Ph.D. Thesis

# Exact and Partial Energy Minimization in Computer Vision

Alexander Shekhovtsov

Supervisor: Václav Hlaváč
Supervisor-specialist: Tomáš Werner

Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics

Ph.D. programme: Electrical Engineering and Information Technology
Branch of study: Mathematical Engineering, 3901V021

# Overview

- Discrete Optimization in Computer Vision (Energy Minimization)
  - Cases Reducible to Minimum Cut
    - Contribution: Distributed mincut/maxflow algorithm
  - General NP-hard case
    - Contribution: Methods to find a Part of Optimal Solution

# Minimum Cut Problem

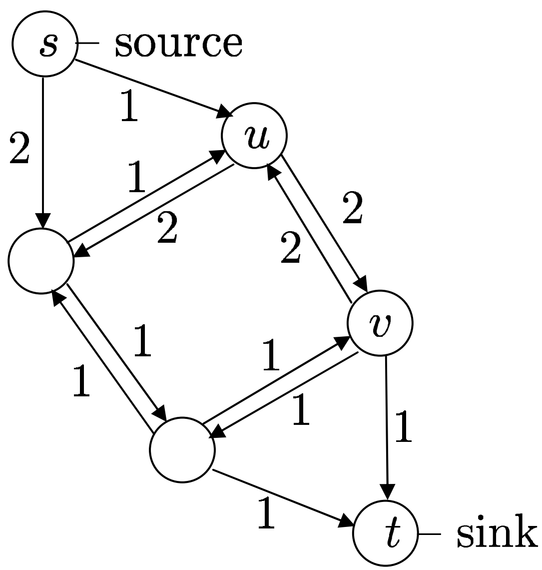


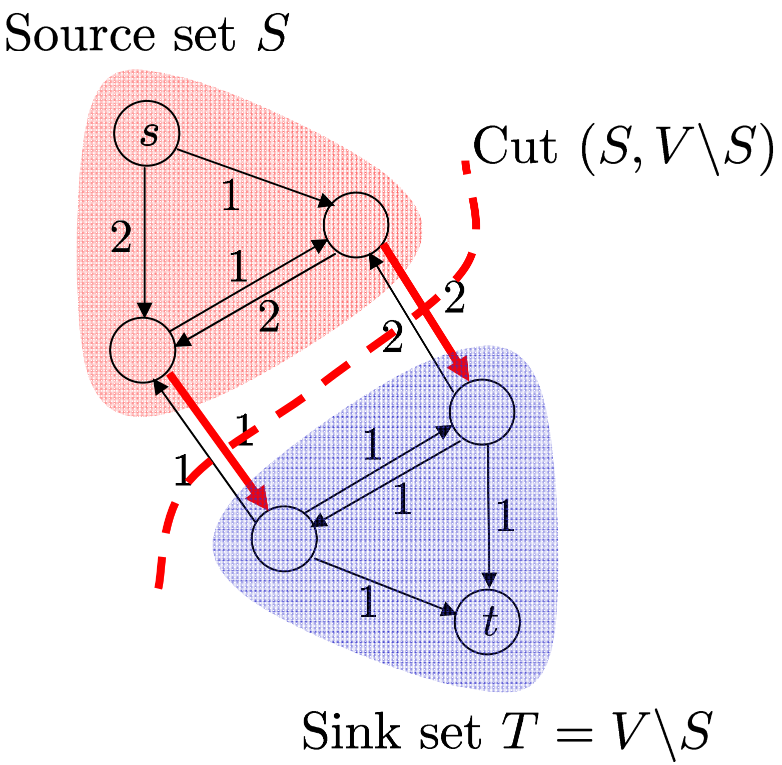
Capacitated network

$G = (V, E, c)$,

$c(u, v) \geq 0$ – arc capacities

Cut cost: $\sum_{\substack{(u,v)\in E \\ u\in S \\ v\notin S}} c(u, v) \to \min_{\substack{S \\ s\in S \\ t\notin S}}$



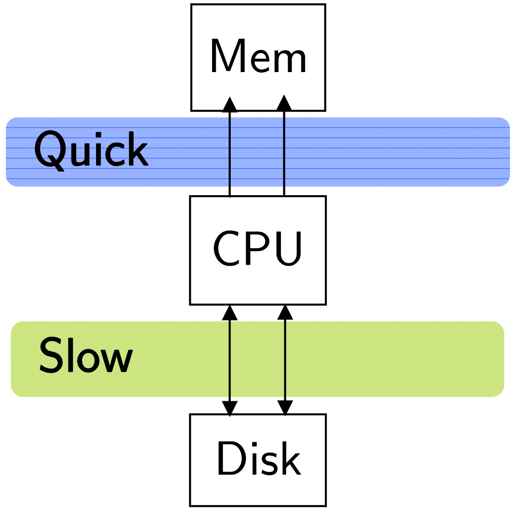



## Introduction



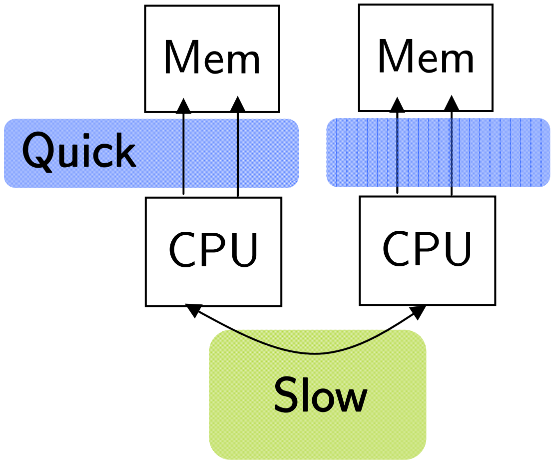
- Distributed Model – Divide Computation AND Memory

Distributed Sequential




Solve large problem on a single computer

Distributed Parallel




on more computers

Split data in parts:

# Introduction



- Sequential Algorithms

  $n = |V|$ – number of vertices

  $m = |E|$ – number of edges

  Orlin (2012), Max flows in $O(nm)$ time or better

  In case $m = O(n)$, in $O(n^2/\log n)$ time

- Parallel Algorithms

  parallel push-relabel, Goldberg (1987) and

  Cheriyan and Maheswari (1988)

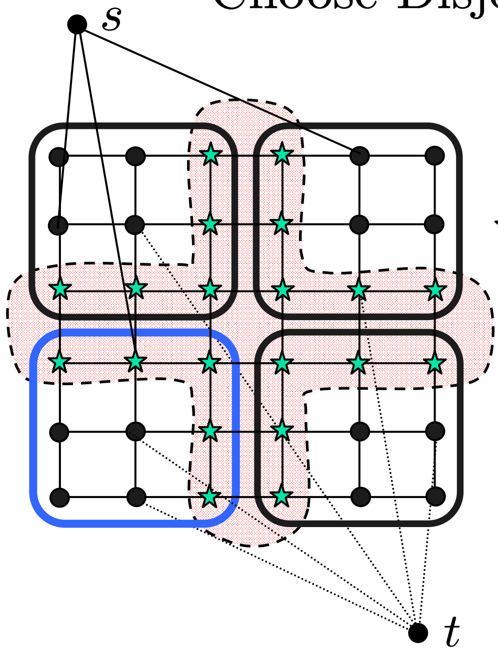
  $O(n^2)$ time using $O(n)$ processors and $\boxed{O(n^2\sqrt{m})}$ messages

Our goal is to improve on this

# Main Result



Choose Disjoint Regions

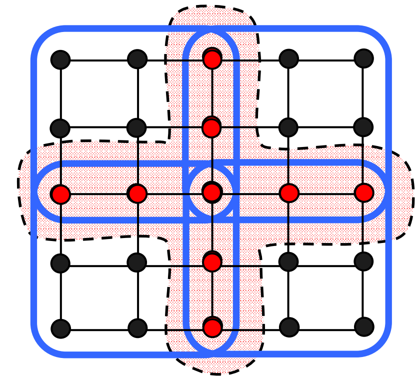



★ boundary vertices $\mathcal{B}$

boundary arcs $E_{\mathcal{B}}$

Alternative Construction:

Choose separator set ●



Overlapping Regions

Only inter-region messages count

Push-relabel approaches (different variants) need $\Omega(n^2)$ messages

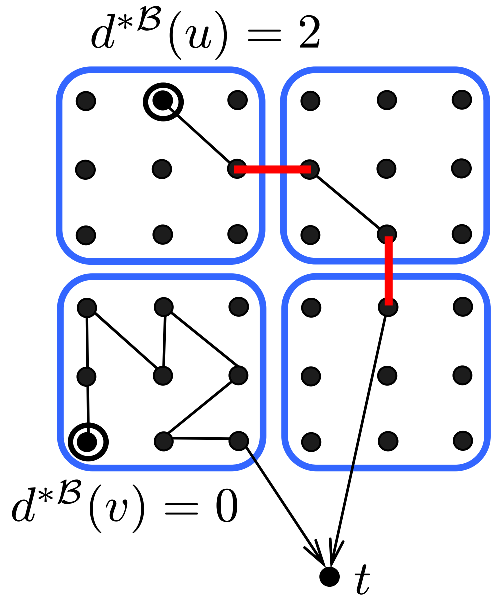
(Proportional to the full problem size)

- Main Result: New Algorithm uses $O(|\mathcal{B}|^2|E_{\mathcal{B}}|)$ messages
  (Proportional to the size of the separator set)

# Main Idea



- New distance function




length of the path = number of boundary edges

distance = length of a shortest path to the sink

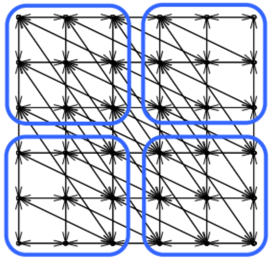
corresponds to costly operations

Algorithm: push-relabel between regions, augmenting path inside regions

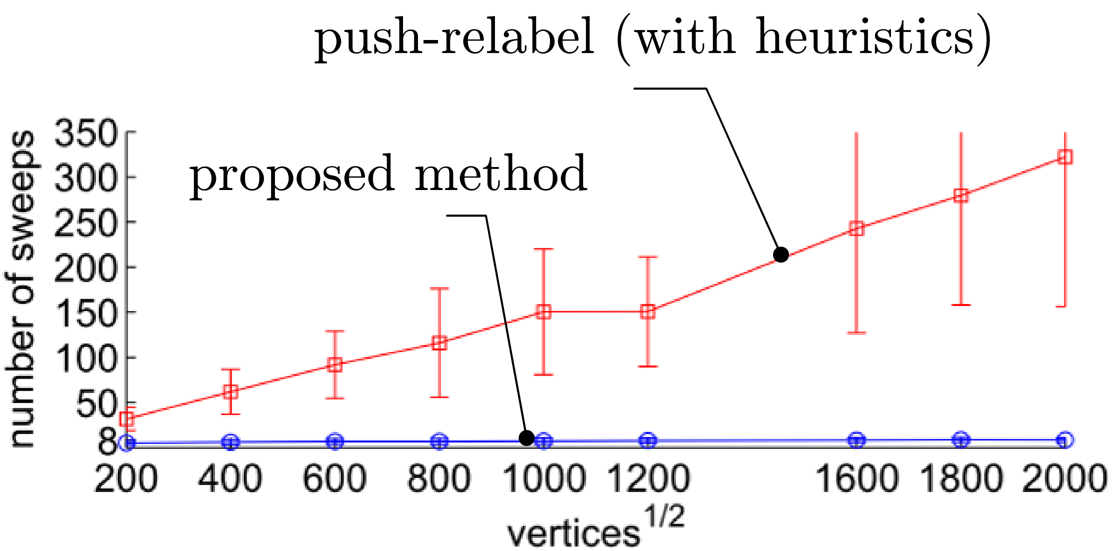
# Experimental Confirmation



Synthetic instances: Grid graph with random capacities, partitioned into 4 regions



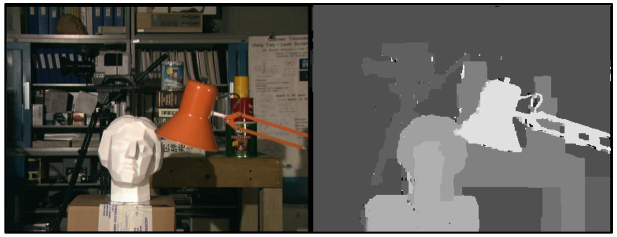
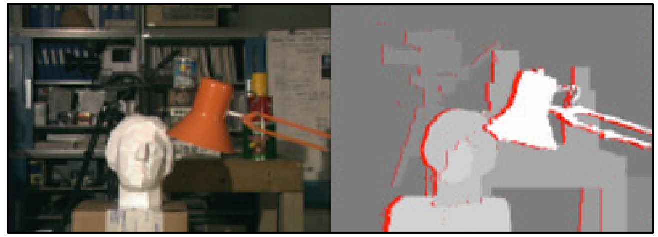
sweep = synchronously send messages on all boundary arcs



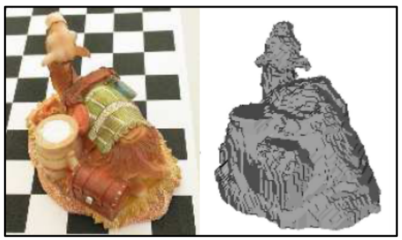
# Instances in Computer Vision



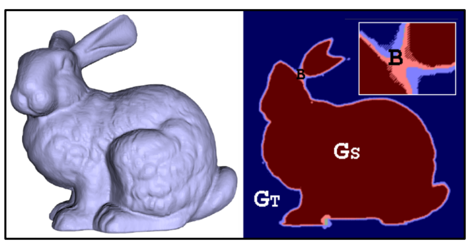
- Dataset Published by Vision Group at University of Western Ontario

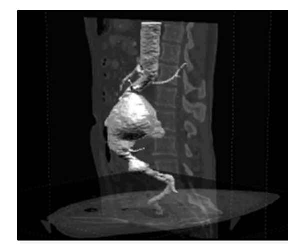
Stereo
Boykov et al. 1998
Kolmogorov and Zabih 2001

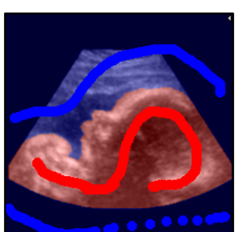
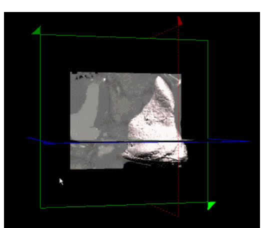
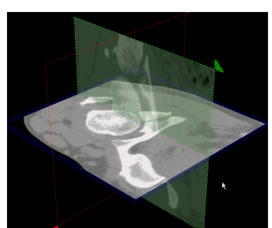



Multiview Reconstruction
Lempitsky et al. 2006
Boykov and Lempitsky 2006

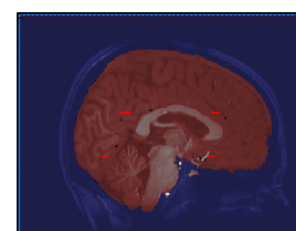



Surface Fitting
Lempitsky and Boykov 2007



3D Segmentation
Boykov and Joly 2001
Boykov and Funka-Lea 2006
Boykov and Kolmogorov 2003

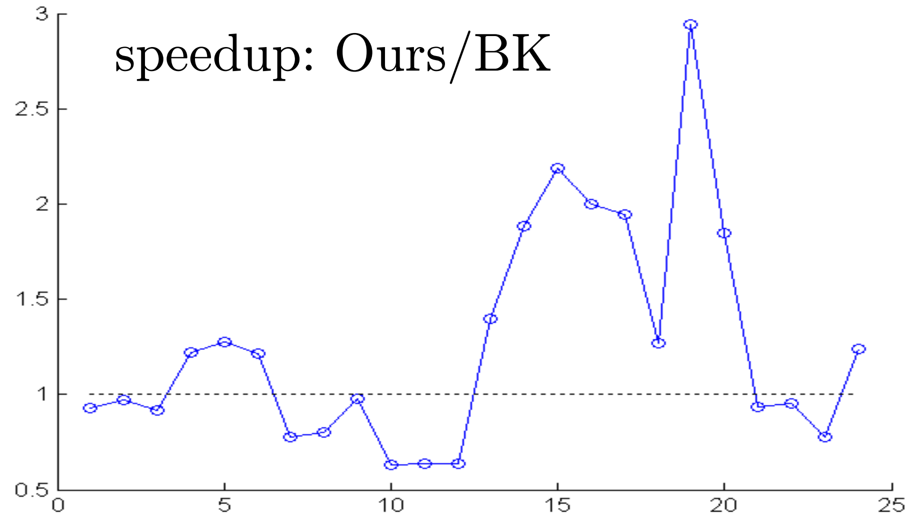


# Sequential Variant for Limited Memory Model

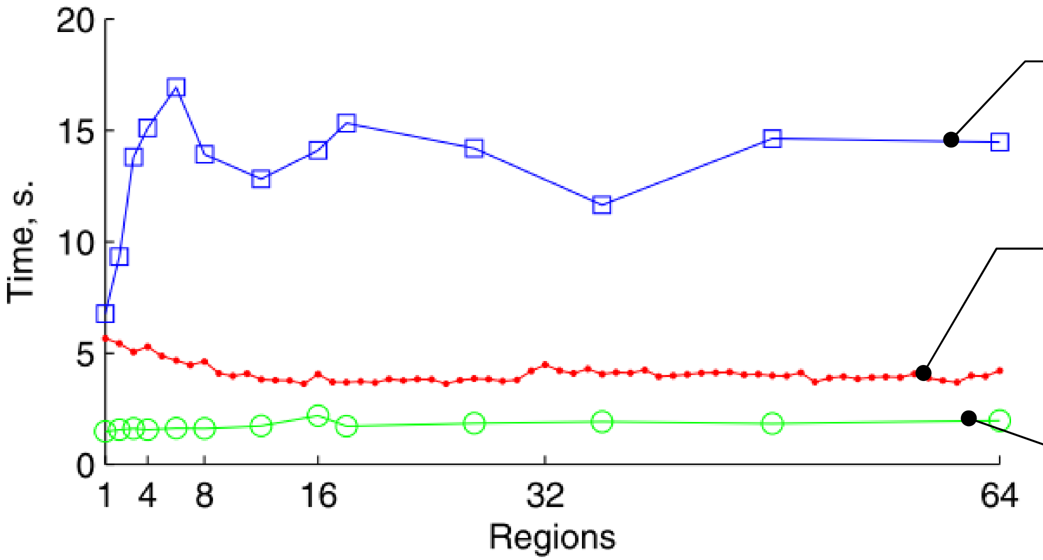


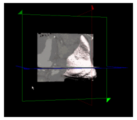
- uses BK (Boykov and Kolmogorov) inside regions
- sometimes faster than BK

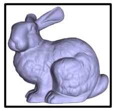
(CPU time excluding I/O)




- robust over partition size



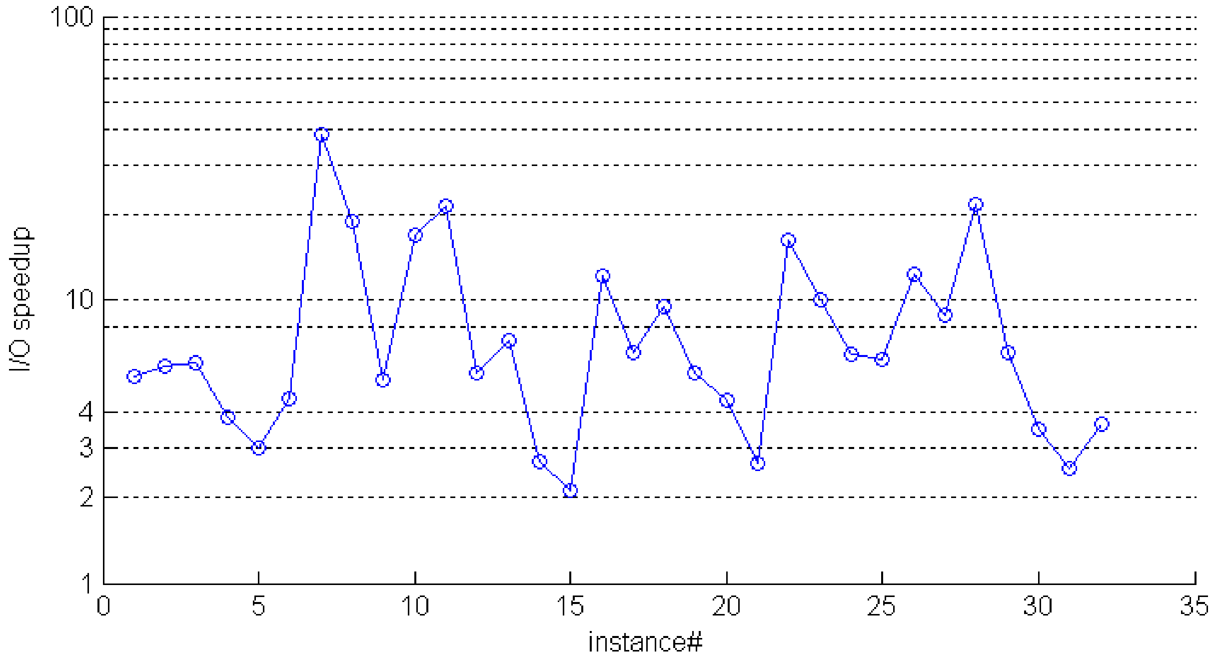





# Sequential Variant for Limited Memory Model

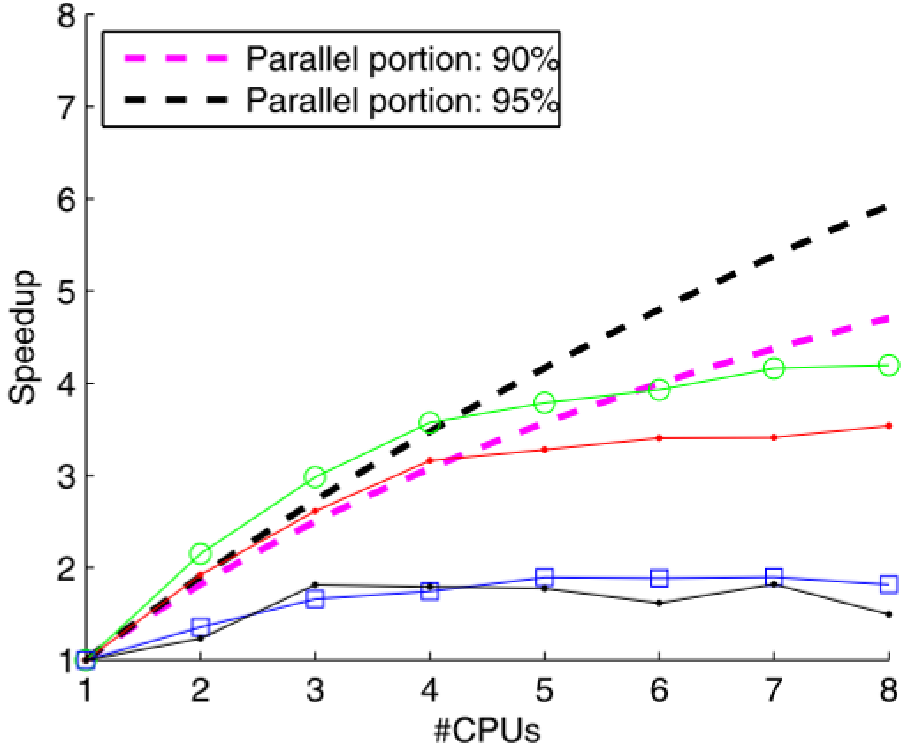


- Messages (sweeps) speedup over push-relabel
  (distributed version of Delong and Boykov 2008)



# Parallel Variant

- Competitive with shared memory model methods
- Speedup bounded by memory bandwidth




| problem | regions | MEM | 1CPU | 8CPUs |
|---|---|---|---|---|
| BL06-gargoyle-lrg | 16 | 1.6GB | 137.4s | 38.82s |
| BL06-camel-lrg | 16 | 1.8GB | 67.14s | 16.00s |
| LB07-bunny-lrg | 64 | 12GB | 12.94s | 7.11s |
| liver.n6c100 | 64 | 0.9GB | 18.41s | 12.21s |

## Conclusion



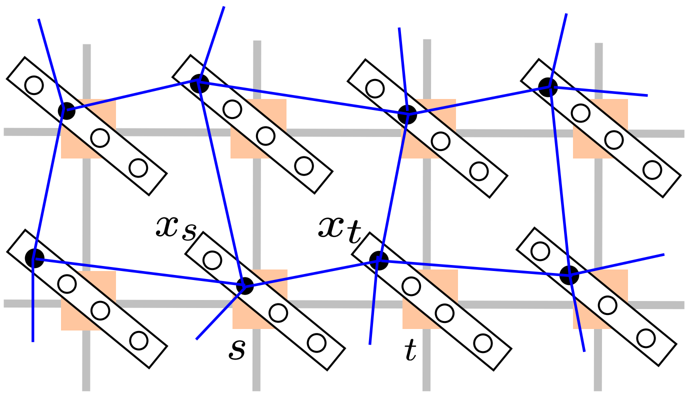
- New distributed algorithm
- Terminates in at most $B^2$+1 sweeps (few in practice)

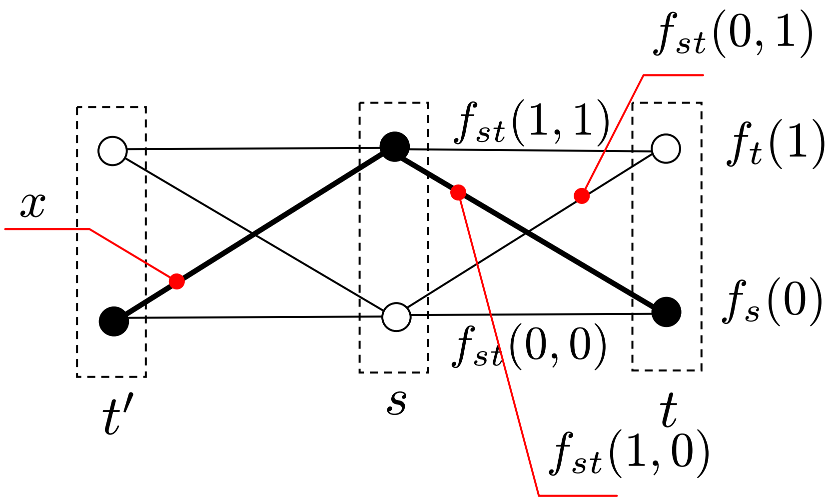
- Sequential Algorithm
  1) competitive with sequential solvers
  2) uses few sweeps (= loads/unloads of regions)
  3) suitable to run in the limited memory model

- Parallel Algorithm
  1) competitive with shared memory algorithms
  2) uses few sweeps (= rounds of message exchange)
  3) suitable for execution on a computer cluster

- Implementation can be specialized for regular grids (less memory/faster)
- (?) no good worst case complexity bound in terms of elementary operations

# Discrete Energy Minimization Problem




$$E_f(x) = f_0 + \sum_{s \in \mathcal{V}} f_s(x_s) + \sum_{st \in \mathcal{E}} f_{st}(x_s, x_t)$$

Minimize over labelings (assignments) $x = (x_s \in \mathcal{L}_s \mid s \in \mathcal{V})$

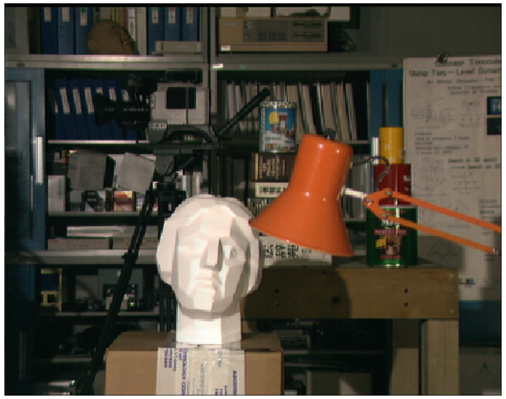
Search space $\mathcal{L} = \prod_s \mathcal{L}_s$

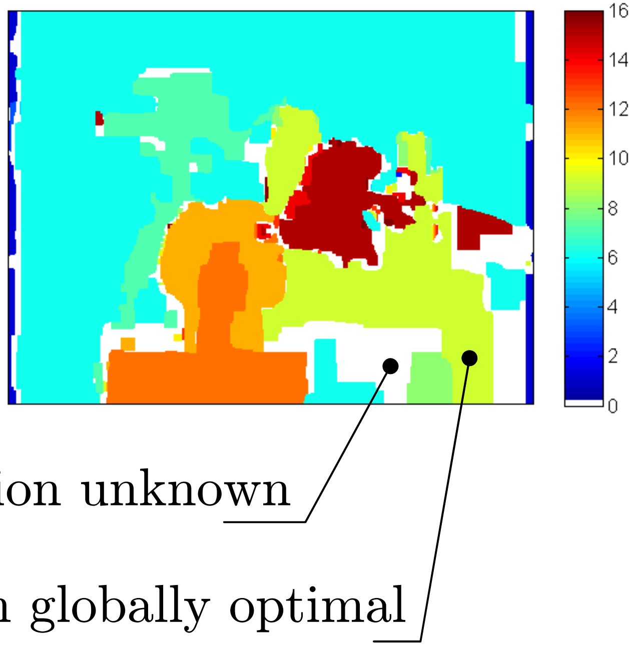
# Partial Optimality



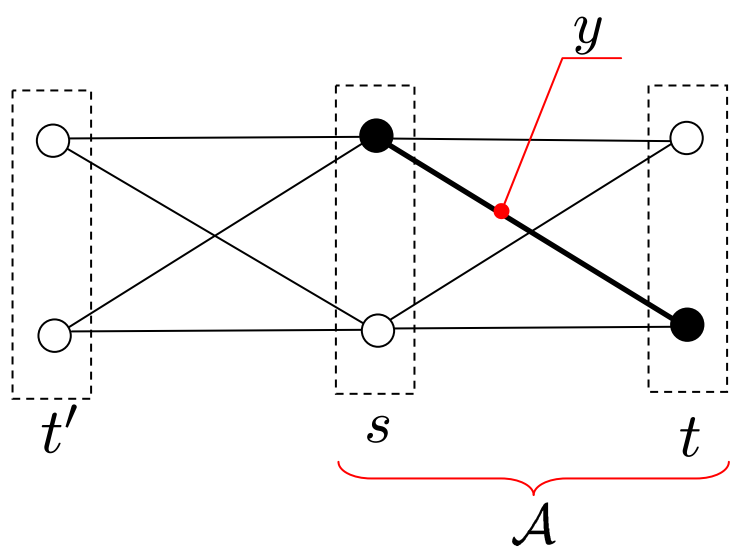
Energy model for stereo, minimization NP-hard

Find optimal solution in some pixels





## Partial Optimality



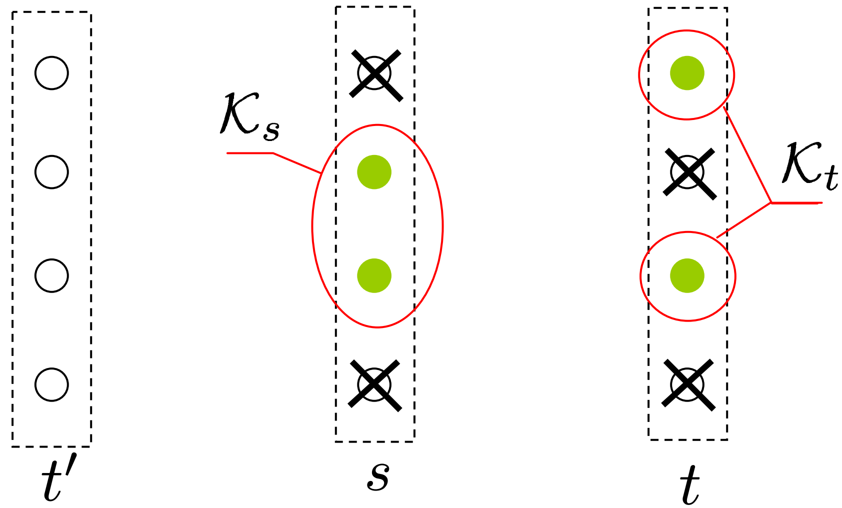

Partial assignment $(\mathcal{A}, y)$, $\mathcal{A} \subset \mathcal{V}$, $y = (y_s \mid s \in \mathcal{A})$:

- Strong optimal
  for **any** minimizers $x^*$: $\quad x^*_{\mathcal{A}} = y$
- Weak optimal
  **exists** minimizer $x^*$: $\quad x^*_{\mathcal{A}} = y \quad$ ($y$ is extendible to a minimizer)

! Verifying weak/strong optimality of a partial assignment is NP-hard

# Partial Optimality (Multilabel)






Domain reduction $K = (K_s \subset \mathcal{L}_s \mid s \in \mathcal{V})$:

- Strong optimal
  for **any** minimizers $x^*$: $\quad x^*_s \in \mathcal{K}_s$
- Weak optimal
  **exists** minimizer $x^*$: $\quad x^*_s \in \mathcal{K}_s$

(remark: can be reformulated as a partial assignment)

**!** Verification is NP-hard

## Overview

**Context:**

- Verification of partial optimality is NP-hard
- There are methods that find some partial optimality
- How this is possible in principle?
- They use different sufficient conditions

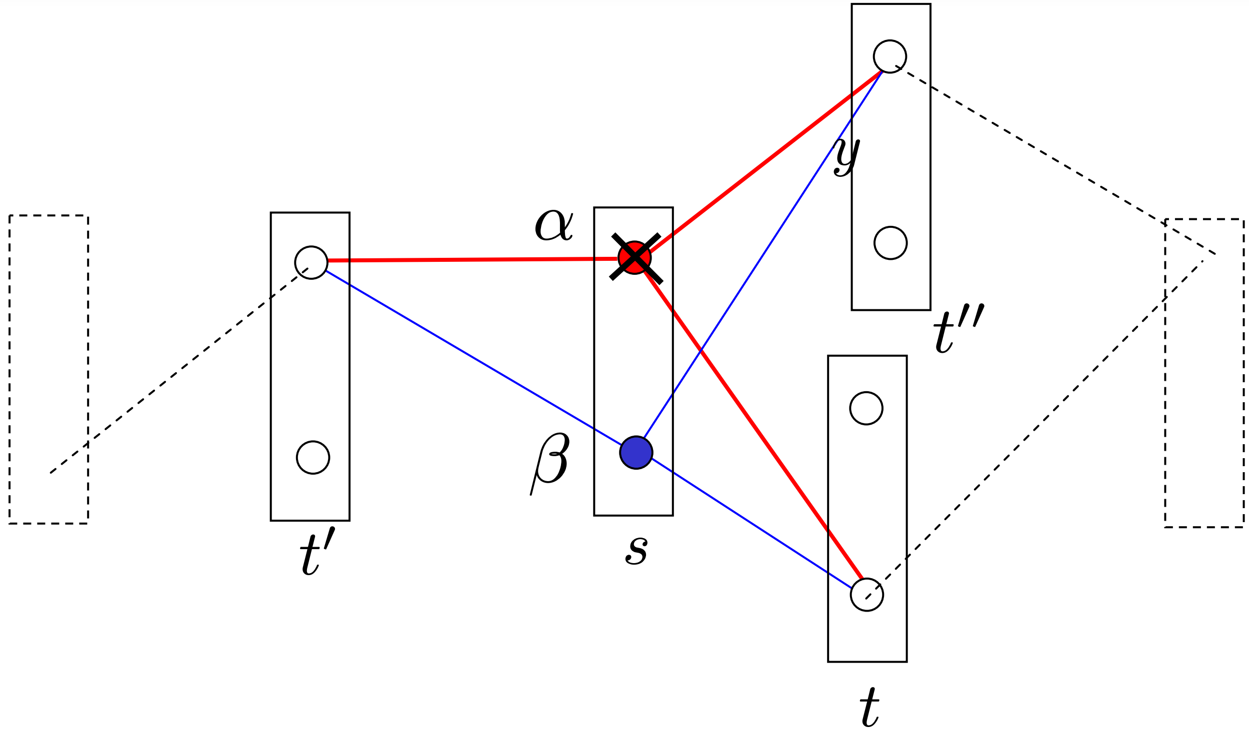
**Contribution:**

- Introduce new sufficient condition verifiable in polynomial time
- Show that it includes different conditions in the literature as special cases.

* Methods for finding the largest partial assignment (identifiable by the proposed condition)

## Dead End Elimination






$$f_s(\alpha) - f_s(\beta) + \sum_{t \in \mathcal{N}(s)} \min_{x_t \in \mathcal{L}_t} [f_{st}(\alpha, x_t) - f_{st}(\beta, x_t)] \geq 0$$

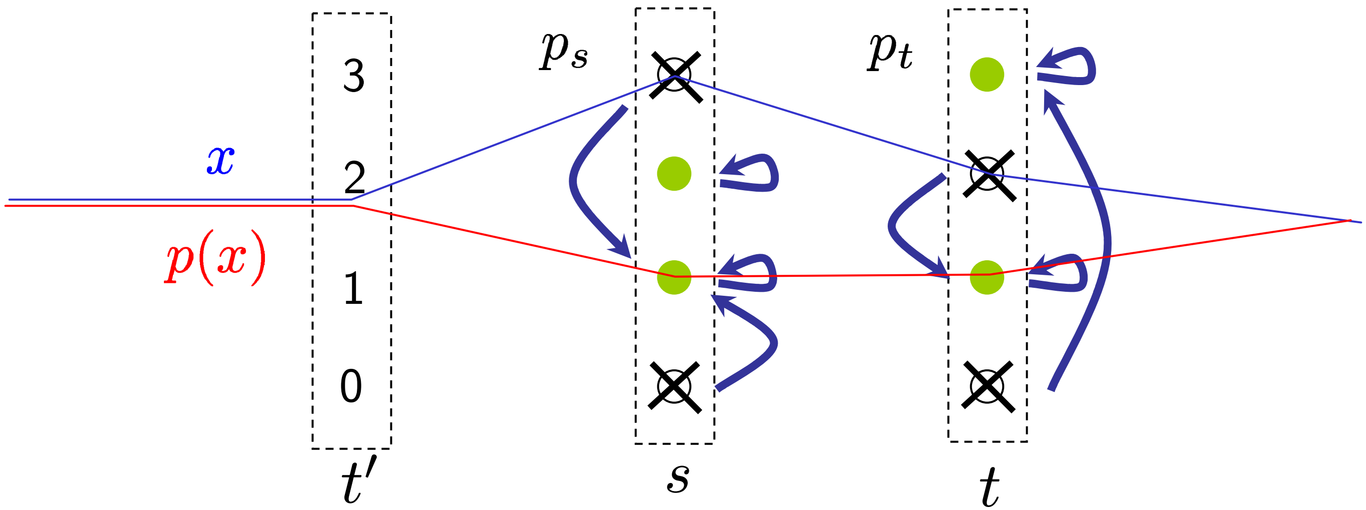
worst case of energy change when replacing $\beta$ with $\alpha$

$\Rightarrow$ eliminate $\alpha$

Desmet et al. (1992), Goldstein (1994), Lasters et al. (1995),Pierce et al. (2000), Georgiev et al. (2006)

## Improving Mapping





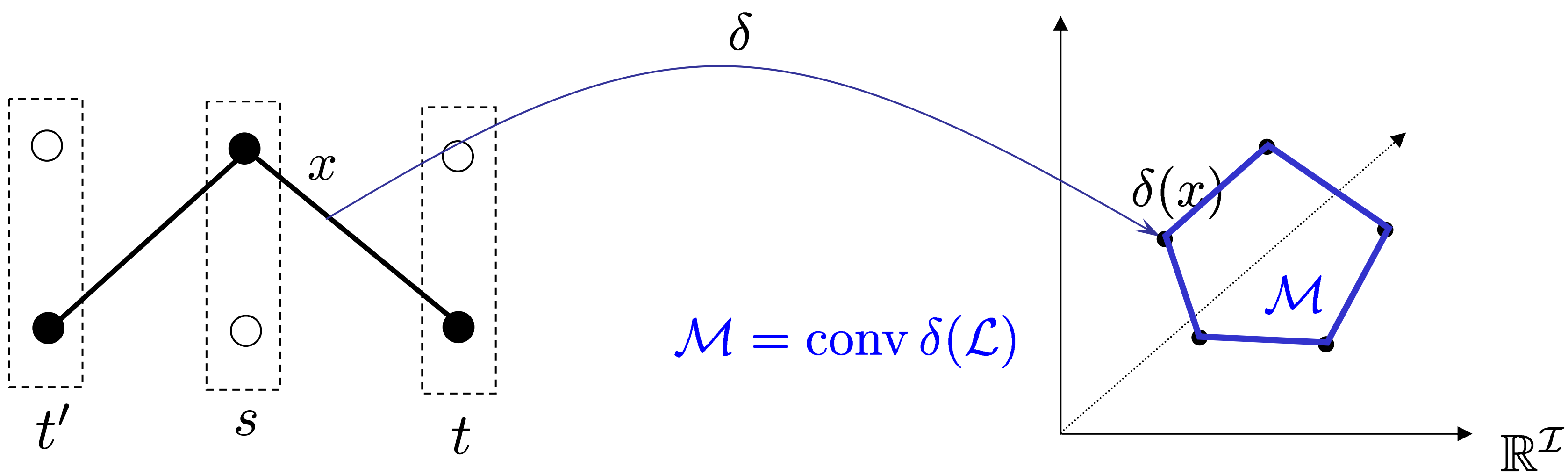

Let $p$ be an improving mapping:

$$(\forall x \in \mathcal{L}) \qquad E_f(p(x)) \le E_f(x) \qquad (*)$$

Then exists a minimizer $x^*$ in $p(\mathcal{L})$

! Verification of improving property $(*)$ is still NP-hard

## Linear Embedding




Embed labelings into Eucledian space $\mathbb{R}^{\mathcal{I}}$ to linearize $E_f$

$$E_f(x) = f_0 + \sum_{s\in\mathcal{V}} f_s(x_s) + \sum_{st\in\mathcal{E}} f_{st}(x_s, x_t) = \langle f, \delta(x)\rangle_{\mathbb{R}^{\mathcal{I}}}$$

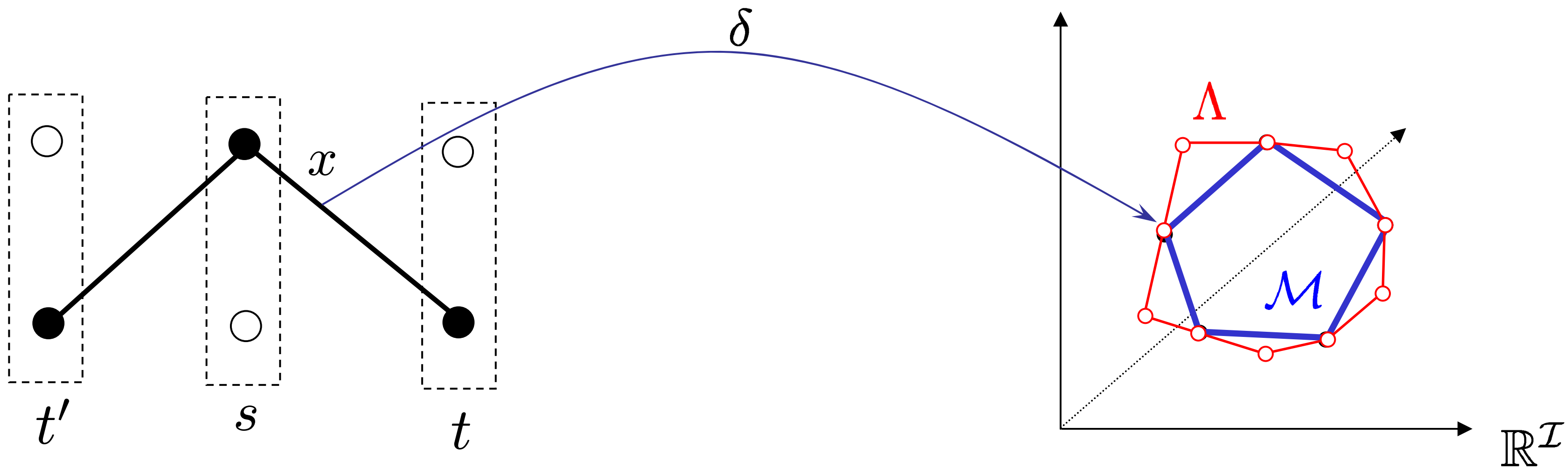
$$\delta(x)_0 = 1$$

$$\delta(x)_s(i) = [\![x_s = i]\!] \longrightarrow f_s(x_s) = \textstyle\sum_i f_s(i)\delta(x)_s(i)$$

$$\delta(x)_{st}(i,j) = [\![x_s = i]\!][\![x_t = j]\!]$$

$$\mathcal{I} = \{0\} \cup \{(s,i) \mid s \in \mathcal{V},\ i \in \mathcal{L}_s\} \cup \{(st, ij) \mid st \in \mathcal{E},\ ij \in \mathcal{L}_{st}\}.$$

## Linear Embedding




$$E_f(x) \longrightarrow \langle f, \delta(x) \rangle$$

$$\mathcal{L} \longrightarrow \delta(\mathcal{L})$$

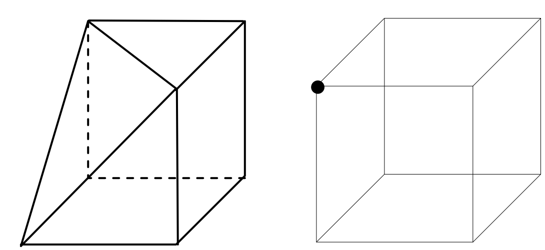
**LP-relaxation**

$$\min_x E_f(x) \longrightarrow \min_{\mu \in \delta(\mathcal{L})} \langle f, \mu \rangle = \underbrace{\min_{\mu \in \operatorname{conv} \delta(\mathcal{L})}}_{\mathcal{M}} \langle f, \mu \rangle \geq \min_{\mu \in \Lambda} \langle f, \mu \rangle$$

$$\Lambda = \operatorname{aff}(\mathcal{M}) \cap \mathbb{R}^{\mathcal{I}}_{+}$$

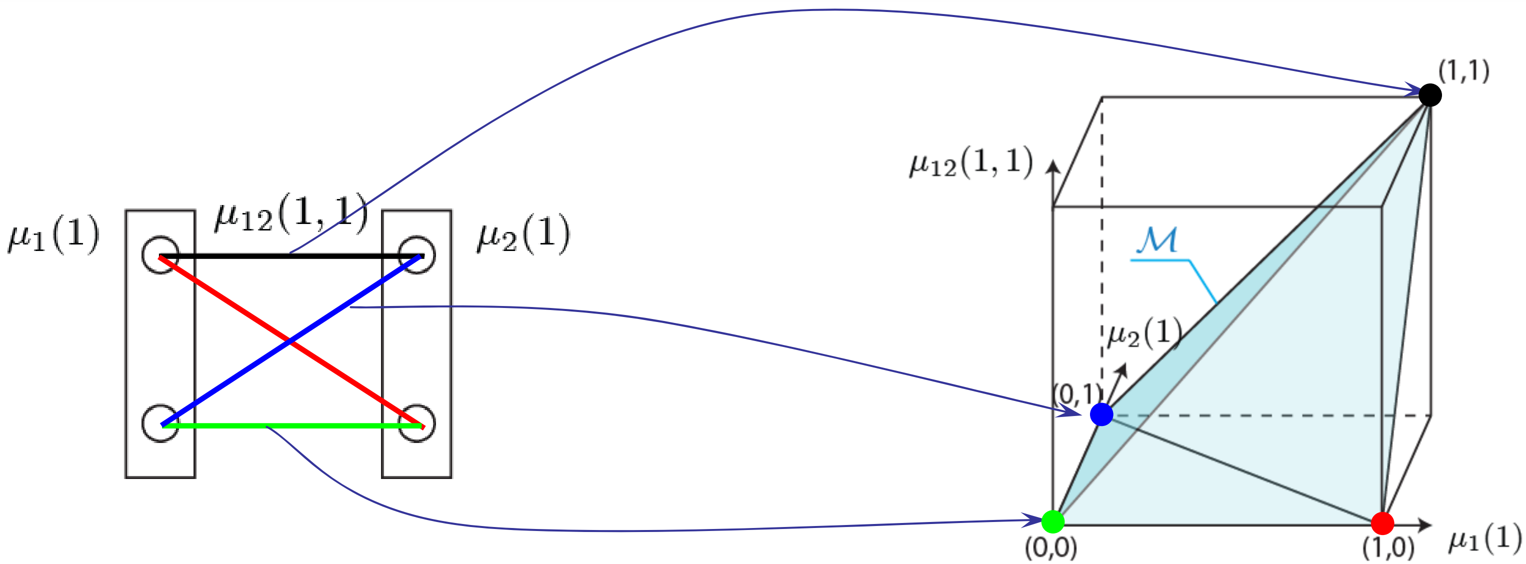
remark:

- more vertices but fewer facets



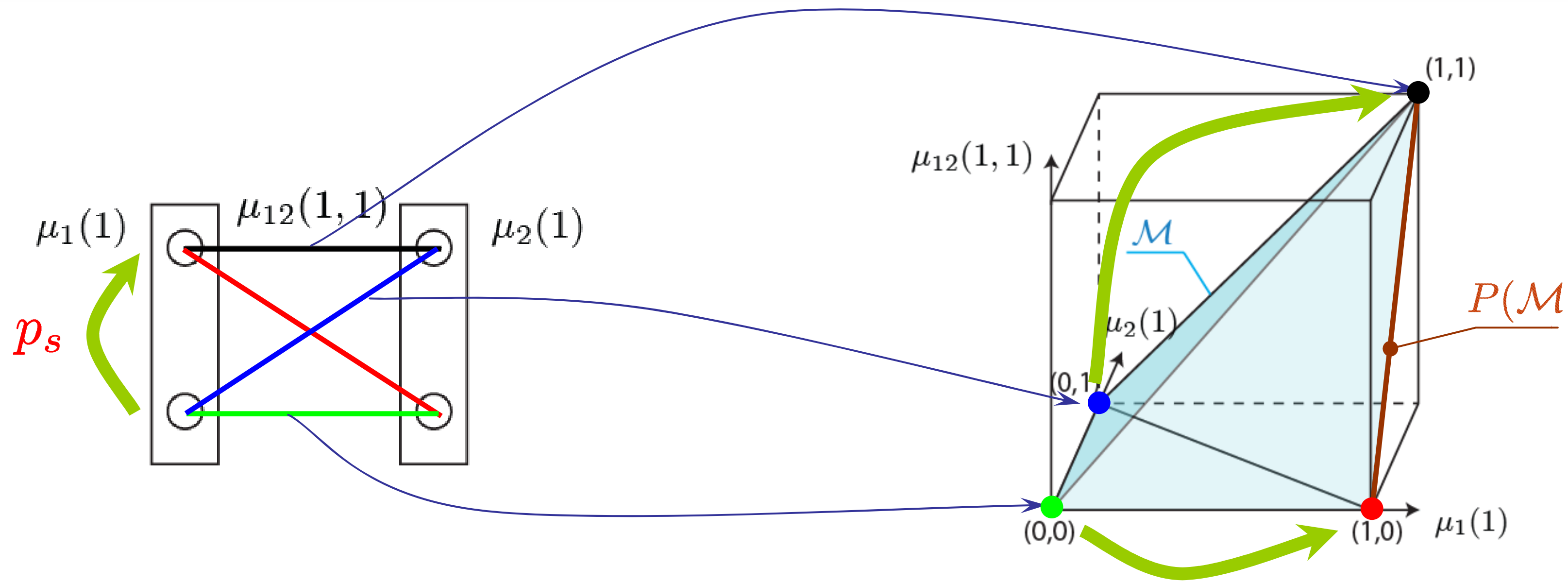
# Linear Embedding of Maps





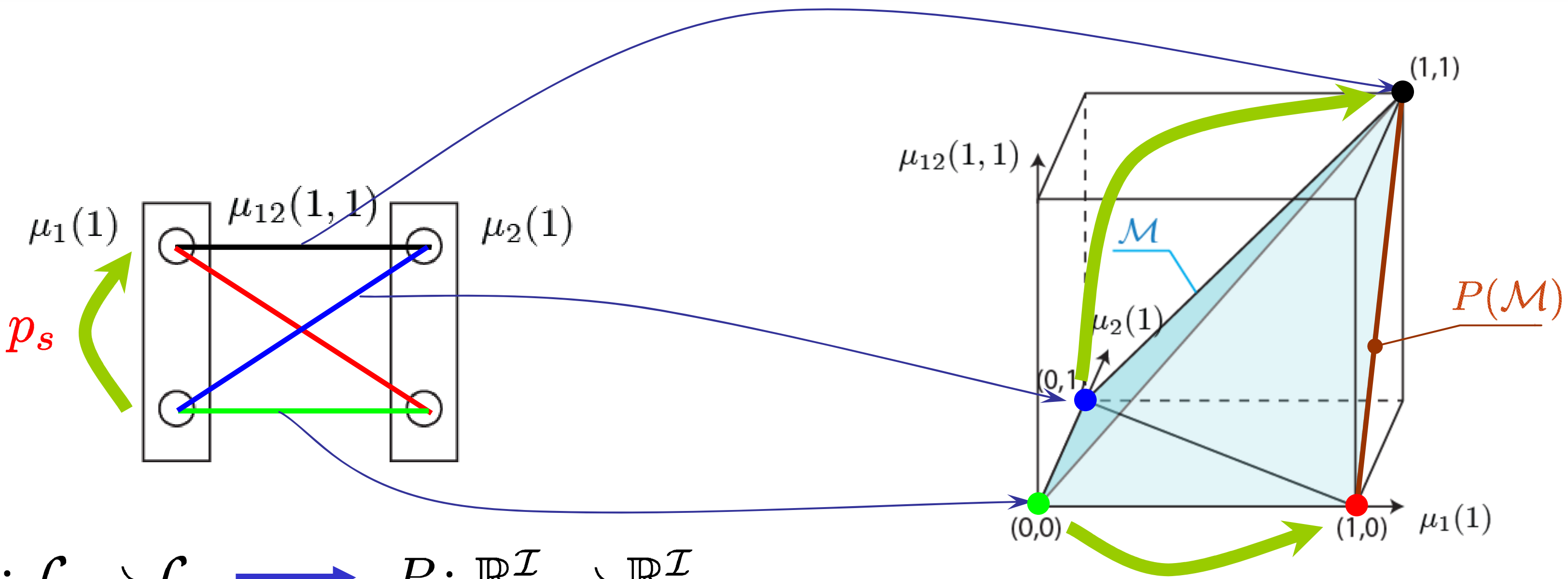

$x \longrightarrow \mu = \delta(x) \in \mathbb{R}^9$

# Linear Embedding of Maps



# Linear Embedding of Maps






$p\colon \mathcal{L} \to \mathcal{L}$ $\longrightarrow$ $P\colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$

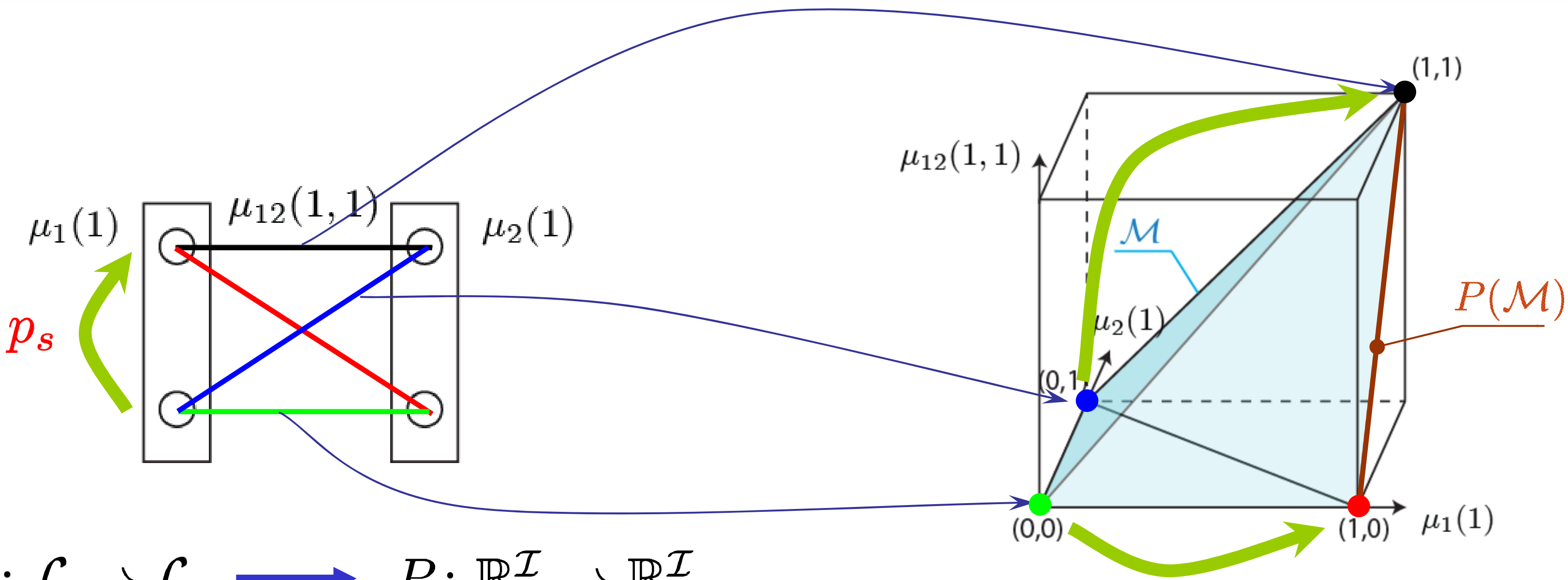
$p$ idempotent $\longrightarrow$ $P^2 = P$, **projection**

Improving property:
$(\forall x \in \mathcal{L}) \quad E_f(p(x)) \leq E_f(x)$ $\quad\Longleftrightarrow\quad$ $(\forall \mu \in \mathcal{M}) \quad \langle f, P\mu \rangle \leq \langle f, \mu \rangle$

Sufficient condition, $\Lambda$-improving: $\quad (\forall \mu \in \Lambda) \quad \langle f, P\mu \rangle \leq \langle f, \mu \rangle$

**!** Verifiable: $\quad \min\limits_{\mu \in \Lambda} \langle f, (I - P)\mu \rangle \geq 0$

## Linear Embedding of Maps




$p\colon \mathcal{L} \to \mathcal{L}$ ⟶ $P\colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$

$p$ idempotent ⟶ $P^2 = P$, **projection**

Improving property:

$(\forall x \in \mathcal{L}) \quad E_f(p(x)) \leq E_f(x) \quad \Leftrightarrow \quad (\forall \mu \in \mathcal{M}) \quad \langle f, P\mu \rangle \leq \langle f, \mu \rangle$

Sufficient condition, $\Lambda$-improving: $(\forall \mu \in \Lambda) \quad \langle f, P\mu \rangle \leq \langle f, \mu \rangle$

**!** Verifiable: $\min\limits_{\mu \in \Lambda} \langle f, (I - P)\mu \rangle \geq 0$

## More General Projections/Maps



Projection $P\colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$

defined pixel-wise, by matricies $P_s$

$$1^\mathsf{T} P_s = 1$$

$$P_s \geq 0$$

Condition:

$$(\forall \mu \in \boldsymbol{\Lambda}) \quad \langle f, P\mu \rangle \leq \langle f, \mu \rangle$$

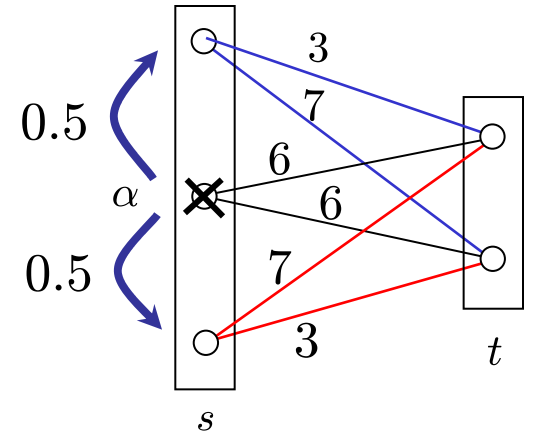
Verifiable:

$$\min_{\mu \in \Lambda} \langle (I - P^\mathsf{T}) f, \mu \rangle \geq 0$$

Provides problem reduction:

$$\min_x E_f(x) = \min_{\mu \in \mathcal{M}} \langle f, \mu \rangle = \min_{\mu \in P(\mathcal{M})} \langle f, \mu \rangle$$

Example of fractional map




$$P_s = \begin{pmatrix} 1 & 0.5 & 0 \\ 0 & 0 & 0 \\ 0 & 0.5 & 1 \end{pmatrix} \qquad P_t = I$$

## $\mathit{\Lambda}$-Improving Characterization



Let $P$ be a pixel-wise projection

satisfying (global) improving property:

$$(\forall \mu \in \mathbf{\Lambda}) \qquad \langle f, P\mu\rangle \leq \langle f, \mu\rangle$$

**Characterization Theorem:** Exists equivalent function $g$

$$(\forall x) \qquad E_f(x) = E_g(x)$$

for which $P$ satisfies local improving property:

$$(Pg)_s(i) \geq g_s(i)$$

$$(Pg)_{st}(i,j) \geq g_{st}(i,j)$$

## Special Cases

Methods that can be explained by the proposed condition:

- DEE conditions by Desmet (1992) and Goldstein (1994)
- (Weak/strong) Persistency in Quadratic Pseudo-Boolean Optimization (QPBO) by Nemhauser & Trotter (1975), Hammerr et al. (1984), Boros et al. (2002)
- Multilabel QPBO Kohli et al. (2008), Shekhovtsov et al. (2008)
- Submodular Auxiliary problems by Kovtun (2003, 2010)

* Iterative Pruninig by Swoboda et al. (2013)

Common properties, Only (M)QPBO was previously related to LP relaxation

# Maximum $\Lambda$-Improving Projections

- **Problem:** Find the mapping that maximizes domain reduction

$$\min_p \sum_{s,i} |p_s(\mathcal{L}_s)| \qquad \text{subject to} \qquad p \in \mathrm{WI}(\Lambda, f) \qquad \text{MAX-WI}$$

(weakly $\Lambda$-improving maps)

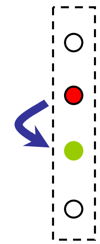
$$\text{subject to} \qquad p \in \mathrm{SI}(\Lambda, f) \qquad \text{MAX-SI}$$

(strictly $\Lambda$-improving maps)

# Maximum $\Lambda$-Improving Projections

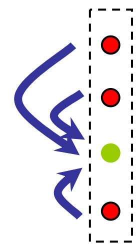
- Thesis

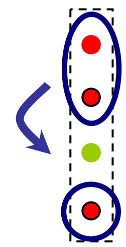
| problem | MAX-SI | MAX-WI |
|---|---|---|
| $\lvert\mathcal{L}_s\rvert = 2$ | P [1] | P [2] |
| 2-label maps | P | P |



* Follow-up work, submitted to CVPR

| problem | MAX-SI | MAX-WI |
|---|---|---|
| $\lvert\mathcal{L}_s\rvert > 3$ | NP | NP |
| subset-to-one | P | P |
| all-to-one unknown | P | NP |





[1] Nemhauser & Trotter (1975), Hammerr et al. (1984), Boros et al. (2002)

[2] Picard & Queyranne (1977) (Vertex Packing)

# Conclusions

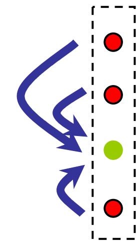
## Higher Order

- Many methods for higher order 0-1 problems not covered in this work:

-Adams, W. P., Lassiter, J. B., and Sherali, H. D. (1998). Persistency in 0-1 polynomial programming.
-Kolmogorov, V. (2012). Generalized roof duality and bisubmodular functions.
-Kahl, F. and Strandmark, P. (2012). Generalized roof duality.
-Lu, S. H. and Williams, A. C. (1987). Roof duality for polynomial 0-1 optimization.
-Ishikawa, H. (2011). Transformation of general binary MRF minimization to the first-order case.
-Fix, A. et al. (2011). A graph cut algorithm for higher-order Markov random fields.

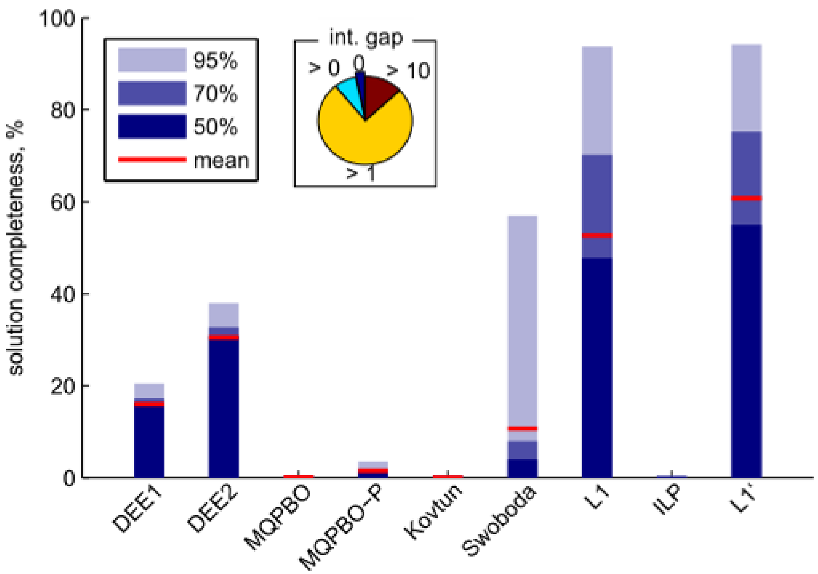
# Follow-up Work

- Algorithm proposed:
  - subset-to-one maps
  - MAX-WI reduced to a linear program (L1)

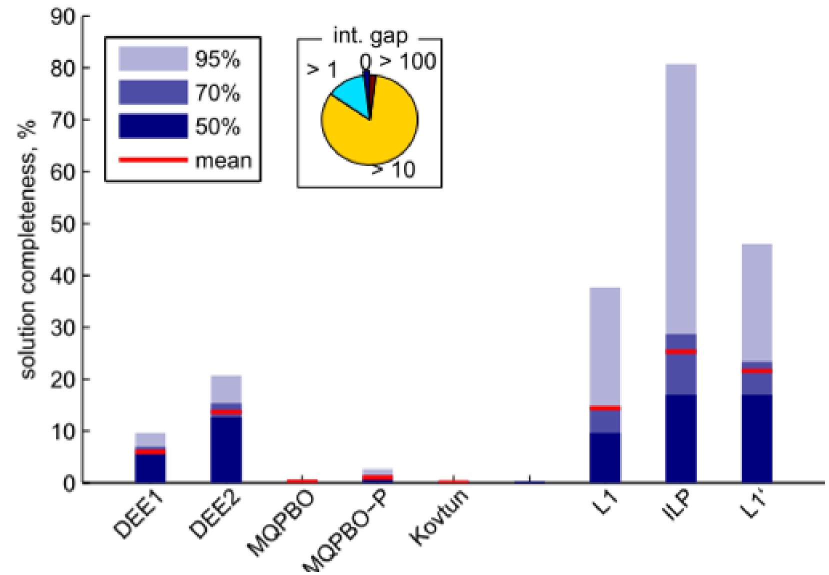


- Experiments: solution completeness on random problems

Generalized Potts (5 labels)




Fully Random (4 labels)

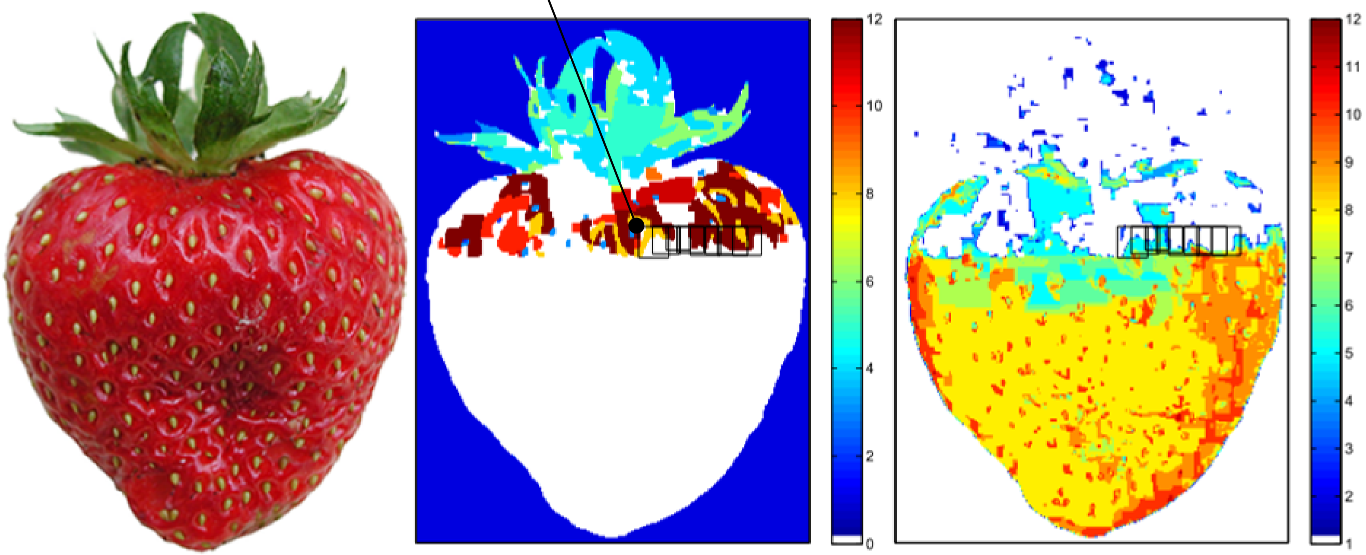


## Follow-up Work

- Experiments: solving large scale problems by parts

Restrict the method to a local window

Find globally optimal reduction




partial labeling

# remaining labels