Normalized Cut with Adaptive Similarity and Spatial Regularization

Jun Liu

School of Mathematical Sciences, Beijing Normal University

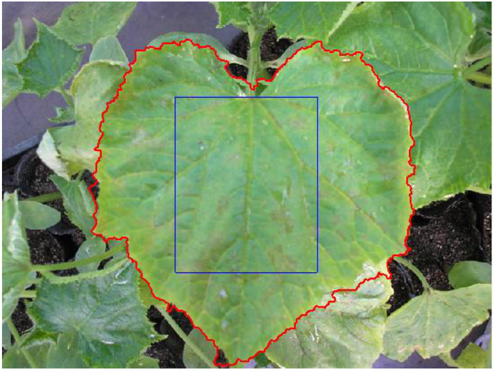
Joint work with Faqiang Wang, Cuicui Zhao and Haiyang Huang

Outline

Outline

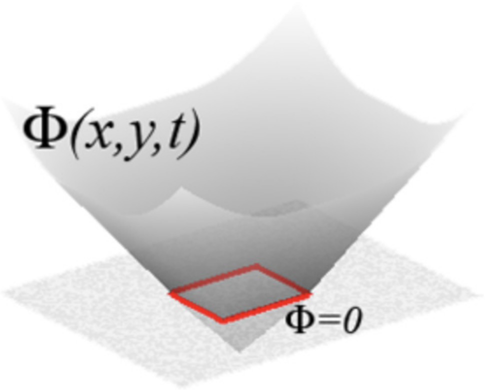
Background

Image segmentation: to extract the object regions to do further research, widely used in computer vision, medical image processing, etc.



Related Work

PDE (level set) method:



Related Work

- PDE Methods
  1. Mumford-Shah (1989)
  2. Chan-Vese (2001)
  3. Piecewise constant level set method (PCLSM) (Lie, Lysaker, Tai, 2006)
  4. Two-phase global segmentation (Bresson *et al.*, 2007)
  5. Convex relaxation method(Pock *et al.*, 2009)
  6. Continuous max-flow (Yuan *et al.*, 2010)
  7. ...

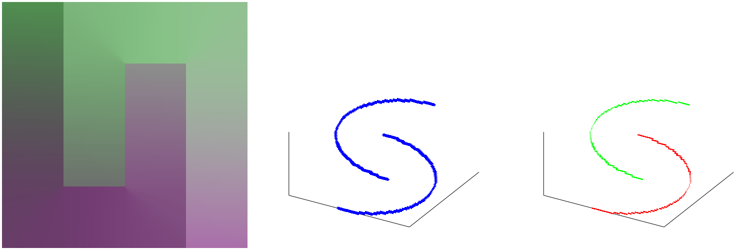
Advantages:

1. High segmentation accuracy.
2. Flexible energy construction.

Flaws:

1. Centers based classification.
2. Data domain based algorithms.

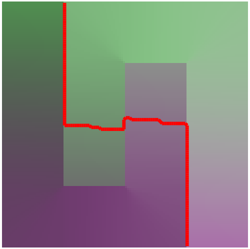
Flaws of Centers based Classification



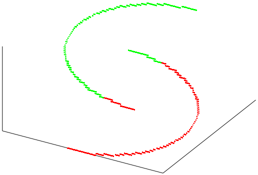
(b) Synthetic image (c) Distribution in RGB space (d) Latent segmentation

Figure 1: A synthetic image and its distributions on RGB space.

Flaws of Centers based Classification



(a) Chan-Vese model

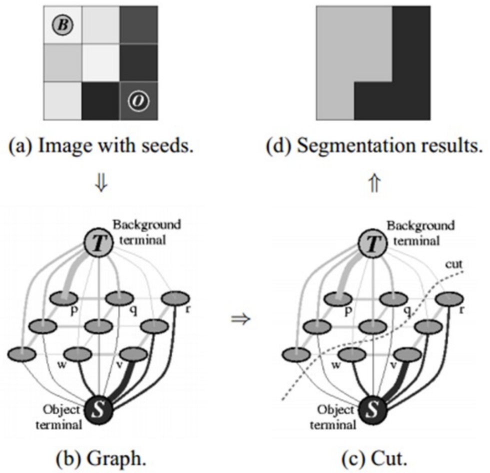


(b) Distributions for Chan-Vese.

Figure 2: Centers based methods are applied to noncentral distributed data.

Related Work

Min-cut: $Cut(B, O)$



- Graph cut Methods

  1. α-expansion and $\alpha - \beta$ swap algorithms. (Boykov,*et al.*, 2001)
  2. Ishikawa's graph cut method. (Ishikawa, 2003)
  3. 4-phase CV graph cut. (Bae, *et al.*, 2011)
  4. PCLSM graph cut. (Liu, *et al.*, 2014)

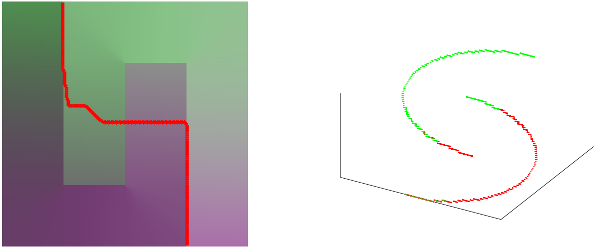
  Advantages:

  Global minimization and fast implementation.

  Flaws:

  1. Suffering from metrication errors.
  2. Only some particular energies can be minimized.
     e.g. anisotropic TV can be minimized by graph cut, but the isotropic TV can not.

Flaws of Centers based Classification



(a) Min-cut method (b) Distribution for Min-cut

Figure 3: Centers based methods are applied to noncentral distributed data.

Another Flaw of Min-cut

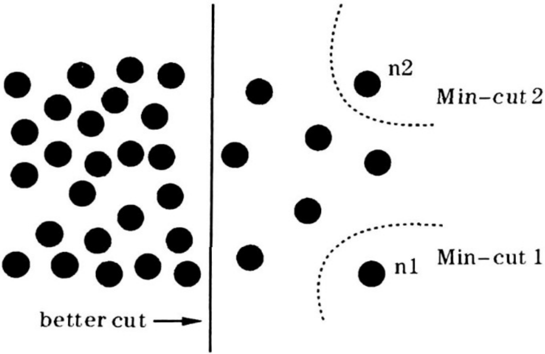


Figure 4: From reference [Shi and Malik, 2000]

Related Work

Normalized Cut (Ncut, [Shi and Malik, 2000])

Given a similarity graph $G =< V, E, W >$, the Normalized cut defined on the graph $G =< V, E, W >$:

$$Ncut(A, B) = cut(A, B)(\frac{1}{assoc(A, V)} + \frac{1}{assoc(B, V)})$$

where

$$assoc(A, V) = \sum_{x \in A} d(x), d(x) = \sum_{y \in V} w(x, y),$$

where V is the set of image pixels, E is the set of edges connecting each data pair and W is the similarity matrix which measures the similarity between data pairs.

- Image segmentation $\Rightarrow$ minimize Normalized cut. Unfortunately, binary segmentation is NP-Hard!

Related Work

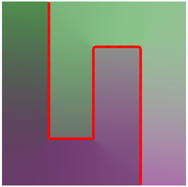
Variational Relaxation Formulation of Ncut

$$\min_{\substack{\mathbf{f}^T D\mathbf{f}=1 \\ \mathbf{f}^T D\mathbf{1}=0}} \mathbf{f}^T(D-W)\mathbf{f}. \tag{1}$$

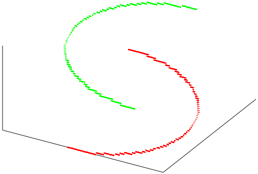
- A minimizer of (1) can be calculated by the "second smallest generalized eigenvector" [Shi and Malik, 2000] of

$$(D-W)\mathbf{f} = \lambda D\mathbf{f} \quad \text{(Spectral clustering).}$$

Noncentral Data Classification



(a) Ncut



(b) Distribution for Ncut

Figure 5: Ncut (spectral clustering) applied to noncentral distributed data.

Motivation

- Normalized cut has better clustering performance.
- Normalized cut lacks of prior information (regularization).
- Similarity W in Normalized cut is manually chosen.

Our Method

- Unify the Normalized Cut, expectation maximum (EM) method and PDE method into a variational framework.
- Spatial regularization for spectrum.
- Similarity W is determined by EM.

Outline

EM based Adaptive Similarity W

- Let Ω be a discrete set in $\mathbb{R}^2$, $I : \Omega \to \{0, 1, 2, ..., 255\}$ stands for the image, the normalized histogram of I can be expressed as

$$p(z) = \frac{1}{|\Omega|} \sum_{y \in \Omega} \delta(z - I(y)), \tag{2}$$

where

$$\delta(x) = \begin{cases} 1, & x = 0, \\ 0, & x \neq 0. \end{cases}$$

EM based Adaptive Similarity W

- $\delta(x)$ can be replaced by a gaussian function $\delta_h(x)$ for convenience of numerical computation

$$\delta_h(x) = \frac{1}{\sqrt{2\pi}h} e^{-\frac{x^2}{2h^2}}.$$

When h is small, $\delta(x)$ can be well approximated.

- Smooth approximation of frequency histogram

$$p(z) = \frac{1}{|\Omega|} \sum_{y \in \Omega} \frac{1}{\sqrt{2\pi}h} e^{-\frac{(z-I(y))^2}{2h^2}}. \tag{3}$$

EM based Adaptive Similarity W

- To estimate the parameter exists in (3), here assume that $\{I(x), x \in \Omega\}$ are samples of random variable $\mathcal{I}$ with the density $p(z)$ given in (3),
- Maximizing the log-likelihood function

$$L(h) = \sum_{x\in\Omega} \ln \frac{1}{|\Omega|} \sum_{y\in\Omega} \frac{1}{\sqrt{2\pi}h} e^{-\frac{(I(x)-I(y))^2}{2h^2}}. \tag{4}$$

- Existence of $log - sum$, similar with Gaussian Mixture Model [luc Gauvain and hui Lee, 1994], Expectation Maximum Algorithm [Bilmes, 1997] is adopted.

EM based Adaptive Similarity W

- EM process: Assume a hidden random variable $\mathcal{Y}$, whose value indicates the sample I comes from which component of the gaussian mixtures, the complete data as $(\mathcal{I}, \mathcal{Y})$, then (4) can be simplified by minimizing

$$L(h) = Q(h; h^{k-1}) - E(h; h^{k-1}), \tag{5}$$

where

$$Q(h; h^{k-1}) = \sum_{x \in \Omega} \sum_{y \in \Omega} \ln\left(\frac{1}{|\Omega|} p_y\left(I(x); h\right)\right) p(y|I(x); h^{k-1}),$$

and

$$E(h; h^{k-1}) = \sum_{x \in \Omega} \sum_{y \in \Omega} \ln p(y|I(x); h) p(y|I(x); h^{k-1}).$$

Variational Method for EM

- $L(h)$ is not easy to be optimized efficiently

- **Theorem (Commutativity of log & sum operations, Liu et al. 2013,IEEE TIP)**

Given a functional $p(x,y) > 0$, we have

$$-\sum_{x\in\Omega} \ln \sum_{y\in\Omega} p(x,y) = \min_{w\in\mathbb{C}_1} \left\{ \begin{array}{c} \underbrace{-\sum_{x\in\Omega}\sum_{y\in\Omega}[\ln p(x,y)]w(x,y)}_{\text{-Q: quadratic}} \\ \underbrace{+\sum_{x\in\Omega}\sum_{y\in\Omega} w(x,y)\ln w(x,y)}_{\text{E: entropy}} \end{array} \right\},$$

where

$$\mathbb{C}_1 = \left\{ w(x,y) : 0 \leqslant w(x,y) \leqslant 1, \sum_{y\in\Omega} w(x,y) = 1, \forall x \in \Omega \right\}.$$

EM based Adaptive Similarity W

- Denote $w : \mathit{\Omega} \times \mathit{\Omega} \to \mathbb{R}, w(x, y) = p(y|I(x); h)$, then the parameter estimation (5) can be converted to an alternating minimization process

$$\min_{h\in\mathbb{H}, w\in\mathbb{C}_1} \left\{ \begin{array}{l} \underbrace{\sum_{x\in\Omega}\sum_{y\in\Omega}\left(\frac{(I(x)-I(y))^2}{2h^2} + \ln(\sqrt{2\pi}h|\Omega|\right) w(x,y)}_{-Q} \\ + \underbrace{\sum_{x\in\Omega}\sum_{y\in\Omega} w(x,y)\ln w(x,y)}_{E}, \end{array} \right\}$$

where $\mathbb{H} = \{h : 0 < h_{min} \leq h \leq h_{max} < +\infty\}$,
$\mathbb{C}_1 = \{w : \Omega \times \Omega \to \mathbb{R} | 0 \leq w(x,y) \leq 1, \sum_{y\in\Omega} w(x,y) = 1, \forall x \in \Omega, \forall y \in \Omega\}$.

- $$\begin{cases} w^k &= \underset{w \in \mathbb{C}_1}{\arg\min} \quad -Q(h^{k-1}, w) + E(w), \\ h^k &= \underset{h}{\arg\min} \quad -Q(h, w^k) + E(w^k). \end{cases} \tag{6}$$

- **Theorem (Energy Descent)**

  The sequence h^k produced by iteration scheme (6) satisfies

  $$L(h^k) \geqslant L(h^{k-1}).$$

EM based Adaptive Similarity W

- Using Lagrange multiplier method,

$$w(x,y) = \frac{1}{S} e^{-\frac{(I(x)-I(y))^2}{2h^2}},$$

  where S serves as a normalization factor.

- $w(x,y)$: measure the similarity between image points, It is model-based, updated adaptively.
- However, w is asymmetric.

Variational Normalized Cut

- Normalized cut

$$\min_{\substack{\mathbf{f}^T D\mathbf{f}=1 \\ \mathbf{f}^T D\mathbf{1}=0}} \mathbf{f}^T(D-W)\mathbf{f}, \tag{7}$$

- Equals to

$$\min_{f\in\mathbb{F}} \left\{ \sum_{x\in\Omega}\sum_{y\in\Omega} w(x,y)(f(x)-f(y))^2 \right\},$$

where
$\mathbb{F} = \{f : \Omega \to \mathbb{R} | \sum_{x\in\Omega} f(x)d(x) = 0, \sum_{x\in\Omega} d(x)f(x)^2 = 1\}$.

Proposed Model

Normalized Cut with Adaptive Similarity and Spatial Regularization

$$\min_{\substack{h\in\mathbb{H}, f\in\mathbb{F}\\ w\in\mathbb{C}_1\bigcap\mathbb{C}_2}} \left\{ E(f,w,h) = \sum_{x\in\Omega}\sum_{y\in\Omega}\{\frac{(I(x)-I(y))^2}{2h^2} + \eta\sum_{x\in\Omega}||\nabla f(x)||_2^p \right.$$

$$+\ln(\sqrt{2\pi}h|\Omega|)\}w(x,y) + \sum_{x\in\Omega}\sum_{y\in\Omega} w(x,y)\ln w(x,y)$$

$$\left. +\lambda\sum_{x\in\Omega}\sum_{y\in\Omega} w(x,y)((K_\epsilon * f)(x) - (K_\epsilon * f)(y))^2 \right\}.$$

- K_ϵ: used to prove the existence of minimizers theoretically in a proper functional space. ϵ small enough, K_ϵ becomes a delta function.

Notation

- $\mathbb{C}_1 = \{w : \Omega \times \Omega \to \mathbb{R} | 0 \leq w(x, y) \leq 1, \sum_{y \in \Omega} w(x, y) = 1, \forall x \in \Omega, \forall y \in \Omega\}$,
- $\mathbb{C}_2 = \{w : \Omega \times \Omega \to \mathbb{R} | 0 \leq w(x, y) \leq 1, w(x, y) = w(y, x), \forall x \in \Omega, \forall y \in \Omega\}$,
- $\mathbb{F} = \{f : \Omega \to \mathbb{R} | \sum_{x \in \Omega} f(x) d(x) = 0, \sum_{x \in \Omega} d(x) f(x)^2 = 1, \forall x \in \Omega\}$,
- $\mathbb{H} = \{h : 0 < h_{min} \leq h \leq h_{max} < +\infty\}$,
- λ, η control the balance of each term, $||.||_2$ is the L^2 norm,
- $p = 1, 2$ corresponding to H^1 regularization (NCASH^1) and TV regularization (NCASTV), respectively.

Outline

Existence of Minimizer

Theorem

There exists a minimizer for NCASTV model

$$\min_{(f,w,h)\in X} E(f,w,h), \tag{8}$$

where $X := \{(f,w,h) : f \in BV(\Omega), w \in L^\infty(\Omega\times\Omega), 0 \le w \le 1, \int_\Omega f(x)dx = 0, \int_\Omega f^2(x)dx = 1, \int_\Omega w(x,y)dy = 1, |f(x)| < C, w(x,y) = w(y,x), a.e. x \in \Omega, y \in \Omega, 0 < h_{min} \le h \le h_{max} < +\infty\}$.

Outline

Algorithm for NCASH[1] Model

- ADMM Scheme:

$$(i)\ \min_{w\in\mathbb{C}_2}\max_{\beta}\Bigg\{\sum_{x\in\Omega}\sum_{y\in\Omega}\{\tfrac{(I(x)-I(y))^2}{2h^2}+\ln(\sqrt{2\pi}h|\Omega|)\}w(x,y)$$

$$+\sum_{x\in\Omega}\sum_{y\in\Omega}w(x,y)\ln w(x,y)+\sum_{x\in\Omega}\beta(x)(\sum_{y\in\Omega}w(x,y)-1)$$

$$+\lambda\sum_{x\in\Omega}\sum_{y\in\Omega}w(x,y)((K_\epsilon * f)(x)-(K_\epsilon * f)(y))^2\Bigg\},$$

$$(ii)\ \min_{h\in\mathbb{H}}\Bigg\{\sum_{x\in\Omega}\sum_{y\in\Omega}\{\tfrac{(I(x)-I(y))^2}{2h^2}+\ln(\sqrt{2\pi}h|\Omega|)\}w(x,y)\Bigg\},$$

$$(iii)\ \min_{f\in\mathbb{F}}\Bigg\{\lambda\sum_{x\in\Omega}\sum_{y\in\Omega}w(x,y)((K_\epsilon * f)(x)-(K_\epsilon * f)(y))^2$$

$$+\eta\sum_{x\in\Omega}||\nabla f(x)||_2^2\Bigg\}.$$

Algorithm for NCASH[1] Model

- Subproblems (i) and (ii) have closed-form solutions by Lagrangian multiplier method and the first order optimal condition

$$h^2 = Proj_{\mathbb{H}}(\frac{\sum_{x\in\Omega}\sum_{y\in\Omega} w(x,y)(I(x)-I(y))^2}{|\Omega|}).$$

$$w(x,y) = Proj_{\mathbb{C}_2}(\frac{e^{\frac{-(I(x)-I(y))^2}{h^2}-\lambda((K_\epsilon * f)(x)-(K_\epsilon * f)(y))^2}}{\sum_{y\in\Omega} e^{\frac{-(I(x)-I(y))^2}{h^2}-\lambda((K_\epsilon * f)(x)-(K_\epsilon * f)(y))^2}}).$$

Algorithm for NCASH[1] Model

- Equivalent form of f subproblem

$$\min_{\substack{\mathbf{f}^T D\mathbf{f}=1 \\ \mathbf{f}^T D\mathbf{1}=0}} \lambda \mathbf{f}^T K^T (D-W) K\mathbf{f} + \eta ||\nabla \mathbf{f}||_2^2, \tag{9}$$

where W is the similarity matrix, and D is the degree matrix, $\mathbf{f}$ is the discretion of f.

- Denote $\mathbf{z} = D^{\frac{1}{2}}\mathbf{f}$ and $\mathbb{S}_2 = \{\mathbf{z} : \mathbf{z}^T D^{\frac{1}{2}}\mathbf{1} = 0\}$, (9) can be converted to the problem concerning about $\mathbf{z}$:

$$\min_{\mathbf{z}\in\mathbb{S}_2} \min_{\mu} \Big\{ \mathbf{z}^T D^{-\frac{1}{2}} (\lambda K^T (D-W) K - \eta \Delta) D^{-\frac{1}{2}} \mathbf{z} \\ - \mu(||\mathbf{z}||^2 - 1) \Big\}.$$

Algorithm for NCASH[1] Model

- Inspired by Dinkelbach Algorithm [Rodenas et al., 1999] [Ghanem and Ahuja, 2010], the numerical scheme is shown below

$$\left\{\begin{array}{lcl} \mu^k & = & \frac{(\mathbf{z}^k)^T D^{-\frac{1}{2}}(\lambda K^T(D-W)K-\eta\Delta)D^{-\frac{1}{2}}\mathbf{z}^k}{(\mathbf{z}^k)^T\mathbf{z}^k}, \\ \widehat{\mathbf{z}}^{k+1} & = & \mathbf{z}^k - \tau\big(D^{-\frac{1}{2}}(\lambda K^T(D-W)K \\ & & -\eta\Delta)D^{-\frac{1}{2}}\mathbf{z}^k - \mu^k\mathbf{z}^k\big), \\ \mathbf{z}^{k+1} & = & Proj_{\mathbb{S}_2}(\widehat{\mathbf{z}}^{k+1}). \end{array}\right. \tag{10}$$

In fact,

$$\mathbf{z}^{k+1} = Proj_{\mathbb{S}_2}(\widehat{\mathbf{z}}^{k+1}) = \widehat{\mathbf{z}}^{k+1} - \frac{(\widehat{\mathbf{z}}^{k+1})^T D^{\frac{1}{2}}\mathbf{1}}{\mathbf{1}^T D\mathbf{1}} D^{\frac{1}{2}}\mathbf{1}.$$

Algorithm 1: NCASH[1]

- Given $\mathbf{f}^0 = \mathbf{1}$, $\mathbf{z}^0 = Proj_{\mathbb{S}_2}(\mathbf{f}^0)$, tolerant error $= \epsilon$; Set τ=2 ,h^0=50, let $k = 0$.
- Update similarity matrix

$$w^{k+1}(x,y) = \frac{e^{\frac{-(I(x)-I(y))^2}{2(h^k)^2} - \lambda(Kf^k(x)-Kf^k(y))^2}}{\sum\limits_{y\in\Omega} e^{\frac{-(I(x)-I(y))^2}{2(h^k)^2} - \lambda(Kf^k(x)-Kf^k(y))^2}}.$$

- Calculate the projection of W in $\mathbb{C}_2$

$$W^{k+1} = \frac{W^{k+1} + (W^{k+1})^T}{2}.$$

- Calculate h

$$(h^{k+1})^2 = Proj_{\mathbb{H}}\left(\frac{\sum\limits_{x\in\Omega}\sum\limits_{y\in\Omega} w^{k+1}(x,y)(I(x)-I(y))^2}{|\Omega|}\right).$$

Algorithm 1: NCASH[1]

- Calculate $\mathbf{z}$

$$\begin{cases} \mu^k & = & \frac{(\mathbf{z}^k)^T D^{-\frac{1}{2}} (\lambda K^T (D - W^{k+1}) K - \eta \Delta) D^{-\frac{1}{2}} \mathbf{z}^k}{(\mathbf{z}^k)^T \mathbf{z}^k}, \\ \widehat{\mathbf{z}}^{k+1} & = & \mathbf{z}^k - \tau (D^{-\frac{1}{2}} (\lambda K^T (D - W^{k+1}) K \\ & & -\eta \Delta) D^{-\frac{1}{2}} \mathbf{z}^k - \mu^k \mathbf{z}^k), \\ \mathbf{z}^{k+1} & = & \widehat{\mathbf{z}}^{k+1} - \frac{(\widehat{\mathbf{z}}^{k+1})^T D^{\frac{1}{2}} \mathbf{1}}{\mathbf{1}^T D \mathbf{1}} D^{\frac{1}{2}} \mathbf{1}. \end{cases}$$

- Reconstruct $\mathbf{f}$

$$\mathbf{f}^{k+1} = D^{-\frac{1}{2}} \mathbf{z}^{k+1}.$$

- If $\frac{||\mathbf{f}^{k+1} - \mathbf{f}^k||^2}{||\mathbf{f}^k||^2} < \epsilon$, stop; Or, set $k = k + 1$, return to step 3.

Algorithm 2: NCASTV

- Given $\mathbf{f}^0 = \mathbf{z}^0 = \mathbf{g}^0 = \mathbf{1}$, tolerant error $= \zeta$; Set τ=2, $h^0 = 50$. Let $k = 0$.
- Update similarity matrix

$$w^{k+1}(x,y) = \frac{e^{\frac{-(I(x)-I(y))^2}{2{h^k}^2} - \lambda(Kf^k(x)-Kf^k(y))^2}}{\sum_{y\in\Omega} e^{\frac{-(I(x)-I(y))^2}{2(h^k)^2} - \lambda(Kf^k(x)-Kf^k(y))^2}}.$$

- Calculate the projection of W in $\mathbb{C}_2$

$$W^{k+1} = \frac{W^{k+1} + (W^{k+1})^T}{2}.$$

- Calculate h

$$(h^{k+1})^2 = Proj_{\mathbb{H}}\left(\frac{\sum_{x\in\Omega}\sum_{y\in\Omega} w^{k+1}(x,y)(I(x)-I(y))^2}{|\Omega|}\right).$$

Algorithm 2: NCASTV

- Calculate $\mathbf{z}$

$$\begin{cases} \mu^k = \frac{(\mathbf{z}^k)^T D^{-\frac{1}{2}}[\lambda K^T(D-W^{k+1})K+\epsilon I]D^{-\frac{1}{2}}\mathbf{z}^k-\epsilon(\mathbf{z}^k)^T D^{-\frac{1}{2}}\mathbf{g}^k}{(\mathbf{z}^k)^T\mathbf{z}^k}, \\ \widehat{\mathbf{z}}^{k+1} = \mathbf{z}^k - \tau(D^{-\frac{1}{2}}[\lambda K^T(D-W^{k+1})K+\epsilon I]D^{-\frac{1}{2}}\mathbf{z}^k \\ \qquad -\mu^k\mathbf{z}^k - \epsilon D^{-\frac{1}{2}}\mathbf{g}^k), \\ \mathbf{z}^{k+1} = \widehat{\mathbf{z}}^{k+1} - \frac{(\widehat{\mathbf{z}}^{k+1})^T D^{\frac{1}{2}}\mathbf{1}}{\mathbf{1}^T D\mathbf{1}} D^{\frac{1}{2}}\mathbf{1}. \end{cases}$$

- Reconstruct $\mathbf{f}$

$$\mathbf{f}^{k+1} = D^{-\frac{1}{2}}\mathbf{z}^{k+1}.$$

- Calculate the auxiliary variable

$$\mathbf{g}^{k+1} = ROF(\mathbf{f}^{k+1}, \frac{\eta}{2\epsilon}).$$

- If $\frac{||\mathbf{f}^{k+1}-\mathbf{f}^k||^2}{||\mathbf{f}^k||^2} < \zeta$, stop; Or, set $k = k+1$, return to step 3.

Outline

Numerical Results 1

- Double-Moon Data-sets consists of 150 points with two labels.

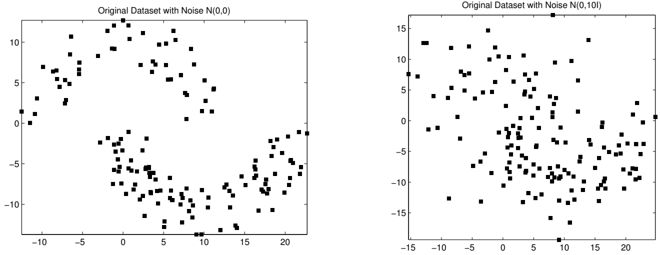


Figure 6: Data: Double-moon data-set and the data-set corrupted by N(0,10) shown in the first and second figure, respectively.

Ncut vs NCASH^1

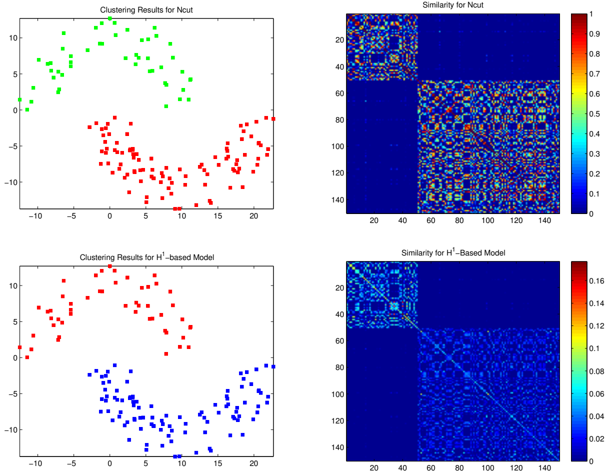


Figure 7: Results: Double-moon data-set by Ncut and the proposed NCASH^1 model and the corresponding similarity matrices.

Ncut vs NCASH[1]

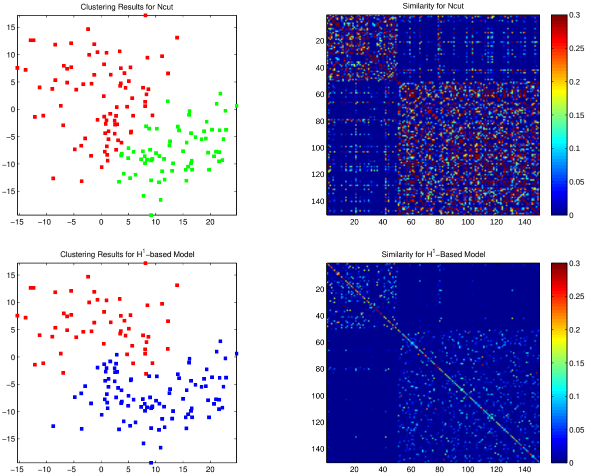


Figure 8: Results: Double-moon data-set corrupted by N(0,10) by Ncut and NCASH[1] model and the corresponding similarity matrices.

Iteration of μ^k

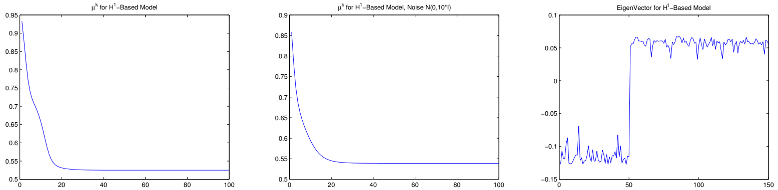
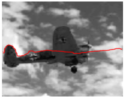


Figure 9: The first two graph shows μ^k by NCASH^{1}model, and the third graph shows eigenvector by NCASH^1 model.

Numerical Results 2

- Images source BSDS500 [Arbelaez et al., 2011].
- Iteration of W.


(a) 1_{st} iter


(b) 2_{nd} iter


(c) 3_{rd} iter


(d) 4_{th} iter


(e) 5_{th} iter


(f) 6_{th} iter


(g) 7_{th} iter


(h) 8_{th} iter


(i) 9_{th} iter


(j) 10_{th} iter

Figure 10: (a)-(j) show the first 10 iteration segmentation results of sample image resized to 100*100 by NCASTV model, with parameters $\lambda = 1$, $\epsilon = 0.001 * \lambda$, $\eta = 0.001 * \epsilon$.

Regularization Term

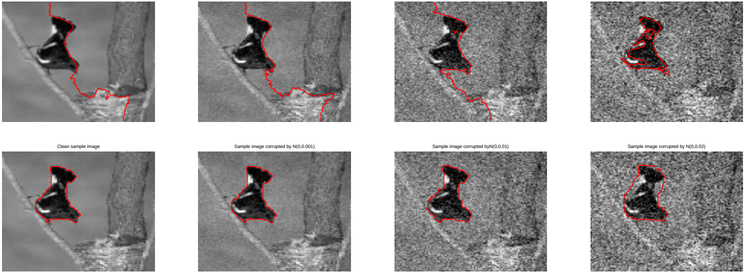
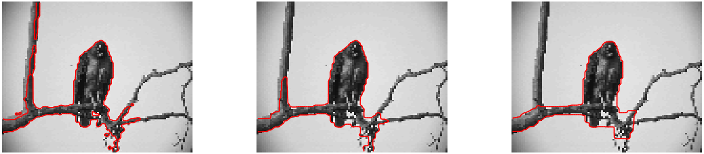


Figure 11: The segmentation results of sample image corrupted by N(0,0), N(0,0.001), N(0,0.01), N(0,0.02) by Ncut model in the first row and by NCASTV model in the second row, with parameters in NCASTV model $\lambda = 1$, $\epsilon = 0.001 * \lambda$, and $\eta = 0.005 * \epsilon$, $\eta = 0.005 * \epsilon$, $\eta = 0.005 * \epsilon$, $\eta = 0.009 * \epsilon$, respectively.

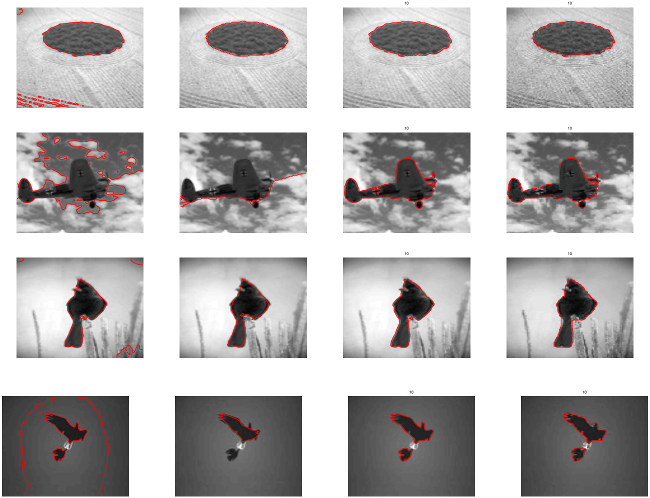
Regularization Parameter η



(a) $\eta = 0.001 * \epsilon$ (b) $\eta = 0.005 * \epsilon$ (c) $\eta = 0.01 * \epsilon$

Figure 12: The segmentation results of sample image by NCASTV model with different regularization parameters: $\eta = 0.001 * \epsilon$, $0.005 * \epsilon$, $0.01 * \epsilon$ respectively. The other parameters: $\lambda = 1$, $\epsilon = 0.001 * \lambda$.

Numerical Results 3



(a) CV model (b) Pre-Ncut (c) NCASH[1] (d) NCASTV

Algorithms evaluation

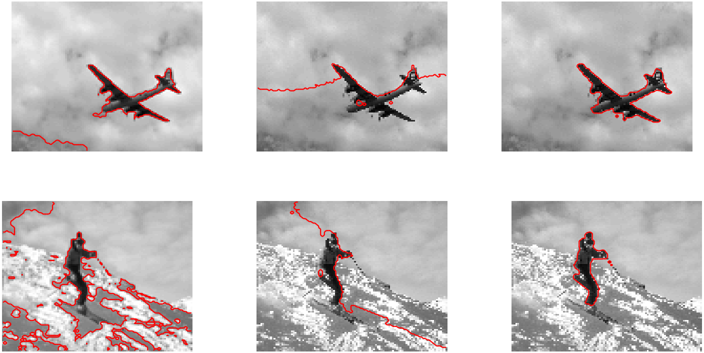
Table 1: Algorithms evaluation.

	Image 1		Image 2		Image 3		Image 4	
	VI	RI	VI	RI	VI	RI	VI	RI
Pre-Ncut	0.1268	0.9843	1.4567	0.5026	0.1760	0.9736	0.2010	0.9624
Chan-Vese	0.2710	0.9551	1.2914	0.5591	0.3811	0.9318	1.2053	0.5224
NCASH[1]	0.1219	0.9830	0.1986	0.9694	0.1311	0.9827	0.1018	0.9862
NCASTV	0.0793	0.9909	0.1930	0.9696	0.1853	0.9736	0.1378	0.9800

VI: Variation of Information.

RI: Rand Index.

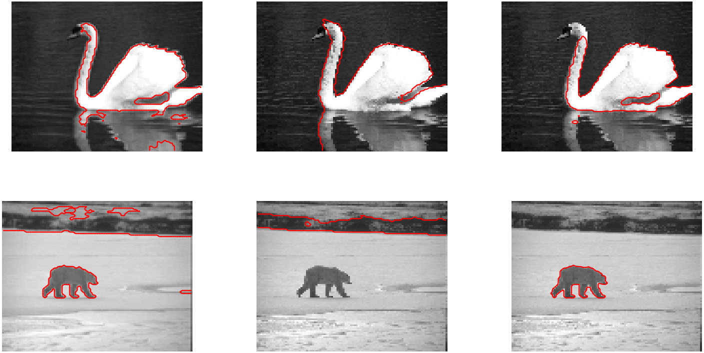
CV vs Pre-Ncut vs NCASTV



(a) CV model (b) Pre-Ncut (c) NCASTV

Figure 14:

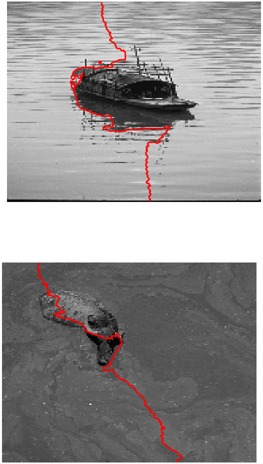
CV vs Pre-Ncut vs NCASTV



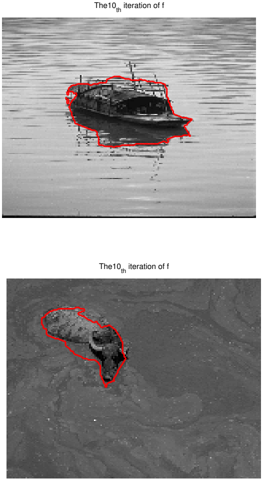
(a) CV model (b) Pre-Ncut (c) NCASTV

Figure 15:

Pre-Ncut vs Pre-NCASTV

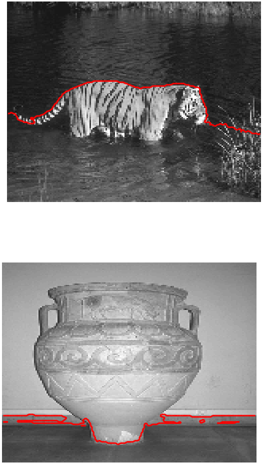


(a) Pre-Ncut

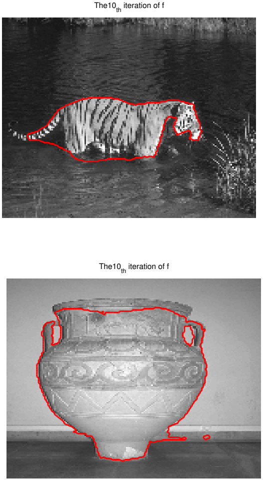


(b) Pre-NCASTV

Pre-Ncut vs Pre-NCASTV



(a) Pre-Ncut



(b) Pre-NCASTV

Outline

Conclusion

- Main contributions:
  1. Unify the Ncut (Spectral clustering),TV (PDE) and EM (statistics) into a variational framework.
  2. EM process gives Ncut good similarity, regularization makes proposed model robust for noise, spectrum clustering can well partition noncentral distributed data.
- Some possible extensions:
  1. Multiphase regularized Ncut with EM.
  2. Priori (e.g. convex)
  3. Convergence.

Thank you!

References

Pablo Arbelaez, Michael Maire, Charless C Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):898–916, 2011.

J. Bilmes. A gentle tutorial on the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. 1997. URL `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.613`.

Bernard Ghanem and Narendra Ahuja. Dinkelbach ncut: An efficient framework for solving normalized cuts problems with priors and convex constraints. *International Journal of Computer Vision*, 89(1):40–55, 2010.

Jean luc Gauvain and Chin hui Lee. Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *IEEE Transactions on Speech and Audio Processing*, 2:291–298, 1994.

Ricardo Garcia Rodenas, Maria Rodriguez, and Doroteo Verastegui Rayo. Extensions of dinkelbach's algorithm for solving non-linear fractional programming problems. *Top*, 7(1): 33–70, 1999.

Jianbo Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Ananlysis and Machine Intelligence*, 22(8):888–905, 2000.