Model fitting and regularization

(discrete optimization approach)

Yuri Boykov
Overview

- Label costs (high-order sparsity prior)
- Model fitting
  - dealing with continuum of labels
  - $K$-means and EM + regularization
- Applications
  - unsupervised image segmentation, compression
  - geometric model fitting (lines, circles, planes, homographies, motion,…)
Potts model \[ E(L) = \sum_p (L_p - I_p)^2 + \sum_{(p,q) \in N} V(L_p, L_q) \]
(piece-wise constant labeling)

- Robust regularization
  - NP-hard, many local minima
  - provably good approximations (a-expansion)

\[ V(\alpha, \beta) = \omega \cdot [\alpha \neq \beta] \]
Potts model

\[ E(L) = \sum_p D_p(L_p) + \sum_{(p,q) \in N} V(L_p, L_q) \]

(piece-wise constant labeling)

- Robust regularization
  - NP-hard, many local minima
  - provably good approximations (a-expansion)

\[ V(\alpha, \beta) = w \cdot [\alpha \neq \beta] \]
Adding label costs

\[ E(L) = \sum_{p} D_p(L_p) + \sum_{(p,q) \in N} V(L_p, L_q) + \sum_{L \in \Lambda} h_L \cdot \delta_L(L) \]

- **Leclerc** [PAMI 89]
  - MDL framework, graduated non-convexity

- **Zhu & Yuille** [PAMI 96]
  - cont. framework (gradient descent + merging heuristics)

- **Torr** [PTRS 98], **Li** [CVPR 2007]
  - AIC/BIC framework, only 1st and 3rd terms
  - Seq. RANSAC heuristic (Torr), LP relaxation w/o any guarantees (Li)

- **Brox & Weikert** [DAGM 04], **Ayed & Mitiche** [TIP’08]
  - Level-sets with merging heuristics (Brox)
  - Multi-level sets (Ayed)

\[ \delta_L(L) = \begin{cases} 
1, & \exists p : L_p = L \\
0, & \text{otherwise}
\end{cases} \]

\( \Lambda \) - set of labels allowed at each point \( p \)
Adding label costs

\[ E(L) = \sum_p D_p(L_p) + \sum_{(p,q)\in N} V(L_p, L_q) + \sum_{L\subseteq \Lambda} h_L \cdot \delta_L(L) \]

\( \Lambda \) - set of labels allowed at each point \( p \)

**Our work** [CVPR 2010, IJCV 2011]

- subsets of labels
- multiple combinatorial algorithms w. optimality bounds
  - \( a\)-expansion++ (3rd term is a high-order clique)
  - UFL heuristics for 1st & 3rd term [Barinova et al., CVPR’10]
- generic model fitting applications

\( \delta_L(L) = \begin{cases} 1, & \exists p : L_p \in L \\ 0, & \text{otherwise} \end{cases} \)
Model fitting

\[ \hat{L} = \arg \min_L \sum_p ||p - L|| \]

\[ ||p - L|| = (p_y - ap_x - b)^2 \]
many outliers

quadratic errors fail

use more robust error measures, e.g.

$$\| p - L \| = | p_y - ap_x - b |$$

gives “MEDIAN” line

- more expensive computations (non-differentiable)
  - still fails if outliers exceed 50%

RANSAC
many outliers

1. sample randomly two points, get a line

RANSAC
many outliers

1. sample randomly two points, get a line
2. count inliers for threshold $T$

RANSAC
many outliers

1. sample randomly two points, get a line
2. count inliers for threshold $T$
3. repeat $N$ times and select model with most inliers

RANSAC
Multiple models and many outliers

Why not RANSAC again?

sequential RANSAC (Torr 98)
Multiple models and many outliers

Why not RANSAC again?

In general, maximization of inliers does not work for outliers + multiple models
Energy-based approach

\[ E(L) = \sum_p \| p - L \| \]

energy-based interpretation of RANSAC criteria for single model fitting:

- find optimal label \( L \)

for one very specific error measure

\[ \| dist \| = \begin{cases} 0, & \text{if } dist \leq T \\ 1, & \text{if } dist > T \end{cases} \]
Energy-based approach

\[ E(L) = \sum_p \| p - L_p \| \]

If multiple models

- assign different models (labels \( L_p \)) to every point \( p \)

- find optimal labeling \( L = \{ L_1, L_2, \ldots, L_n \} \)

Need regularization!
Energy-based approach

\[ E(L) = \sum_p \| p - L_p \| + \sum_{L \in \Lambda} h_L \cdot \delta_L(L) \]

If **multiple** models

- assign different models (labels \( L_p \)) to every point \( p \)

- **find optimal labeling**
  \( L = \{ L_1, L_2, \ldots, L_n \} \)

\( \Lambda \) - set of labels allowed at each point \( p \)

\[ \delta_L(L) = \begin{cases} 1, & \exists p : L_p = L \\ 0, & \text{otherwise} \end{cases} \]
Energy-based approach

\[ E(L) = \sum_{p} \| p - L_p \| + \sum_{pq \in N} w \cdot [L_p \neq L_q] \]

If multiple models

- assign different models (labels \( L_p \)) to every point \( p \)

- find optimal labeling \( L = \{ L_1, L_2, \ldots, L_n \} \)
Energy-based approach

\[ E(L) = \sum_{p} | | p - L_p | | + \sum_{pq \in N} w \cdot [L_p \neq L_q] + \sum_{L \in \Lambda} h_L \cdot \delta_L(L) \]

If multiple models

- assign different models (labels \( L_p \)) to every point \( p \)
- find optimal labeling \( L = \{ L_1, L_2, \ldots, L_n \} \)

Practical problem: number of potential labels (models) is huge, how can we use a-expansion designed for a finite set of labels?
Discrete optimization for continuum of labels?

**example:** line detection
Discrete optimization for continuum of labels?

**example:** line detection

Hough transform (that is, *space of lines*)

Uniform discretization of label space

Adaptive exploration of label space (PEARL)
PEARL

Propose
Expand
And
Reestimate
Labels

data points
PEARL

Propose
Expand
And
Reestimate
Labels

Sample data to generate a **finite set of initial labels**

\[ \Lambda \]

data points + randomly sampled models
PEARL

\[
E(L) = \sum_p \| p - L_p \| + \sum_{pq \in N} w \cdot [L_p \neq L_q] + \sum_{L \in \Lambda} h_L \cdot \delta_L[L]
\]

\[L_p \in \Lambda\]

Propose

Expand

And

Reestimate

Labels

iteration 1: optimize labeling \( L \)

**a-expansion:** minimize \( E(L) \) over a fixed set of labels \( \Lambda \)
PEARL

Propose
Expand
And
Reestimate
Labels

$E(L) = \sum_{p} || p - L_p || + \sum_{pq \in N} w \cdot [L_p \neq L_q] + \sum_{L \in \Lambda} h_L \cdot \delta_L[L]$

iteration 1: reestimate models

reestimating labels in $\Lambda$ for given inliers

minimizing the first term of energy $E(L)$
### PEARL

\[ E(L) = \sum_{p} || p - L_{p} || + \sum_{pq \in N} w \cdot [L_{p} \neq L_{q}] + \sum_{L \in \Lambda} h_{L} \cdot \delta_{L}[L] \]

**Propose**
- **Expand**
- **And**
- **Reestimate**
**Labels**

**a-expansion:** minimize \( E(L) \) over a fixed set of labels \( \Lambda \)

**Iteration 2:** optimize labeling \( L \)
\[ E(L) = \sum_p \| p - L_p \| + \sum_{pq \in N} w \cdot [L_p \neq L_q] + \sum_{L \in \Lambda} h_L \cdot \delta_L[L] \]
\[ E(L) = \sum_p \| p - L_p \| + \sum_{pq \in N} w \cdot [L_p \neq L_q] + \sum_{L \in \Lambda} h_L \cdot \delta_L[L] \]

**PEARL**

- Propose
- Expand
- And
- Reestimate Labels

**iteration 3: optimize labeling \( L \)**
$E(L) = \sum_p || p - L_p || + \sum_{pq \in N} w \cdot [L_p \neq L_q] + \sum_{L \in \Lambda} h_L \cdot \delta_L [L]$
\[ E(L) = \sum_p \| p - L_p \| + \sum_{pq \in N} w \cdot [L_p \neq L_q] + \sum_{L \in \Lambda} h_L \cdot \delta_L [L] \]

PEARL
\[ E(L) = \sum_p \| p - L_p \| + \sum_{pq \in N} w \cdot [L_p \neq L_q] + \sum_{L \in \Lambda} h_L \cdot \delta_L[L] \]

**PEARL**
\[ E(L) = \sum_p || p - L_p || + \sum_{pq \in N} w \cdot [L_p \neq L_q] + \sum_{L \in \Lambda} h_L \cdot \delta_L [L] \]

PEARL

iteration 15... converged.
PEARL can significantly improve initial models

![Graph showing performance comparison between RANSAC and PEARL]

- Single line fitting with 80% outliers
- Deviation (from ground truth)

Number of initial samples
Comparison for multi-model fitting

Low noise

original data points
Comparison for multi-model fitting

Low noise

sequential RANSAC
Comparison for multi-model fitting
Comparison for 
multi-model fitting

original data points

High noise
Comparison for multi-model fitting

High noise

sequential RANSAC
Comparison for multi-model fitting

Other generalization of RANSAC (J-linkage, Toldo & Fusiello, ECCV’08)

High noise
Comparison for multi-model fitting

Finding modes in Hough-space, e.g. via mean-shift
(also maximizes the number of inliers)

Hough transform

High noise
Comparison for multi-model fitting
Automatic noise level estimation by fitting models $L= (a, b, \sigma)$

various level of noise

our result

Each model $L_k$ gets its own $\sigma_k$
K-means vs. PEARL

$$E(L) = \sum_p \| p - L_p \| + \text{hard constraint on number of models}$$

K-means

5 random initial lines + outlier model

gets stuck in local minima
K-means vs. PEARL

\[ E(L) = \sum_{p} \| p - L_p \| + \sum_{L \in \Lambda} h_L \cdot \delta_L(L) \]

K-means gets stuck in local minima

PEARL \( h_L = 1000 \)

better explores label space
**K-means vs. PEARL**

\[
E(L) = \sum_p \| p - L_p \| + \sum_{L \in \Lambda} h_L \cdot \delta_L(L)
\]

**K-means**

- 5 random initial lines + outlier model
- Gets stuck in local minima

**PEARL**

- \( h_L = 500 \)
- 1000 initial lines + outlier model
- Better explores label space
K-means vs. PEARL

\[ E(L) = \sum_{p} \| p - L_p \| + \sum_{L \in \Lambda} h_L \cdot \delta_L(L) \]

K-means

PEARL \( h_L = 2000 \)

5 random initial lines + outlier model

gets stuck in local minima

1000 initial lines + outlier model

better explores label space
Fitting circles

\[ E(L) = \sum_{p} \| p - L_p \| + \sum_{L \in \Lambda} h_L \cdot \delta_L(L) \]
EM vs K-means

EM, with 5 models

K-means, with 5 models
EM vs K-means

EM, with 4 models

K-means, with 4 models
EM vs K-means

EM, with 7 models

K-means, with 7 models
EM vs K-means + sparsity + many proposals

EM + dirichlet, with 50 models
[Figueiredo & Jain, PAMI 2002]

K-means + label cost, with 50 models
[Delong, Osokin, Isack, Boykov, IJCV 2012] (PEARL)
EM vs K-means + sparsity + many proposals

EM + dirichlet, with 50 models
[Figueiredo & Jain, PAMI 2002]

K-means + label cost, with 50 models
[Delong, Osokin, Isack, Boykov, IJCV 2012] (PEARL)
EM vs PEARL

Models in vision have non-overlapping support (since non-transparent models occlude each other)

PEARL can integrate both sparsity and spatial regularity [Delong, Osokin, Isack, Boykov, IJCV 2012]

Q: spatial regularity + EM?
Fitting planes (homographies)

\[ E(L) = \sum_{p} || p - L_p || + \sum_{(p,q) \in N} w \cdot [L_p \neq L_q] + \sum_{L \in \Lambda} h_L \cdot \delta_L(L) \]
Fitting planes (homographies)
same scene
from a different view point…

Note very small steps between each floor
Fitting planes (homographies)

Original image (one of 2 views)
Fitting planes (homographies)

(a) Label costs only
Fitting planes (homographies)

(b) Spatial regularity only
Fitting planes (homographies)

(c) Spatial regularity + label costs
Comparison

based on spectral clustering - Chin, Wang, Sutter ICCV 2009
Fitting Rigid Motions (fundamental matrices)
Fitting Rigid Motions (fundamental matrices)

(a) Label costs only
Fitting Rigid Motions (fundamental matrices)

(b) Spatial regularity only
Fitting Rigid Motions (fundamental matrices)

(c) Spatial regularity + label costs
Fitting Rigid Motions (fundamental matrices)
Fitting Rigid Motions (fundamental matrices)
Fitting Rigid Motions (fundamental matrices)
(unsupervised image segmentation)

Fitting color models

label $L$ represents parameters (e.g. mean) of a Gaussian $N(I|L)$

$$E_I(L) = \sum_p (I_p - L_p)^2 + \sum_{(p,q) \in N} w \cdot [L_p \neq L_q]$$

color consistency model (Chan-Vese)
(unsupervised image segmentation)

Fitting color models

more generally...

label \( L \) represents parameters of an arbitrary distribution \( Pr(I|L) \)

\[
E_I(L) = \sum_p \left| \left| p - L_p \right| \right| + \sum_{(p,q) \in N} w \cdot [L_p \neq L_q] + \sum_{L \in \Lambda} h_L \cdot \delta_L(L)
\]

information theory (MDL) interpretation:

\( \text{number of bits to compress image } I \text{ losslessly} \)
(unsupervised image segmentation)

Fitting color models

Spatial smoothness + label costs

Zhu & Yuille, PAMI 1996
used continuous formulation (gradient descent)
(unsupervised image segmentation)

Fitting color models

Spatial smoothness only [Zabih & Kolmogorov, CVPR 04]
(unsupervised image segmentation)

Fitting color models

Label costs only
(unsupervised image segmentation)

Fitting color models

Spatial smoothness + label costs
Lossy image compression

\[ E(\bar{I}, L) = E_{\bar{I}}(L) + \lambda \cdot \sum_p \| \bar{I}_p - I_p \| \]

- color model fitting (optimal bits for \( \bar{I} \))
- distortion of \( I \)
Lossy image compression

\[ E(\bar{I}, L) = E_l(L) + \lambda \cdot \sum_p \| \bar{I}_p - I_p \| \]

- color model fitting
- (optimal bits for \( \bar{I} \))
- distortion of \( I \)
Lossy image compression

\[ E(\bar{I}, L) = E_{\bar{I}}(L) + \lambda \cdot \sum_p \| \bar{I}_p - I_p \| \]

color model fitting (optimal bits for \( \bar{I} \))
distortion of \( I \)
Lossy image compression

\[ E(\bar{I}, L) = E_{\bar{I}}(L) + \lambda \cdot \sum_{p} \| \bar{I}_p - I_p \| \]

color model fitting (optimal bits for \( \bar{I} \))

\( \lambda \) distortion of \( I \)
Lossy image compression

\[
E(\bar{I}, L) = E_{\bar{I}}(L) + \lambda \cdot \sum_{p} \| \bar{I}_p - I_p \| \]

color model fitting (optimal bits for \( \bar{I} \))

distortion of \( I \)
Rate-Distortion Plot

- Chickens
- Museum
- Zebra
- Tiger

Compression rate (in bpp)

Average squared-error distortion measure
Conclusions

- Energy-based multi-model fitting

- Algorithms for minimizing label-costs energies with global optimality guarantees
  - extended *a-expansion*, standard *UFL heuristics*

- Exploring a continuum of labels, *PEARL*
Extensions

Piece-wise smooth model fitting

\[ k = \frac{|q - q'|}{2 |p - q|^2} \]

[Olsson, Boykov CVPR12]
Extensions

Piece-wise smooth model fitting

[Olsson, Boykov CVPR12]
Extensions
Piece-wise smooth stereo

Labels are tangents (incl. orientation)

[Olsson, Ulen, Boykov CVPR13]
Extensions

Piece-wise smooth stereo

First-order smoothness

Second-order smoothness

Unlike Woodford et al.'08, first-order interactions

[Olsson, Ulen, Boykov CVPR13]
Extensions
Piece-wise smooth stereo

First-order smoothness

Second-order smoothness
Unlike Woodford et al.'08, first-order interactions

[Olsson, Ulen, Boykov CVPR13]
://www.youtube.com/watch?v=2HAFSwFRoR8&list=UUVS7P9dioyjoN7j9mHStQ_Q&feature=player_detailpage&t=7