorSegmentation (N8) *J. Lellmann et.al.* *converted by J. Lellmann and J.H. Kappes*	76800	3-12	2	grid8	potts	2	[40]	
ColorSegmentation *K. Alahari et.al.* *converted by J.H. Kappes*	21000-424720	3-4	2	grid8	potts	2	[3]	
Object Segmentation *K. Alahari et.al.* *converted by J.H. Kappes*	68160	4-8	2	grid4	potts	2	[3]	
MRF Photomontage *R. Szeliski et.al.* *converted by J.H. Kappes*	425632-514080	5-7	2	grid4	explicite	2	[48]	Models include soft-constraints.
MRF Stereo	~100000	16-60	2	grid4	trunc.	3	[48]	

Open Library for Graphical Models:

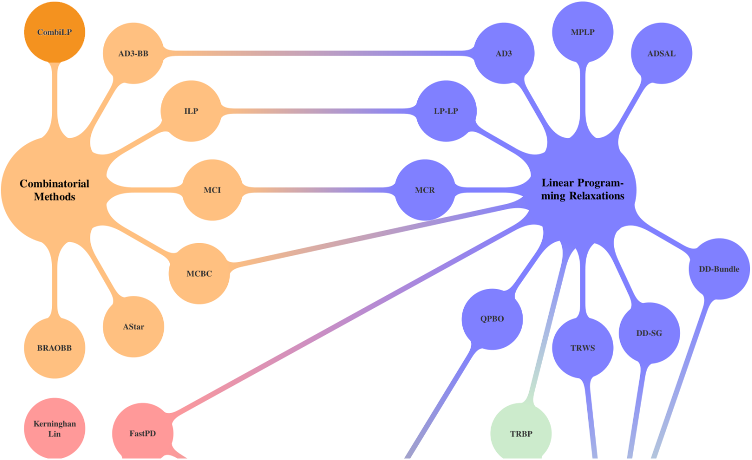
- inference algorithms;
- benchmark data;
- includes Middlebury MRF benchmark
- comparison tables;

Just type `'OpenGM'` in Google...

Our code is freely available as a part of OpenGM!

Experimental Evaluation: Methods

Algorithms, available in OpenGM:



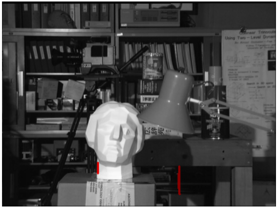
We used:

- **TRW-S** [Kolmogorov 2005] as LP solver
- **CPLEX** [IBM] as ILP solver.

Middlebury MRF Benchmark

tsukuba

384×288, 16 labels



$$E_{min} = 369218$$
$$E_{TRWS} = 369218$$

venus

434×383, 20 labels

$$E_{min} = 3048043$$
$$E_{TRWS} = 3048296$$

family

752×566, 5 labels

$$E_{min} = 184813$$
$$E_{TRWS} = 184825$$

Middlebury MRF Benchmark

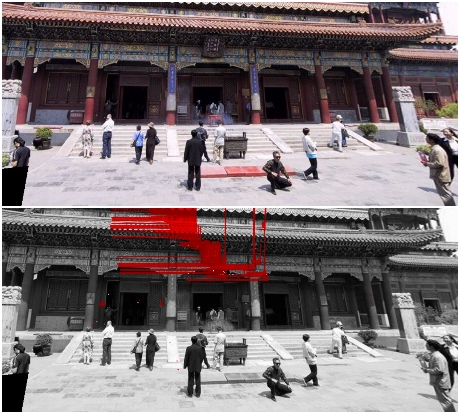
teddy

450×375, 60 labels

1 iteration of ILP
= out of memory

panorama

1071×480, 7 labels



1 iteration of ILP
= out of memory

Middlebury MRF Benchmark

Dataset	Step (1) LP (TRWS)			Step (3) ILP (CPLEX)			$\lvert\mathcal{B}\rvert$	
name	it	time, s	E	it	time, s	E	min	max
tsukuba	250	186	369537	24	36	369218	130	656
venus	2000	3083	3048296	10	69	3048043	66	233
teddy	10000	14763	1345214	1	–	–	2062	–
family	10000	20156	184825	18	2	184813	11	109
pano	10000	34092	169224	1	–	–	24474	–

Table : Results on Middlebury datasets

Color Segmentation: 26 Potts models

Solved

Heidelberg Collaboratory for Image Processing

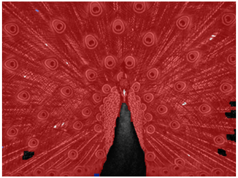
Potts Models: Comparison to State-of-the-Art

Dataset	LP step 0		ILP steps 1-3		MCA	MPLP		
	it	time, s	it	time, s	time, s	LP it	LP time, s	ILP time, s
pfau	1000	**276**	14	**14**	> 55496	10000	> 15000	
palm	200	**65**	17	**93**	561	700	1579	3701
clownfish	100	**32**	8	**10**	328	350	790	181
crops	100	**32**	6	**6**	355	350	797	1601
strawberry	100	**29**	8	**31**	483	350	697	1114

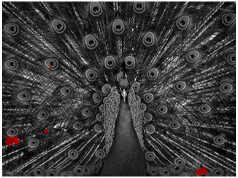
Table : Exemplary Potts model comparison on `Color segmentation (N8)` dataset.
Our method is **the fastest**.

MCA $=$ Multiway cut: [Kappes et al. 2011],[Kappes et al. 2013]
MPLP: [Globerson, Jaakkola 2007]+[Sonntag et al. 2008]

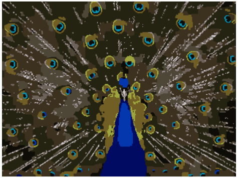
Comparison to Partial Optimality by [Kovtun 2003]



Method of Kovtun



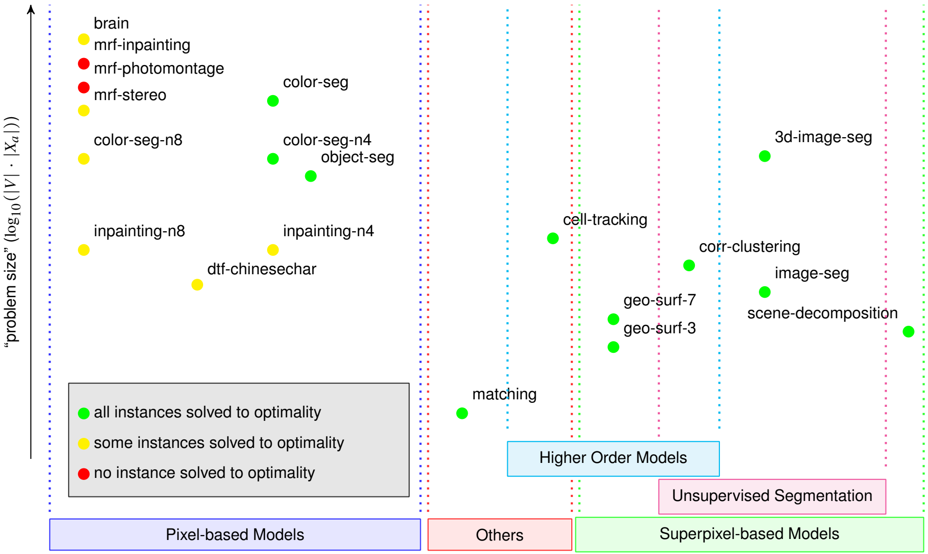
Our approach



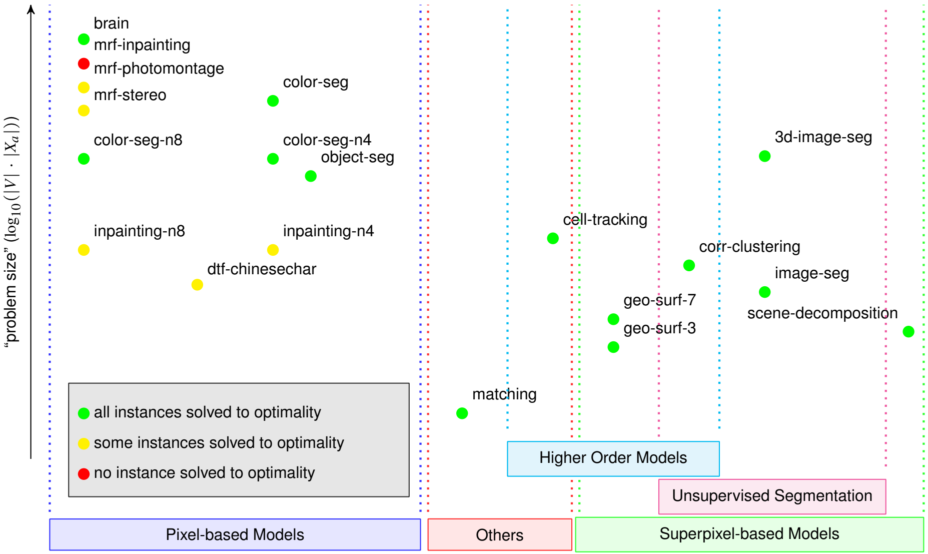
Solution

Figure : Red pixels mark nodes that need to be labeled by an ILP solver.

OpenGM Models: w/o our results



OpenGM Models: with our results



Conclusions and Future Work

Our approach

- does efficient extraction of the complex, combinatorial subproblem;
- is generic: allows almost any combination of LP and ILP solvers;
- makes the problems, which are easy in practice, easy in theory.

Limitations:

- sparse graphs;
- LP relaxation is *almost* tight.

Future work:

- Alternative and specialized solvers for LP and ILP.
- Higher order models.
- Tighter convex relaxations.

Proof of the Main Theorem

Definition $\mathcal{V}_\mathcal{A} \subset \mathcal{V}$, $\mathcal{V}_\mathcal{C} = \{v \in \mathcal{V}_\mathcal{A} \colon \exists uv \in \mathcal{E}_\mathcal{G} \colon u \in \mathcal{V}_\mathcal{G} \backslash \mathcal{V}_\mathcal{A}\}$, $\mathcal{V}_\mathcal{B} = \mathcal{V}_\mathcal{C} \cup (\mathcal{V}_\mathcal{G} \backslash \mathcal{V}_\mathcal{A})$, $\mathcal{Q} = (\mathcal{V}_\mathcal{Q}, \mathcal{E}_\mathcal{Q})$, $\mathcal{E}_\mathcal{Q} = \{uv \in \mathcal{E}_\mathcal{G} \colon u, v \in \mathcal{V}_\mathcal{Q}\}$.

Theorem. Let

- $x^*_\mathcal{A}$ and $x^*_\mathcal{B}$ minimize the energy on $\mathcal{A}$ and $\mathcal{B}$ resp.,
- $x^*_\mathcal{A}|_\mathcal{C} = x^*_\mathcal{B}|_\mathcal{C}$,
- problem $\min_{x_\mathcal{A}} E_\mathcal{A}(x_\mathcal{A})$ is trivial.

Then $x^* = (x^*_\mathcal{A}, x^*_\mathcal{B}|_{\mathcal{B}\backslash\mathcal{C}})$ is optimal on $\mathcal{G}$.

Proof. $E_\mathcal{G}(x) \to E^\theta_\mathcal{G}(x)$

$$\theta'_w(x_w) := \begin{cases} 0, & w \in \mathcal{V}_\mathcal{C} \cup \mathcal{E}_\mathcal{C} \\ \theta_w(x_w), & w \notin \mathcal{V}_\mathcal{C} \cup \mathcal{E}_\mathcal{C} \end{cases}$$

$E^\theta_\mathcal{G}(x) = E^{\theta'}_\mathcal{A}(x_\mathcal{A}) + E^\theta_\mathcal{B}(x_\mathcal{B})$

$\min_{x_\mathcal{A}} E_\mathcal{A}(x_\mathcal{A})$ is trivial $\Rightarrow x^*_\mathcal{A} \in \arg\min_{x_\mathcal{A}} E^{\theta'}_\mathcal{A}(x)$

$$\min_x E^\theta_\mathcal{G}(x) = \{\min_{x_\mathcal{A}, x_\mathcal{B}} E^{\theta'}_\mathcal{A}(x_\mathcal{A}) + E^\theta_\mathcal{B}(x_\mathcal{B}) | \text{ s.t. } x_\mathcal{A}|_\mathcal{C} = x_\mathcal{B}|_\mathcal{C}\}$$

$$= \min_{x'_\mathcal{C}} \min_{x_\mathcal{A} \colon x_\mathcal{A}|_\mathcal{C} = x'_\mathcal{C}} E^{\theta'}_\mathcal{A}(x_\mathcal{A}) + \min_{x_\mathcal{B} \colon x_\mathcal{B}|_\mathcal{C} = x'_\mathcal{C}} E^\theta_\mathcal{B}(x_\mathcal{B})$$

$$\geqslant \min_{x_\mathcal{A}} E^{\theta'}_\mathcal{A}(x_\mathcal{A}) + \min_{x_\mathcal{B}} E^\theta_\mathcal{B}(x_\mathcal{B}) = E^{\theta'}_\mathcal{A}(x^*_\mathcal{A}) + E^\theta_\mathcal{B}(x^*_\mathcal{B}) = E^\theta_\mathcal{G}(x^*)$$