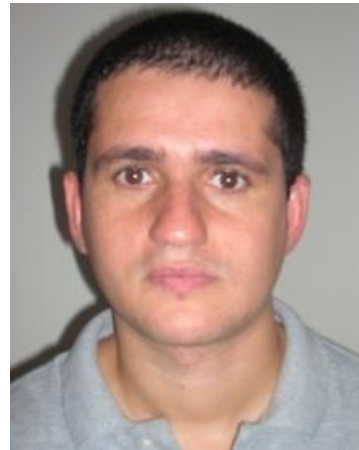
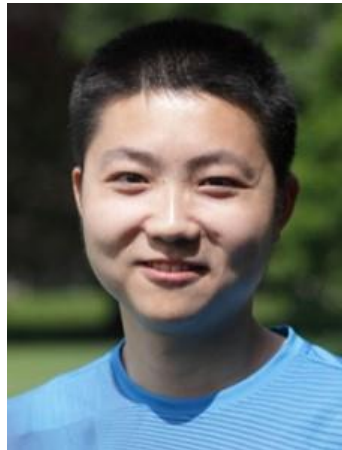


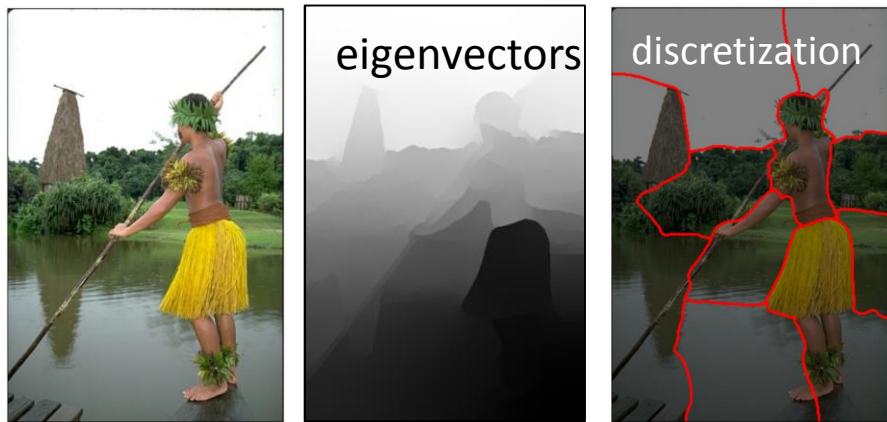
Normalized Cut Meets MRF

Meng Tang, Dmitrii Marin, Ismail Ben Ayed, Yuri Boykov



Normalized Cut (NC)

$$NC(S) = - \sum_k \frac{assoc(S^k, S^k)}{assoc(S^k, \Omega)}$$



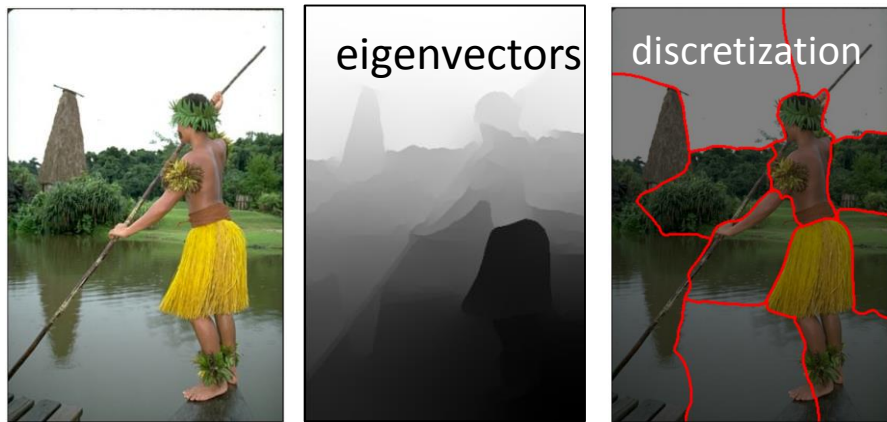
[Arbelaez, Maire, Fowlkes & Malik, 2010]



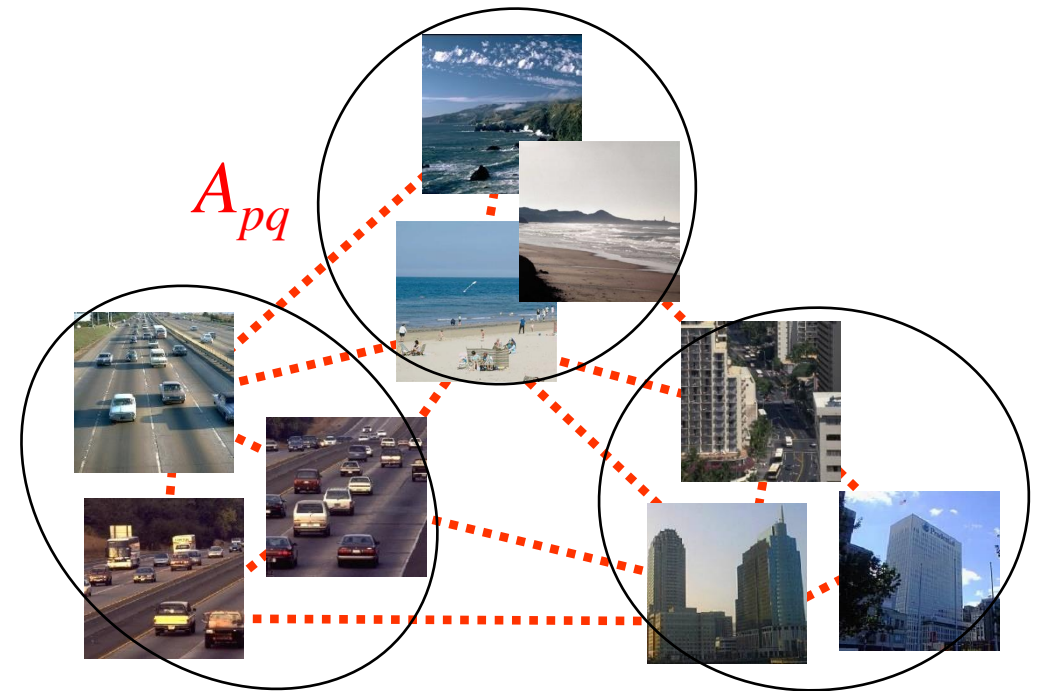
[Shi & Malik, 2000] [Ng, Jordan & Weiss, 2002] [Belkin & Niyogi, 2003] [Luxburg, 2007]

Normalized Cut (NC)

$$NC(S) = - \sum_k \frac{S^{k'} A S^k}{d' S^k}$$



[Arbelaez, Maire, Fowlkes & Malik, 2010]



[Shi & Malik, 2000] [Ng, Jordan & Weiss, 2002] [Belkin & Niyogi, 2003] [Luxburg, 2007]

Markov Random Field (**MRF**)

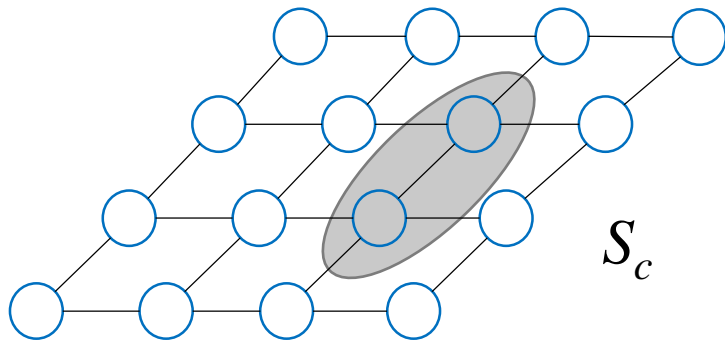
(graphical models)

$$\sum_{c \in \mathcal{F}} E_c(S_c)$$

Markov Random Field (MRF)

(graphical models)

$$\sum_{c \in \mathcal{F}} E_c(S_c)$$



Potts Model: edge alignment

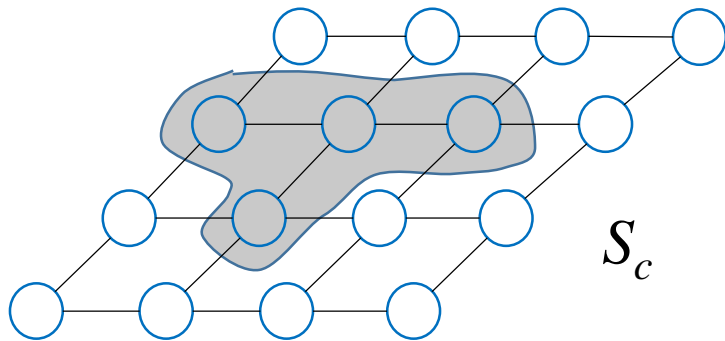


interactive segmentation [*Boykov & Jolly, 2001*]

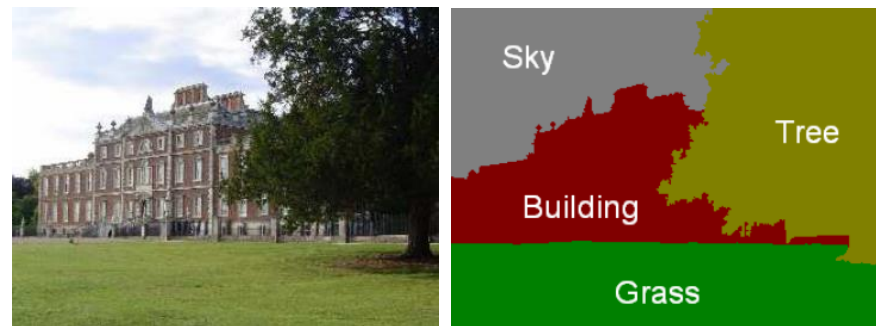
Markov Random Field (MRF)

(graphical models)

$$\sum_{c \in \mathcal{F}} E_c(S_c)$$



Robust P^n Potts: bin consistency

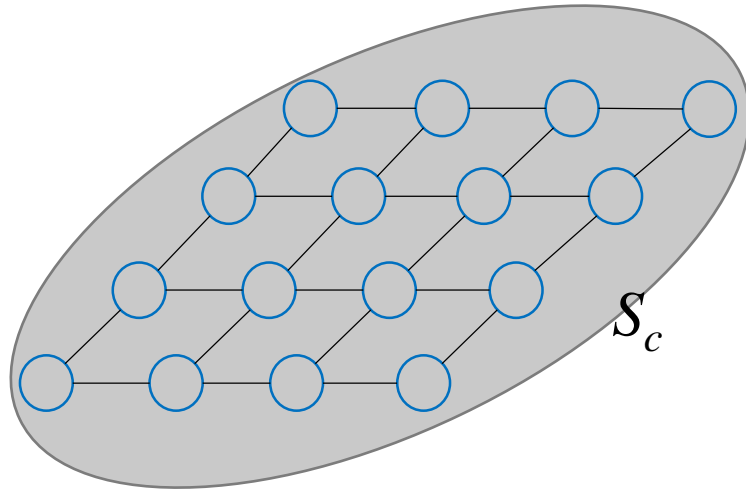


semantic segmentation [Kohli & Torr 2009] [Gould 2014]

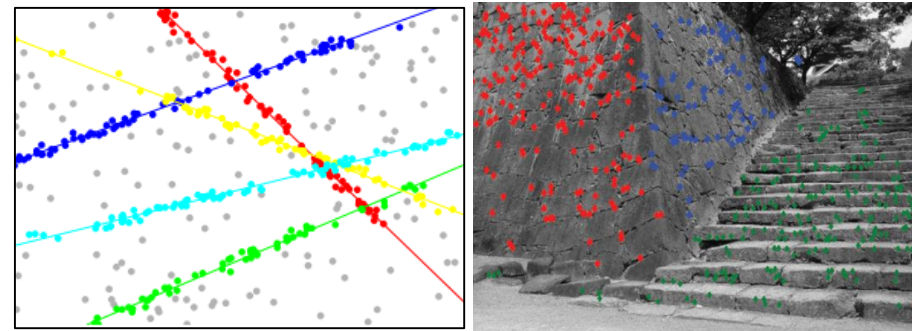
Markov Random Field (MRF)

(graphical models)

$$\sum_{c \in \mathcal{F}} E_c(S_c)$$



Label Cost: sparsity



geometric model fitting [DeLong et al., 2012]

Our proposal: **Normalized Cut** + **MRF**

$$E(S) = \underbrace{\sum_k -\frac{S^{k'} A S^k}{d' S^k}}_{\text{balanced clustering}} + \gamma \underbrace{\sum_{c \in \mathcal{F}} E_c(S_c)}_{\text{regularization constraints}}$$

Why *MRF* for Normalized Cut?

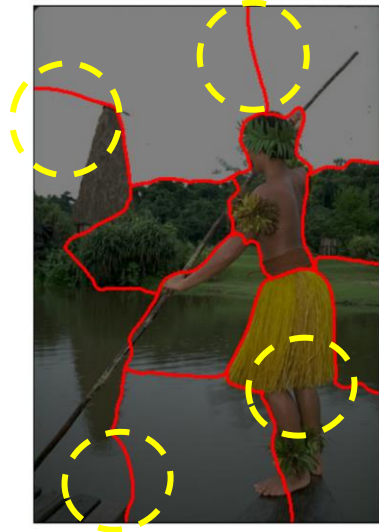
Why *MRF* for Normalized Cut?

previous NC approach:



Why *MRF* for Normalized Cut?

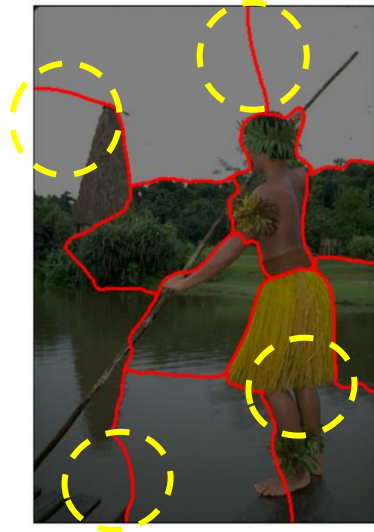
weak edge alignment



previous NC approach:

Why *MRF* for Normalized Cut?

weak edge alignment



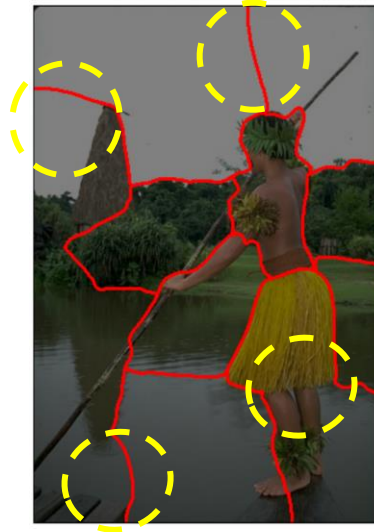
previous NC approach:

post-processing

(e.g. [\[Arbelaez et al., 2011\]](#))

Why *MRF* for Normalized Cut?

weak edge alignment



previous NC approach:

post-processing

(e.g. [\[Arbelaez et al., 2011\]](#))

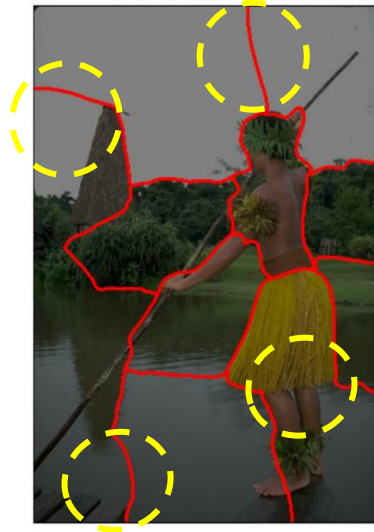
Our approach:

NC + Potts

MRF

Why *MRF* for Normalized Cut?

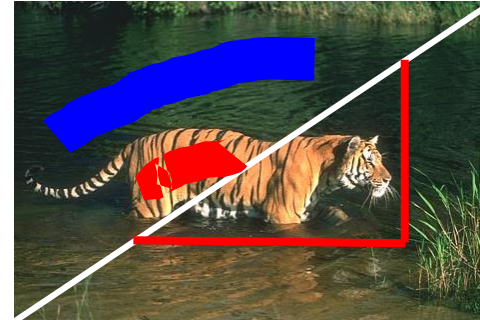
weak edge alignment



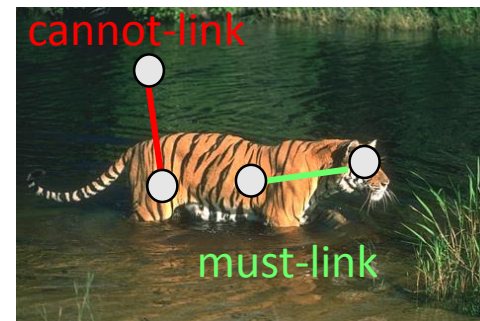
previous NC approach:

post-processing
(e.g. [Arbelaez et al., 2011])

semi-supervision is challenging



[Yu & Shi 2004]
[Eriksson et al. 2010]
[Maji et al. 2011]
[Chew et al. 2015]



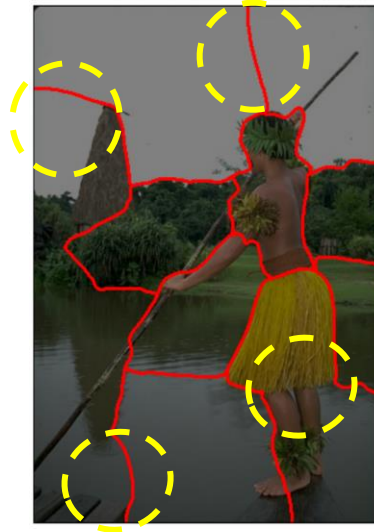
reformulation of NC
constrained eigen problem

Our approach:

NC + Potts
MRF

Why *MRF* for Normalized Cut?

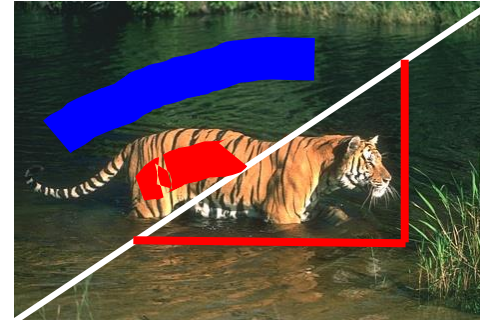
weak edge alignment



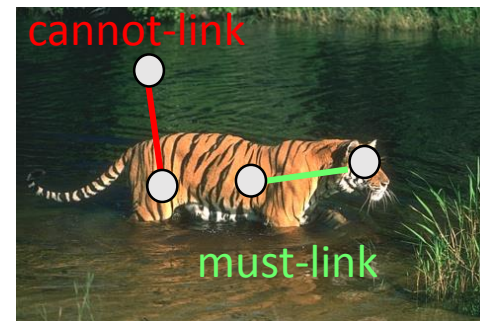
previous NC approach:

post-processing
(e.g. [Arbelaez et al., 2011])

semi-supervision is challenging



[Yu & Shi 2004]
[Eriksson et al. 2010]
[Maji et al. 2011]
[Chew et al. 2015]



reformulation of NC
constrained eigen problem

Our approach:

NC + Potts

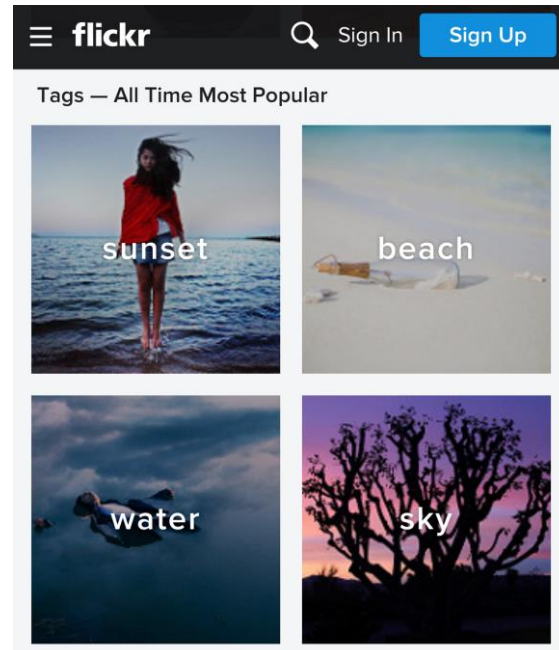
MRF

NC + Potts + seeds

MRF

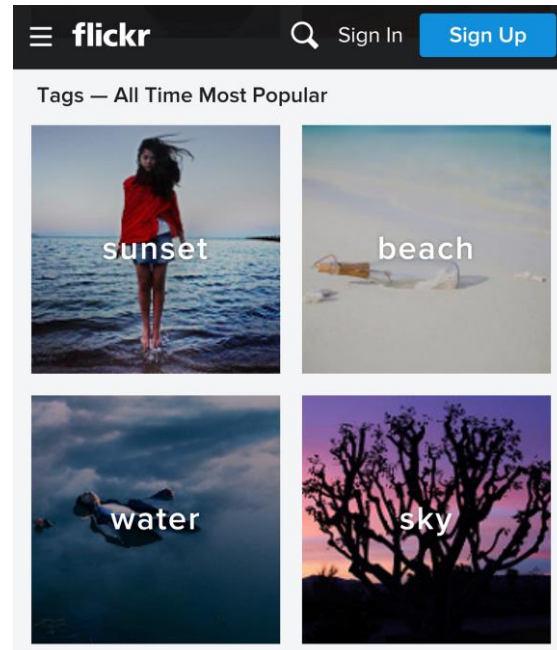
Why *MRF* for Normalized Cut?

How to incorporate group priors?



Why *MRF* for Normalized Cut?

How to incorporate group priors?



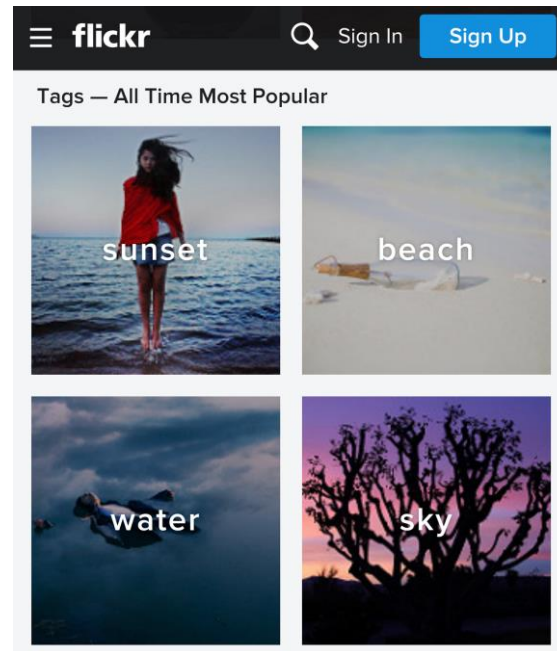
Our approach:

NC + Robust P^n Potts

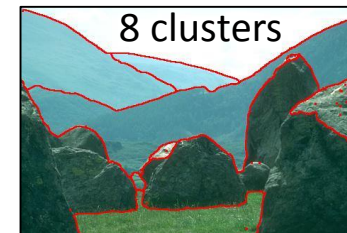
MRF

Why *MRF* for Normalized Cut?

How to incorporate group priors?



How many clusters ?



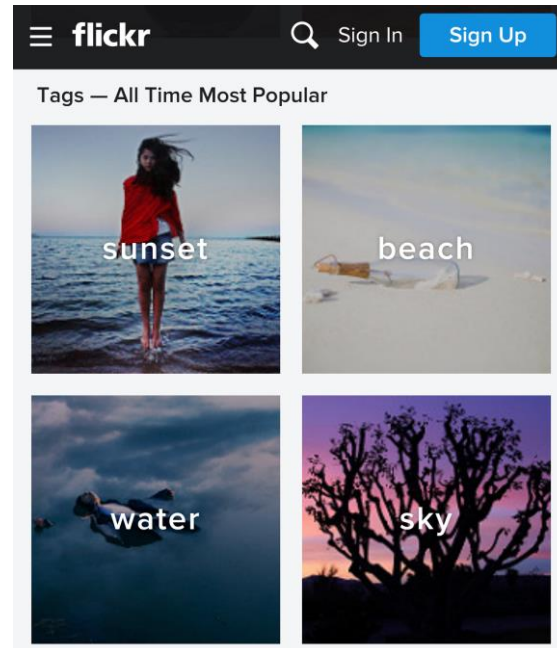
Our approach:

NC + Robust P^n Potts

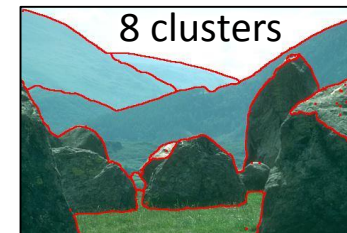
MRF

Why *MRF* for Normalized Cut?

How to incorporate group priors?



How many clusters ?



Our approach:

NC + Robust P^n Potts
MRF

NC + label costs
MRF

Why *Normalized Cut* for MRF?

Why *Normalized Cut* for MRF?

typical MRF for segmentation:

$$-\sum_k \sum_{p \in S^k} \ln P(I_p | \theta^k) + \sum_{pq \in \mathcal{N}} w \cdot [s_p \neq s_q]$$

ML term for θ^k

Why *Normalized Cut* for MRF?

typical MRF for segmentation:

$$\sum_k \sum_{p \in S^k} \ln P(I_p | \theta^k) + \sum_{pq \in \mathcal{N}} w \cdot [s_p \neq s_q]$$

ML term for θ^k

Why *Normalized Cut* for MRF?

typical MRF for segmentation:

$$\sum_k \sum_{p \in S^k} \ln P(I_p | \theta^k) + \sum_{pq \in \mathcal{N}} w \cdot [s_p \neq s_q]$$

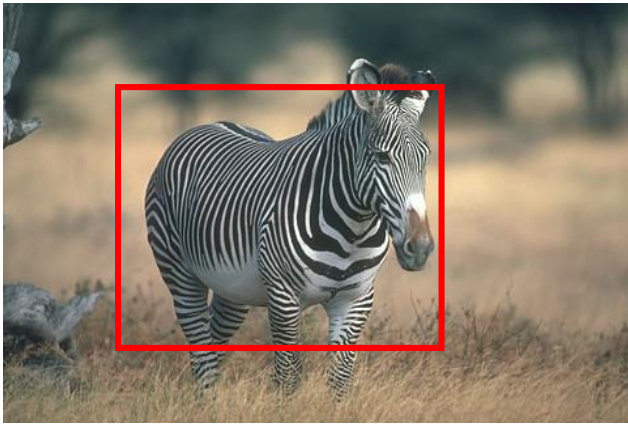
ML term for θ^k

Why *Normalized Cut* for MRF?

typical MRF for segmentation:

$$\sum_k \sum_{p \in S^k} \ln P(I_p | \theta^k) + \sum_{pq \in \mathcal{N}} w \cdot [s_p \neq s_q]$$

ML term for θ^k



model fitting (*e.g.* GMM)

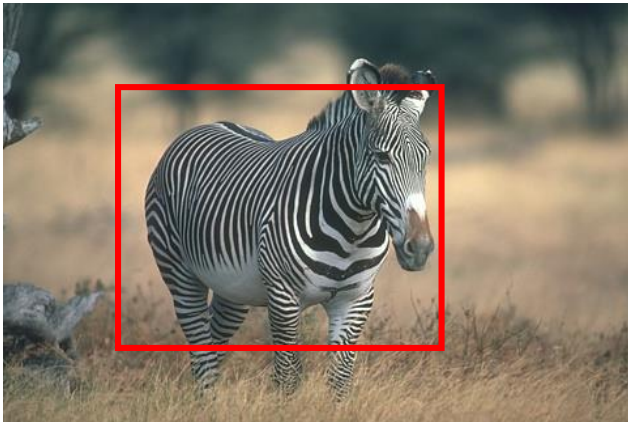
GrabCut [*Rother, Kolmogorov, Blake, 2004*]

Why *Normalized Cut* for MRF?

typical MRF for segmentation:

$$\sum_k \sum_{p \in S^k} \ln P(I_p | \theta^k) + \sum_{pq \in \mathcal{N}} w \cdot [s_p \neq s_q]$$

ML term for θ^k



model fitting (e.g. GMM)

GrabCut [Rother, Kolmogorov, Blake, 2004]

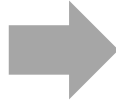
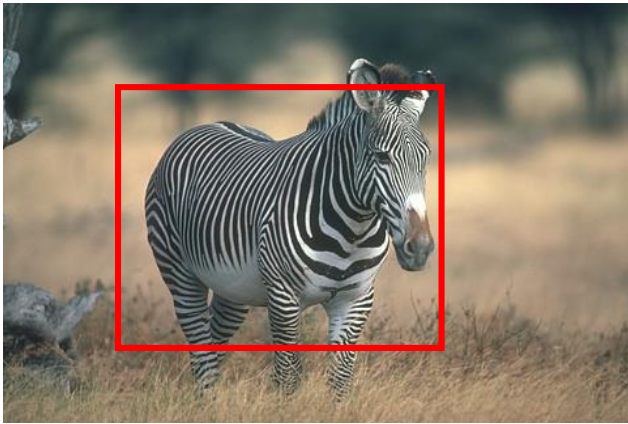
color space **clustering**

Why *Normalized Cut* for MRF?

typical MRF for segmentation:

$$\sum_k \sum_{p \in S^k} \ln P(I_p | \theta^k) + \sum_{pq \in \mathcal{N}} w \cdot [s_p \neq s_q]$$

ML term for θ^k

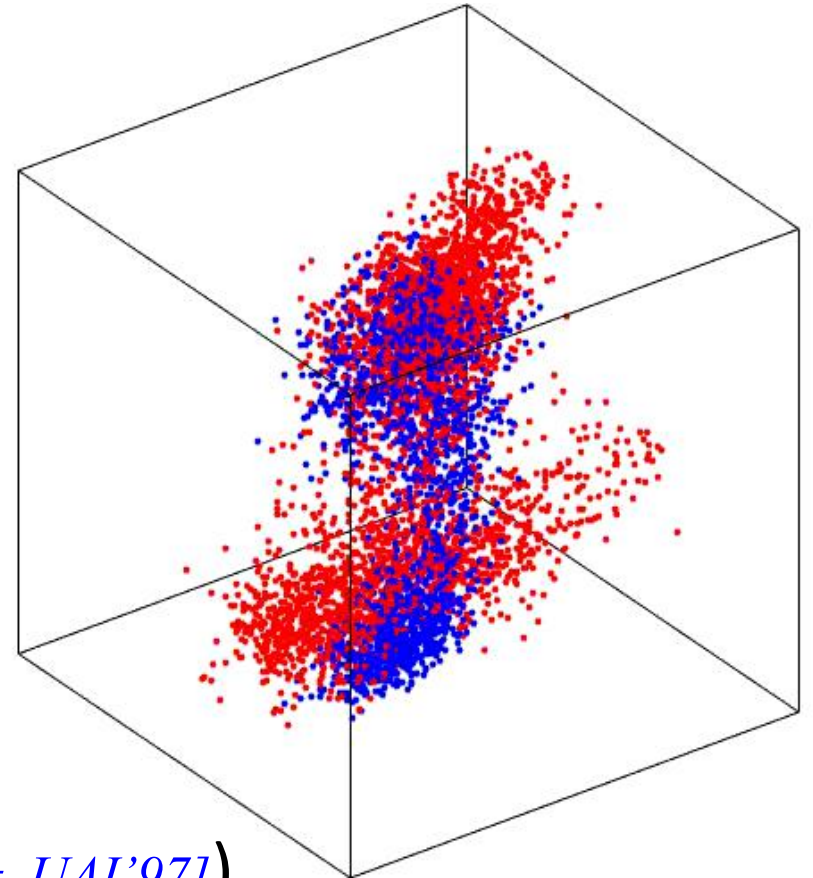


model fitting (e.g. GMM)

GrabCut [Rother, Kolmogorov, Blake, 2004]

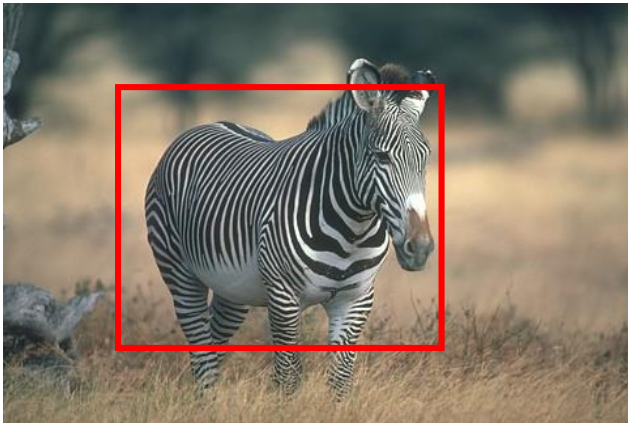
color space **clustering**

poor clustering
(*overfitting & local minima*)



Why *Normalized Cut* for MRF?

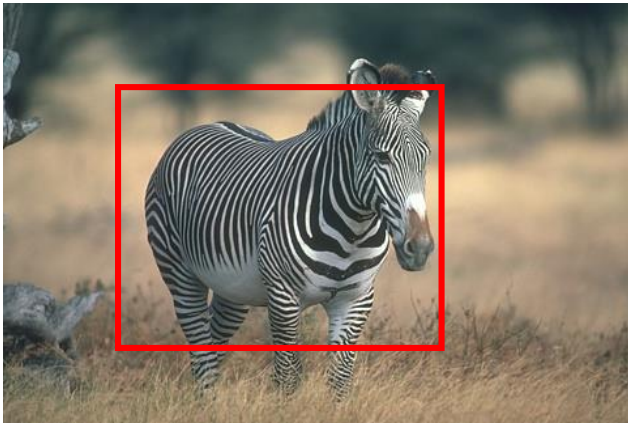
$$\underbrace{NC(S)}_{\text{NC for colors}} + \sum_{pq \in \mathcal{N}} w \cdot [s_p \neq s_q]$$



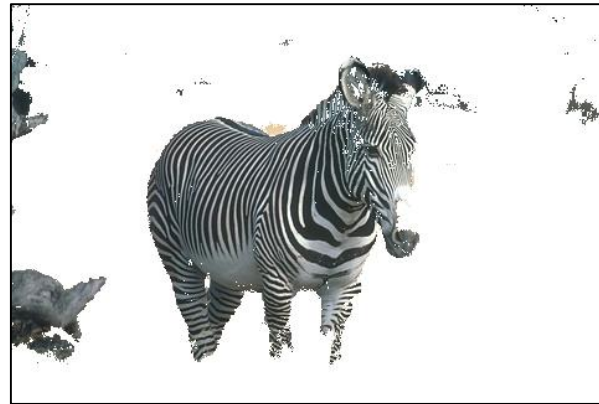
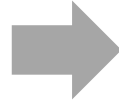
Normalized Cut

Why *Normalized Cut* for MRF?

$$\underbrace{NC(S)}_{\text{NC for colors}} + \sum_{pq \in \mathcal{N}} w_{pq} [s_p \neq s_q]$$

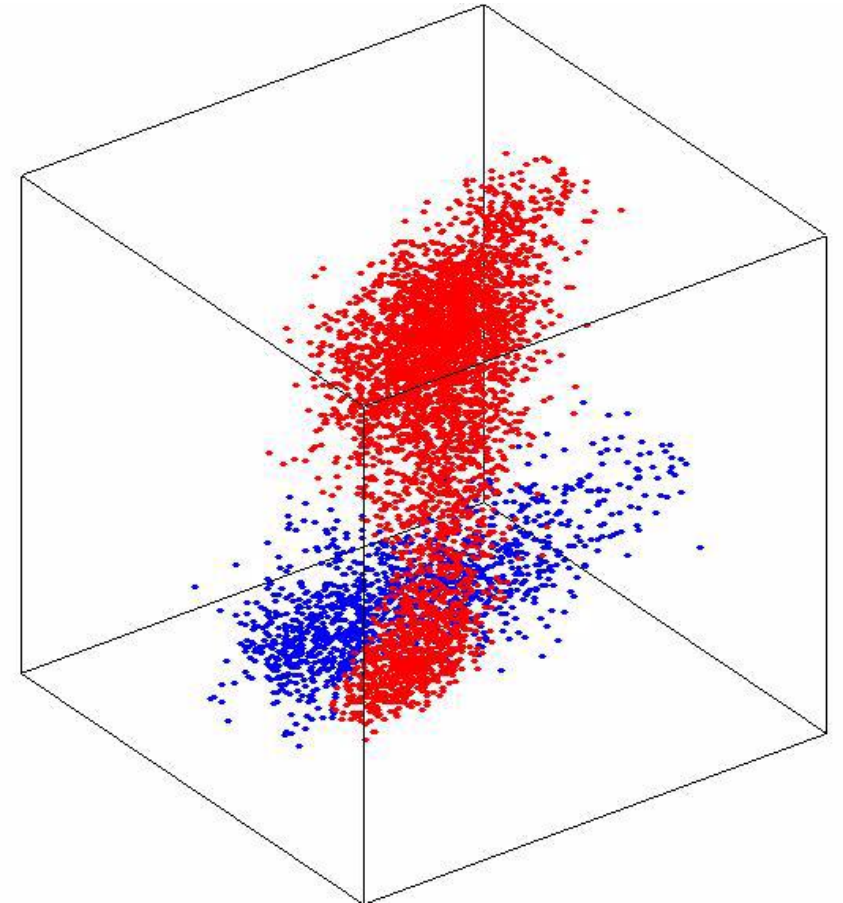


Normalized Cut



color space **clustering**

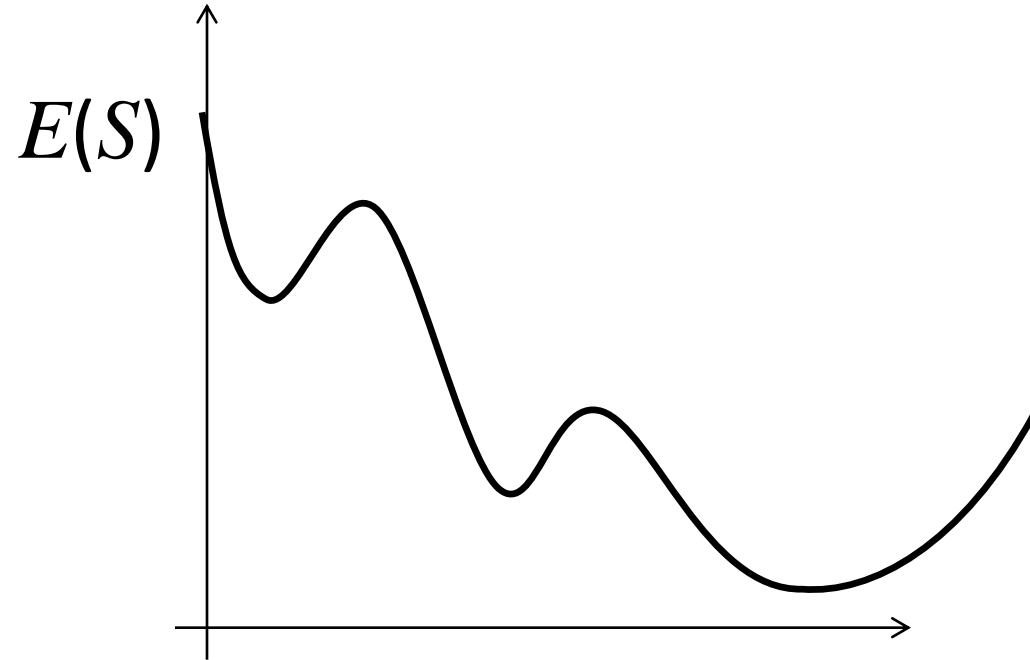
good clustering



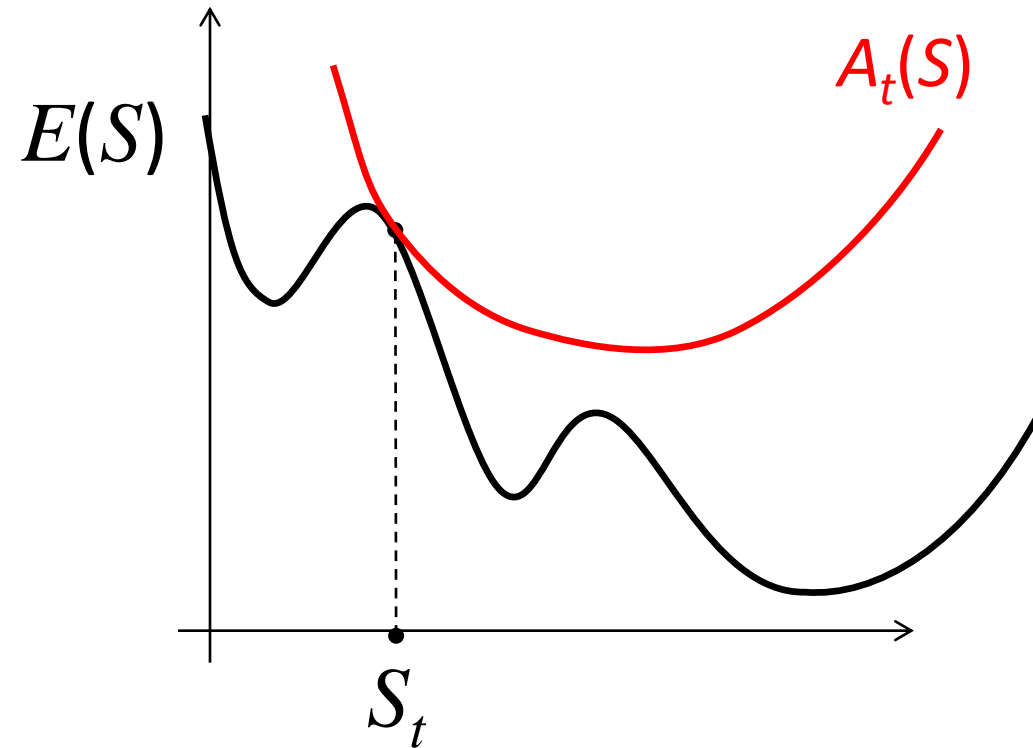
Our proposal: **Normalized Cut** + **MRF**

$$E(S) = \underbrace{\sum_k -\frac{S^{k'} A S^k}{d' S^k}}_{\text{balanced clustering}} + \gamma \underbrace{\sum_{c \in \mathcal{F}} E_c(S_c)}_{\text{regularization constraints}}$$

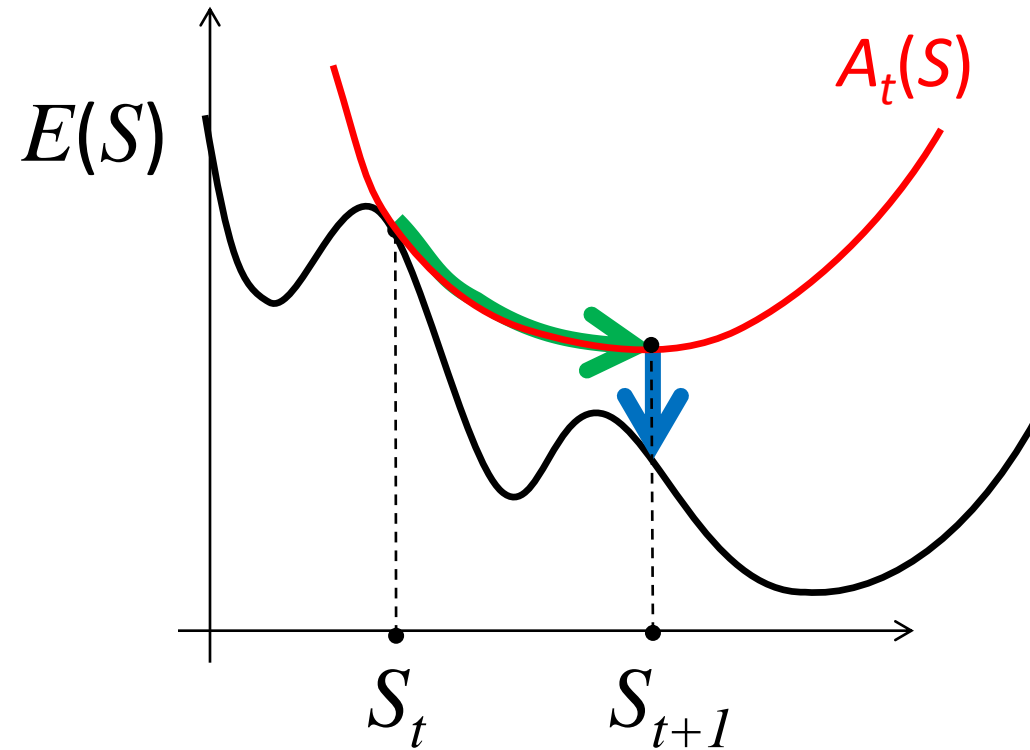
Bound optimization, in general



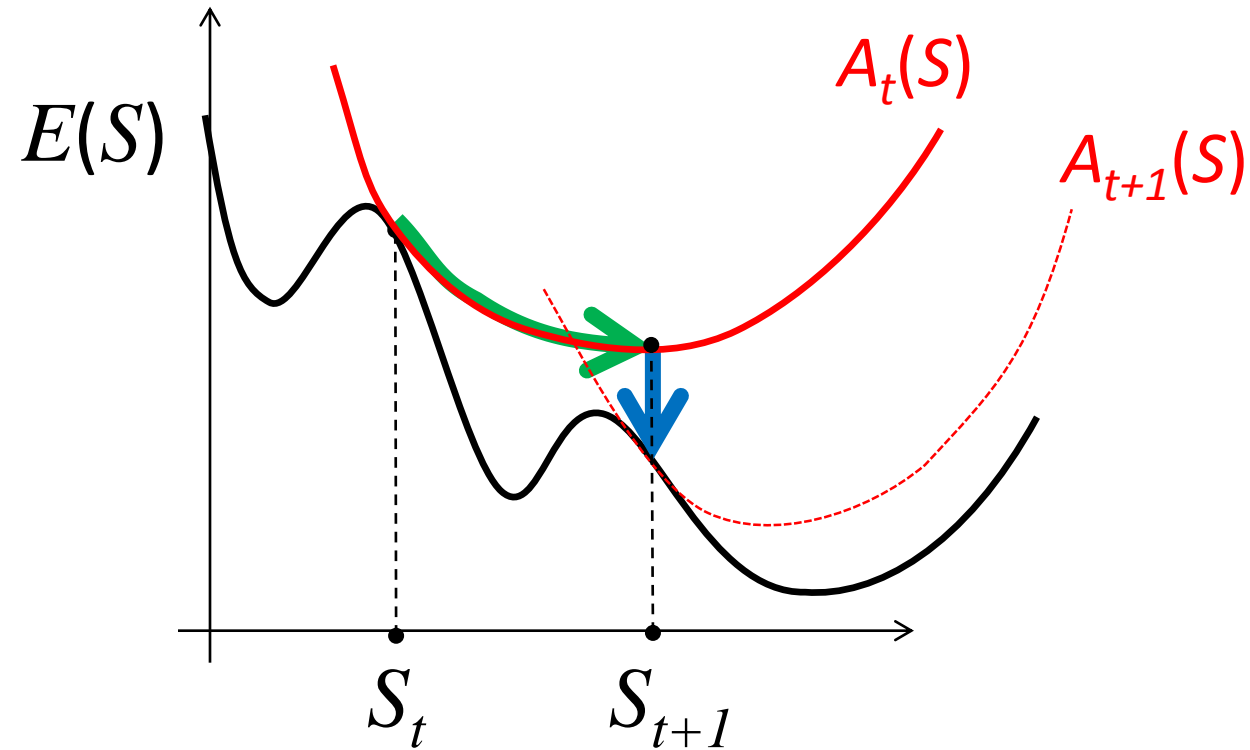
Bound optimization, in general



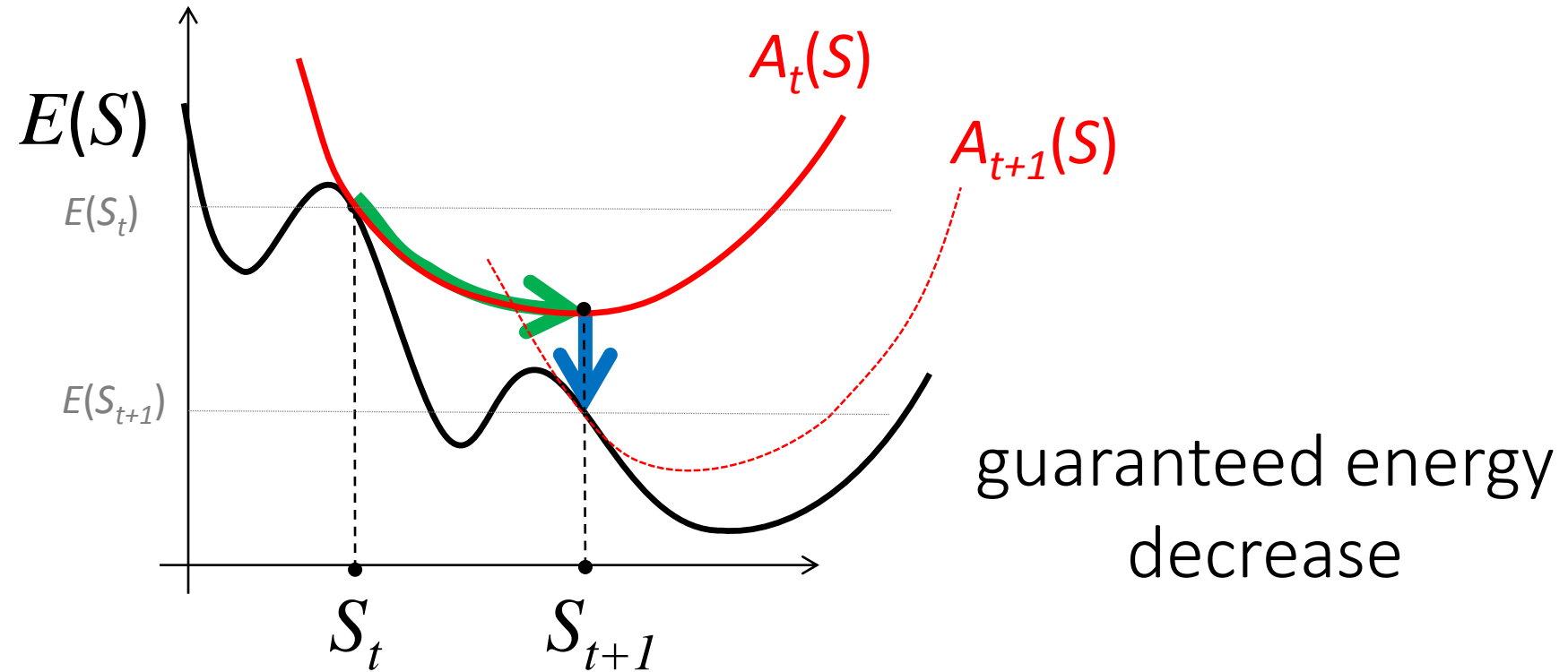
Bound optimization, in general



Bound optimization, in general



Bound optimization, in general



Bound for our joint energy

$$E(S) = \sum_k -\frac{S^{k'} A S^k}{d' S^k} + \gamma \sum_{c \in \mathcal{F}} E_c(S_c)$$

// \wedge unary bound for NC

$$A_t(S) = \sum_{p \in \Omega} U_p(S_p) + \gamma \sum_{c \in \mathcal{F}} E_c(S_c)$$

Bound for our joint energy

$$E(S) = \boxed{\sum_k -\frac{S^{k'} A S^k}{d' S^k}} + \boxed{\gamma \sum_{c \in \mathcal{F}} E_c(S_c)}$$

// \wedge unary bound for NC

$$A_t(S) = \boxed{\sum_{p \in \Omega} U_p(S_p)} + \boxed{\gamma \sum_{c \in \mathcal{F}} E_c(S_c)}$$

we propose **kernel bound** and **spectral bound** for NC

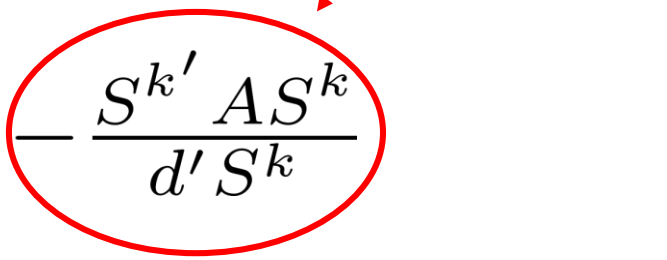
Kernel bound for NC

Lemma 1 (concavity)

$$NC(S) = \sum_k -\frac{S^{k'} AS^k}{d' S^k}$$

Kernel bound for NC

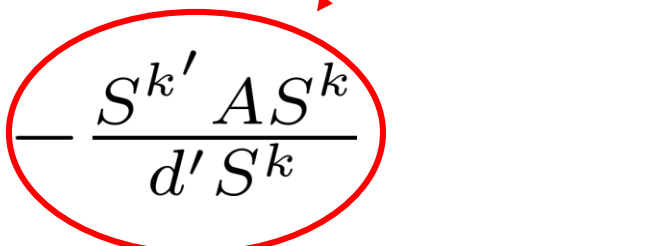
Lemma 1 (concavity)

$$NC(S) = \sum_k \frac{S^{k'} A S^k}{d' S^k}$$


$e(S^k)$

Kernel bound for NC

Lemma 1 (concavity)

$$NC(S) = \sum_k \frac{S^{k'} A S^k}{d' S^k}$$


$e(S^k)$

Function $e : \mathbf{R}^{|\Omega|} \rightarrow \mathbf{R}$ is concave over region $S^k > 0$ given p.s.d. affinity matrix $A := [A_{pq}]$.

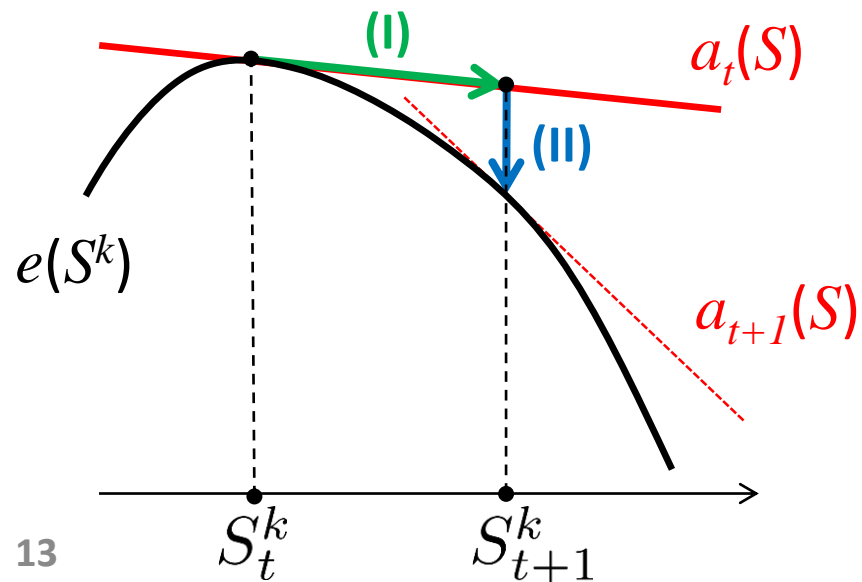
Kernel bound for NC

Lemma 1 (concavity)

$$NC(S) = \sum_k \frac{S^{k'} A S^k}{d' S^k}$$

$e(S^k)$

Function $e : \mathbf{R}^{|\Omega|} \rightarrow \mathbf{R}$ is concave over region $S^k > 0$ given p.s.d. affinity matrix $A := [A_{pq}]$.



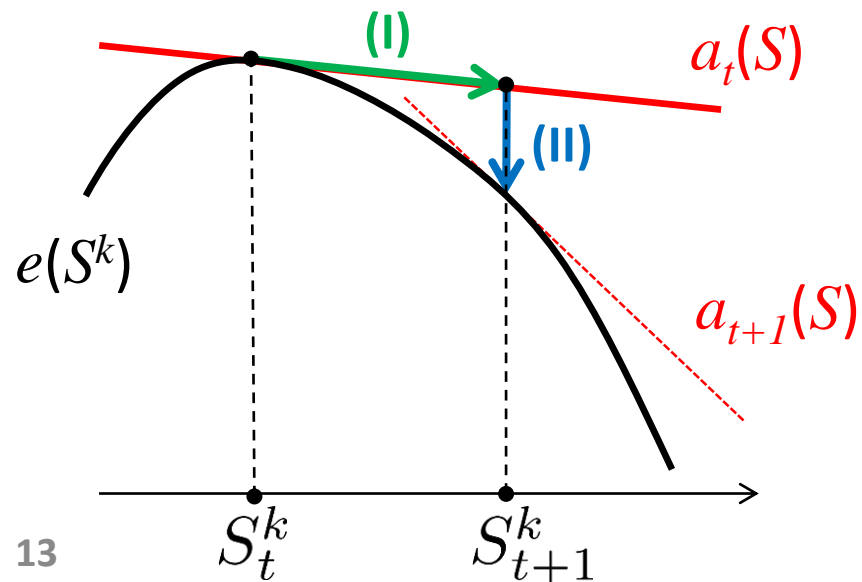
Kernel bound for NC

Lemma 1 (concavity)

$$NC(S) = \sum_k \frac{S^{k'} A S^k}{d' S^k}$$

$e(S^k)$

Function $e : \mathbf{R}^{|\Omega|} \rightarrow \mathbf{R}$ is concave over region $S^k > 0$ given p.s.d. affinity matrix $A := [A_{pq}]$.



first-order Taylor expansion:

$$e(S_t^k) + \nabla e(S_t^k) \cdot (S^k - S_t^k)$$

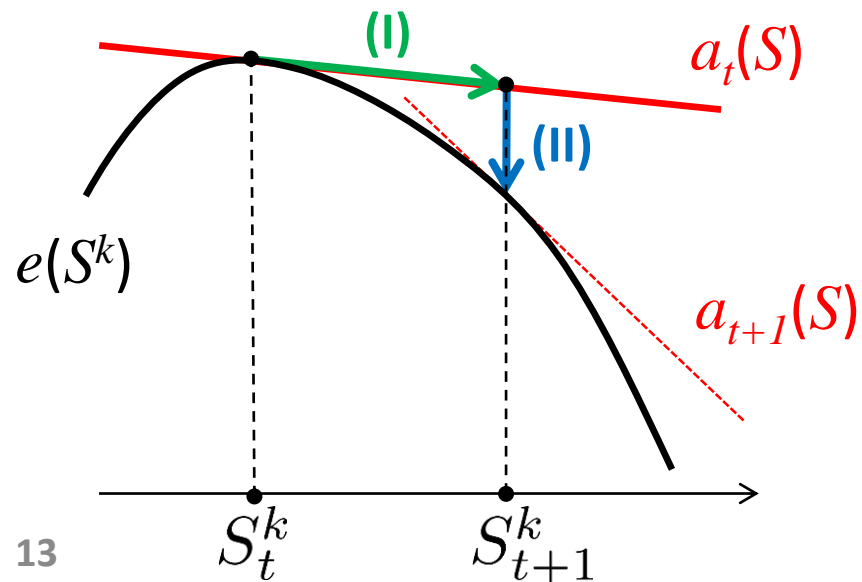
Kernel bound for NC

Lemma 1 (concavity)

$$NC(S) = \sum_k \frac{S^{k'} A S^k}{d' S^k}$$

$e(S^k)$

Function $e : \mathbf{R}^{|\Omega|} \rightarrow \mathbf{R}$ is concave over region $S^k > 0$ given p.s.d. affinity matrix $A := [A_{pq}]$.



first-order Taylor expansion:

$$e(S_t^k) + \nabla e(S_t^k) \cdot (S^k - S_t^k)$$

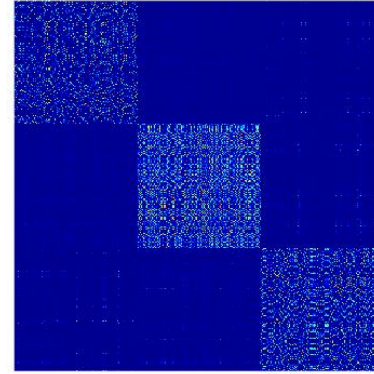
equivalently kernel k-means for NC [Dhillon et al., 2004]

Low-rank Approximation

Low-rank Approximation

Consider rank m approximation $\tilde{A} \approx A$:

$$\min_{\tilde{A}} \|A - \tilde{A}\|_F$$



A

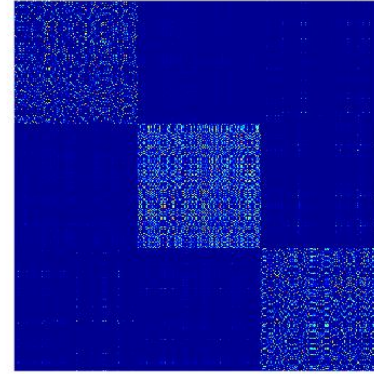
Low-rank Approximation

Consider rank m approximation $\tilde{A} \approx A$:

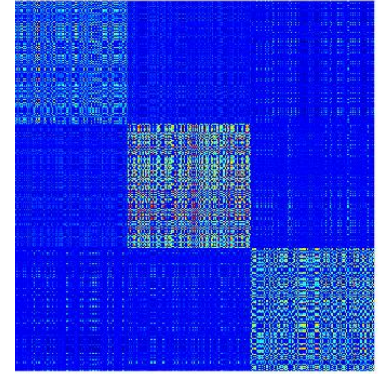
$$\min_{\tilde{A}} \|A - \tilde{A}\|_F$$

using top m eigenvectors and eigenvalues of A :

$$\tilde{A} = V^m \Lambda^m V^m$$



A



\tilde{A}

Low-rank Approximation

Consider rank m approximation $\tilde{A} \approx A$:

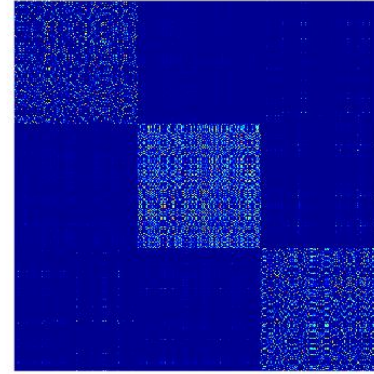
$$\min_{\tilde{A}} \|A - \tilde{A}\|_F$$

using top m eigenvectors and eigenvalues of A :

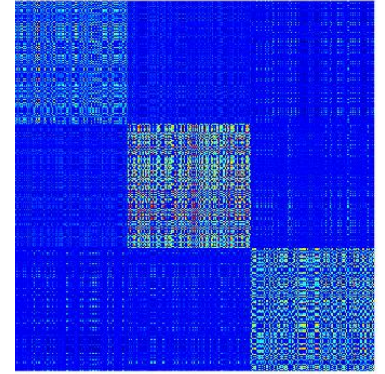
$$\tilde{A} = V^m \Lambda^m V^m$$

low-dimensional points:

$$\tilde{\phi}_p := \sqrt{\Lambda^m} V_p^m \in \mathcal{R}^m$$



A



\tilde{A}

Low-rank Approximation

Consider rank m approximation $\tilde{A} \approx A$:

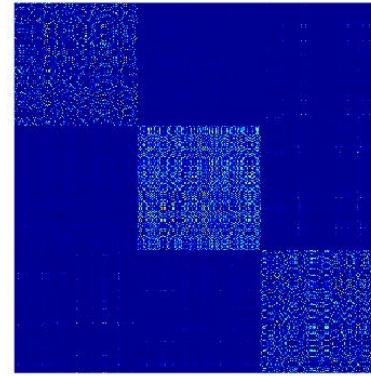
$$\min_{\tilde{A}} \|A - \tilde{A}\|_F$$

using top m eigenvectors and eigenvalues of A :

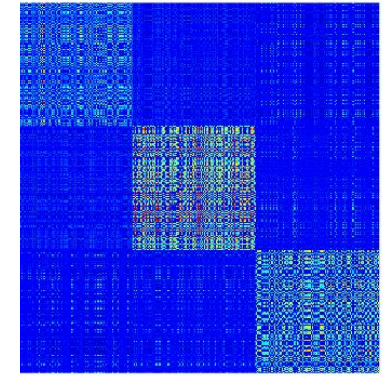
$$\tilde{A} = V^m \Lambda^m V^m$$

low-dimensional points:

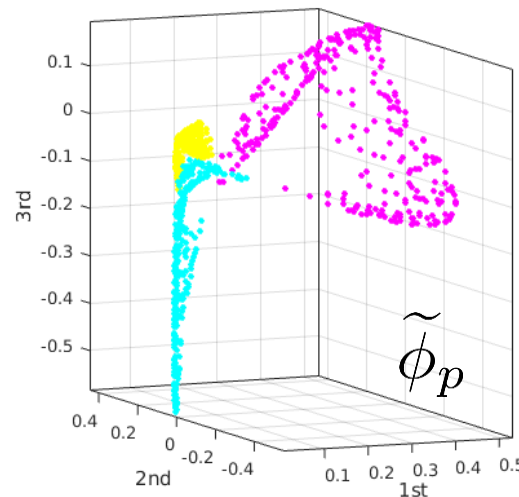
$$\tilde{\phi}_p := \sqrt{\Lambda^m} V_p^m \in \mathcal{R}^m$$



A



\tilde{A}



Low-rank Approximation

Consider rank m approximation $\tilde{A} \approx A$:

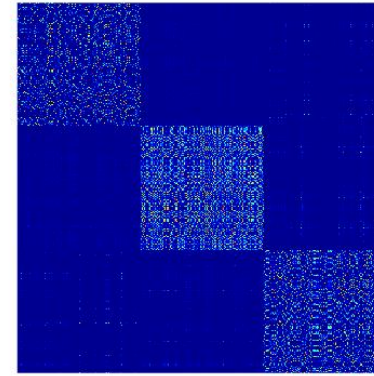
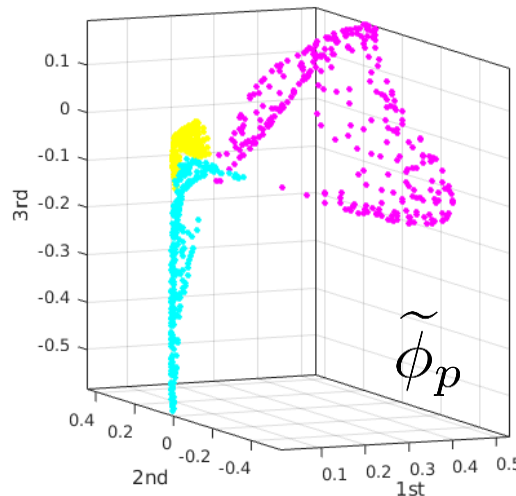
$$\min_{\tilde{A}} \|A - \tilde{A}\|_F$$

using top m eigenvectors and eigenvalues of A :

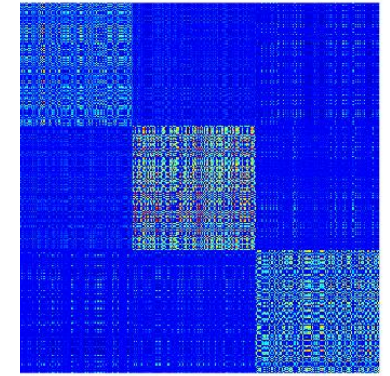
$$\tilde{A} = V^m \Lambda^m V^m$$

low-dimensional points:

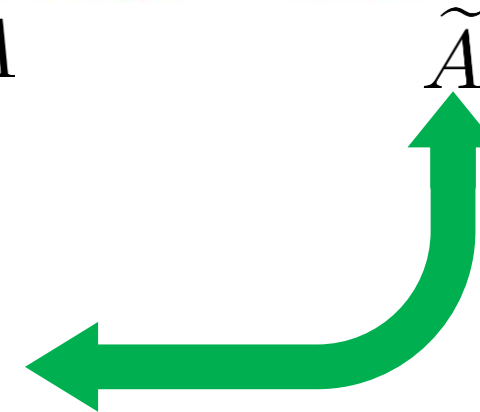
$$\tilde{\phi}_p := \sqrt{\Lambda^m} V_p^m \in \mathcal{R}^m$$



A



\tilde{A}



isometric Euclidean Embedding

$$A_{pq} \approx \tilde{A}_{pq} = \langle \tilde{\phi}_p, \tilde{\phi}_q \rangle$$

Low-rank Approximation

Consider rank m approximation $\tilde{A} \approx A$:

$$\min_{\tilde{A}} \|A - \tilde{A}\|_F$$

using top m eigenvectors and eigenvalues of A :

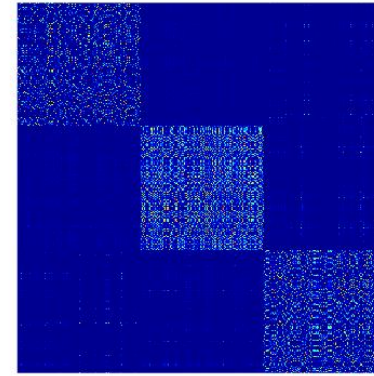
$$\tilde{A} = V^m \Lambda^m V^m$$

low-dimensional points:

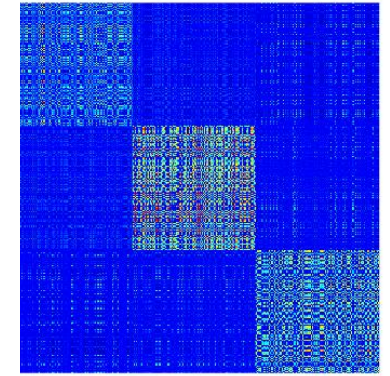
$$\tilde{\phi}_p := \sqrt{\Lambda^m} V_p^m \in \mathcal{R}^m$$

MDS [Cox & Cox., 2000]

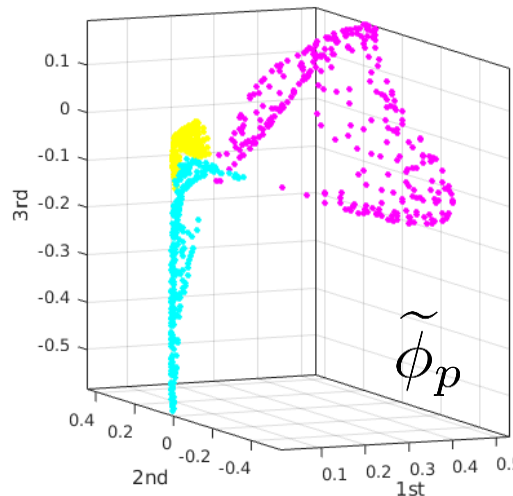
Kernel PCA [Schölkopf, Smola, Müller, 1998]



A



\tilde{A}



isometric Euclidean Embedding

$$A_{pq} \approx \tilde{A}_{pq} = \langle \tilde{\phi}_p, \tilde{\phi}_q \rangle$$

NC $\tilde{A} \approx A$

$$\sum_k \frac{S^{k'} A S^k}{d' S^k} \approx \sum_k \frac{S^{k'} \tilde{A} S^k}{d' S^k}$$

NC

$$\tilde{A} \approx A$$

$$\sum_k -\frac{S^{k'} A S^k}{d' S^k} \approx \sum_k -\frac{S^{k'} \tilde{A} S^k}{d' S^k} \stackrel{c}{=} \sum_k \sum_{p \in S^k} d_p \|\tilde{\phi}_p - \mu^k\|^2$$

K-means

NC

$$\tilde{A} \approx A$$

$$\sum_k -\frac{S^{k'} A S^k}{d' S^k}$$

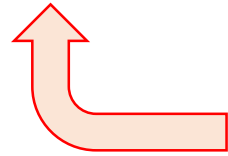


$$\approx \sum_k -\frac{S^{k'} \tilde{A} S^k}{d' S^k}$$

$$\stackrel{c}{=}$$

$$\sum_k \sum_{p \in S^k} d_p \|\tilde{\phi}_p - \mu^k\|^2$$

K-means



$$\mu^k := \text{mean}\{\tilde{\phi}_p \mid p \in S^k\}$$

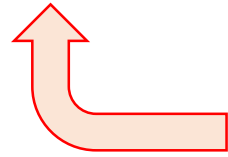
$$\tilde{A}_{pq} = \langle \tilde{\phi}_p, \tilde{\phi}_q \rangle$$

NC

$$\tilde{A} \approx A$$

K-means

$$\sum_k -\frac{S^{k'} A S^k}{d' S^k} \approx \sum_k -\frac{S^{k'} \tilde{A} S^k}{d' S^k} \stackrel{c}{=} \sum_k \sum_{p \in S^k} d_p \|\tilde{\phi}_p - \mu^k\|^2$$



$$\mu^k := \text{mean}\{\tilde{\phi}_p \mid p \in S^k\}$$

$$\tilde{A}_{pq} = \langle \tilde{\phi}_p, \tilde{\phi}_q \rangle$$

□ **NC = K-means** for exact embedding $\phi_p \in \mathcal{R}^{|\Omega|}$ [Bach & Jordan 2003, Dhillon et al., 2004]

NC $\tilde{A} \approx A$ **K-means**

$$\sum_k \frac{S^{k'} A S^k}{d' S^k} \approx \sum_k \frac{S^{k'} \tilde{A} S^k}{d' S^k} \stackrel{c}{=} \sum_k \sum_{p \in S^k} d_p \|\tilde{\phi}_p - \mu^k\|^2$$

$$\mu^k := \text{mean}\{\tilde{\phi}_p \mid p \in S^k\}$$

$$\tilde{A}_{pq} = \langle \tilde{\phi}_p, \tilde{\phi}_q \rangle$$

□ **NC = K-means** for exact embedding $\phi_p \in \mathcal{R}^{|\Omega|}$ [Bach & Jordan 2003, Dhillon et al., 2004]

□ We propose **NC \approx K-means** for **low-dimensional embedding** $\tilde{\phi}_p \in \mathcal{R}^m$ for $m \ll |\Omega|$

NC

$$\sum_k \frac{S^{k'} A S^k}{d' S^k}$$

K-means

$$\approx^c \sum_k \sum_{p \in S^k} d_p \|\tilde{\phi}_p - \mu^k\|^2$$

NC

K-means

$$\sum_k -\frac{S^{k'} A S^k}{d' S^k} \approx^c \sum_k \sum_{p \in S^k} d_p \|\tilde{\phi}_p - \mu^k\|^2$$

NC

K-means

$$\begin{aligned} \sum_k -\frac{S^{k'} A S^k}{d' S^k} &\stackrel{c}{\approx} \sum_k \sum_{p \in S^k} d_p \|\tilde{\phi}_p - \mu^k\|^2 \\ &= \sum_k \sum_{p \in \Omega} d_p \|\tilde{\phi}_p - \mu^k\|^2 \cdot S_p^k \end{aligned}$$

NC

K-means

$$\sum_k -\frac{S^{k'} A S^k}{d' S^k} \approx^c \sum_k \sum_{p \in S^k} d_p \|\tilde{\phi}_p - \mu^k\|^2$$

unary spectral bound for NC

$$\leq \sum_k \sum_{p \in \Omega} d_p \|\tilde{\phi}_p - \mu_t^k\|^2 \cdot S_p^k$$

NC

K-means

$$\sum_k -\frac{S^{k'} A S^k}{d' S^k} \approx^c \sum_k \sum_{p \in S^k} d_p \|\tilde{\phi}_p - \mu^k\|^2$$

unary spectral bound for NC

$$\leq \sum_k \sum_{p \in \Omega} d_p \|\tilde{\phi}_p - \mu_t^k\|^2 \cdot S_p^k$$

□ **K-means** on $\phi_p^* \in \mathcal{R}^K$ - discretization heuristic for a spectral relaxation of **NC**

[Shi & Malik 2000, Yu & Shi 2003]

NC

K-means

$$\sum_k -\frac{S^{k'} A S^k}{d' S^k} \approx^c \sum_k \sum_{p \in S^k} d_p \|\tilde{\phi}_p - \mu^k\|^2$$

unary spectral bound for NC

$$\leq \sum_k \sum_{p \in \Omega} d_p \|\tilde{\phi}_p - \mu_t^k\|^2 \cdot S_p^k$$

□ **K-means** on $\phi_p^* \in \mathcal{R}^K$ - discretization heuristic for a spectral relaxation of **NC**

[Shi & Malik 2000, Yu & Shi 2003]

□ We justify **NC** \approx **K-means** via low-rank (m) approximation using $\tilde{\phi}_p \in \mathcal{R}^m$

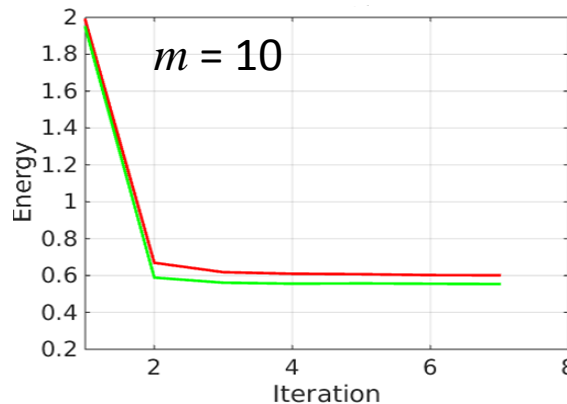
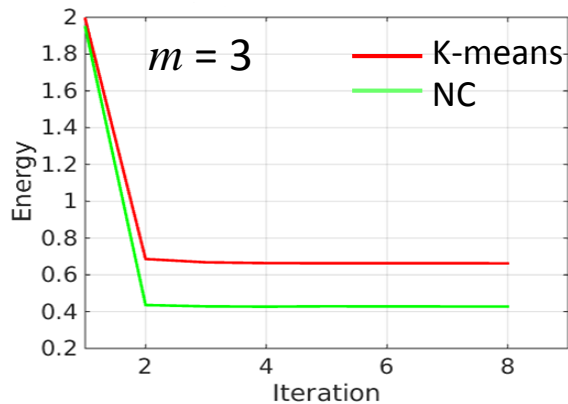
NC

K-means

$$\sum_k -\frac{S^{k'} AS^k}{d' S^k} \approx^c \sum_k \sum_{p \in S^k} d_p \|\tilde{\phi}_p - \mu^k\|^2$$

unary spectral bound for NC

$$\leq \sum_k \sum_{p \in \Omega} d_p \|\tilde{\phi}_p - \mu_t^k\|^2 \cdot S_p^k$$



□ We justify **NC** \approx **K-means** via low-rank (m) approximation using $\tilde{\phi}_p \in \mathcal{R}^m$

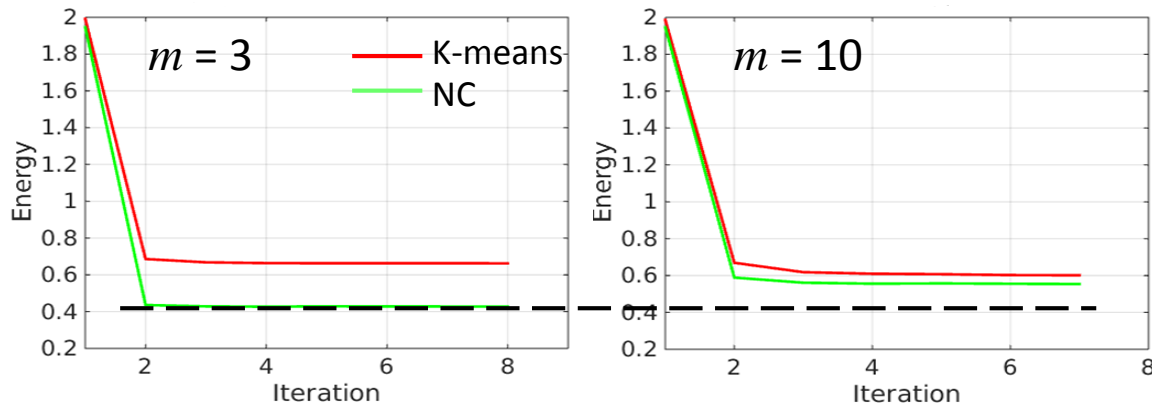
NC

K-means

$$\sum_k -\frac{S^{k'} AS^k}{d' S^k} \approx^c \sum_k \sum_{p \in S^k} d_p \|\tilde{\phi}_p - \mu^k\|^2$$

unary spectral bound for NC


$$\leq \sum_k \sum_{p \in \Omega} d_p \|\tilde{\phi}_p - \mu_t^k\|^2 \cdot S_p^k$$




□ We justify **NC** \approx **K-means** via low-rank (m) approximation using $\tilde{\phi}_p \in \mathcal{R}^m$

$$E(S) = \sum_k -\frac{S^{k'} A S^k}{d' S^k} + \gamma \sum_{c \in \mathcal{F}} E_c(S_c)$$

// \wedge unary bound for NC

iterate 

$$A_t(S) = \sum_{p \in \Omega} U_p(S_p) + \gamma \sum_{c \in \mathcal{F}} E_c(S_c)$$





$$S_{t+1} = \arg \min_S A_t(S) \text{ (move-making and graph cuts [Boykov, Veksler, Zabih, 2001])}$$

Our Kernel Cut and Spectral Cut

$$E(S) = \sum_k -\frac{S^{k'} A S^k}{d' S^k} + \gamma \sum_{c \in \mathcal{F}} E_c(S_c)$$

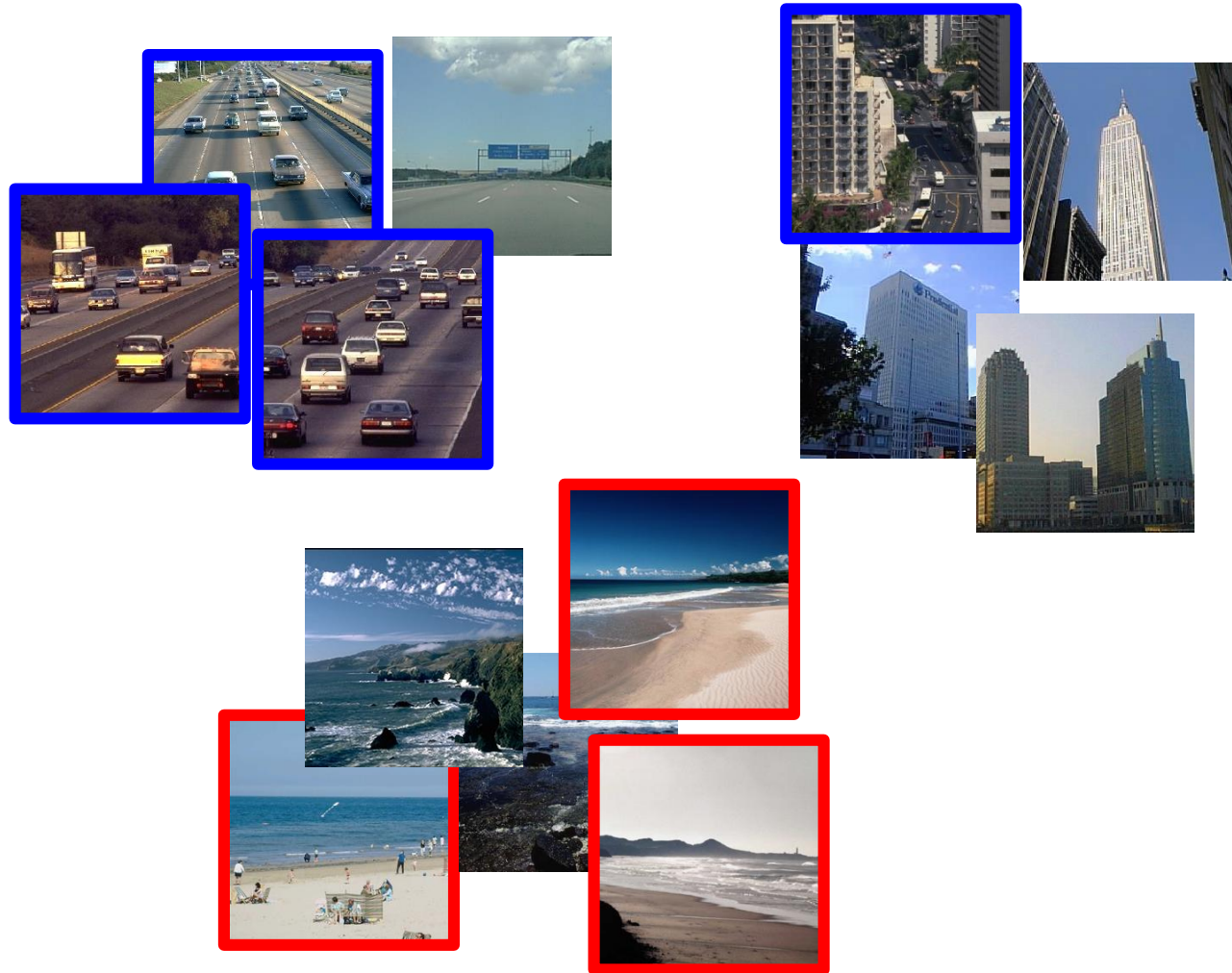
// \wedge unary bound for NC (Kernel Bound or Spectral Bound)


$$A_t(S) = \sum_{p \in \Omega} U_p(S_p) + \gamma \sum_{c \in \mathcal{F}} E_c(S_c)$$


$$S_{t+1} = \arg \min_S A_t(S) \text{ (move-making and graph cuts [Boykov, Veksler, Zabih, 2001])}$$

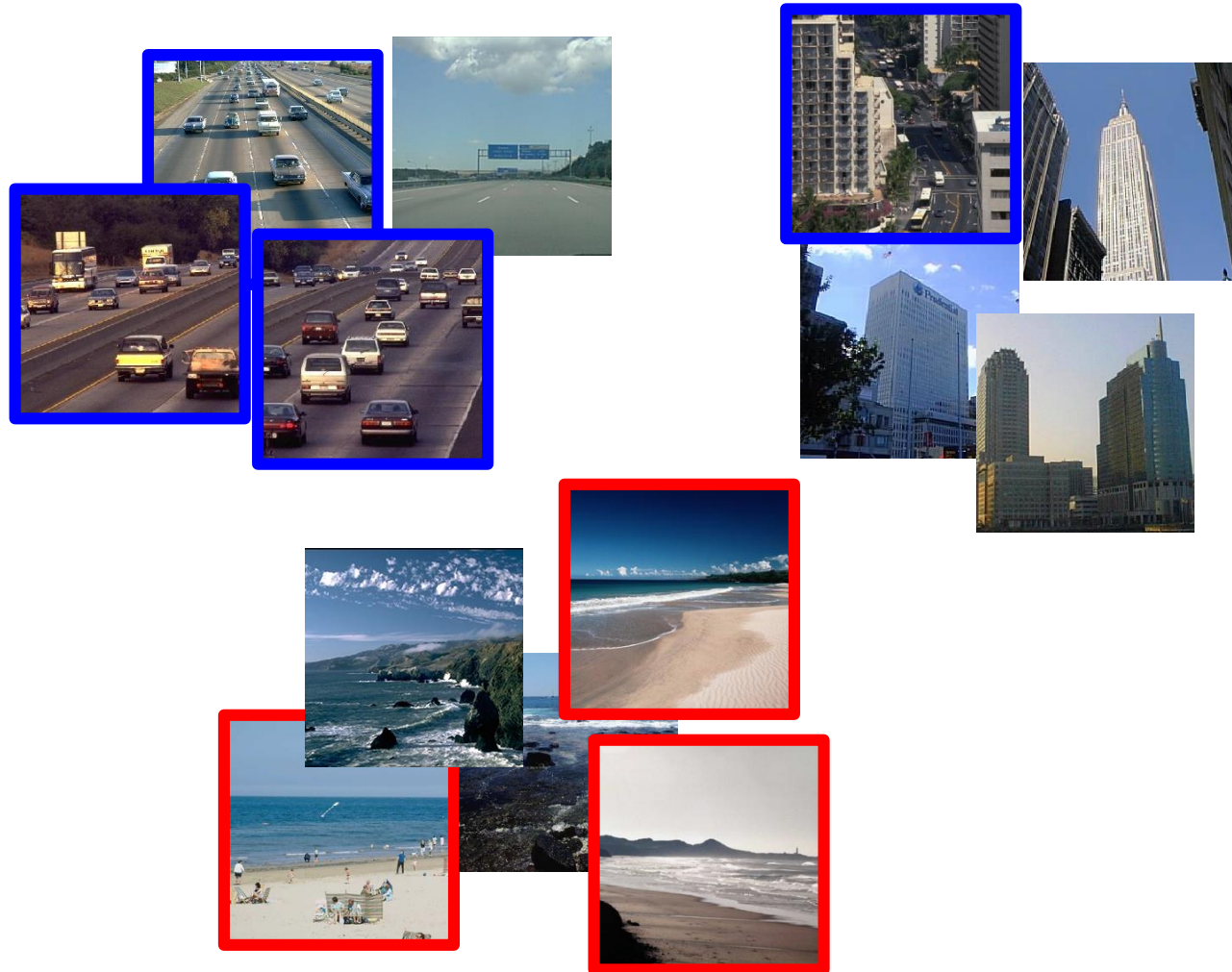
Experiments: MRF helps Normalized Cut

using image tags (e.g. **beach**, **car**) to help image clustering

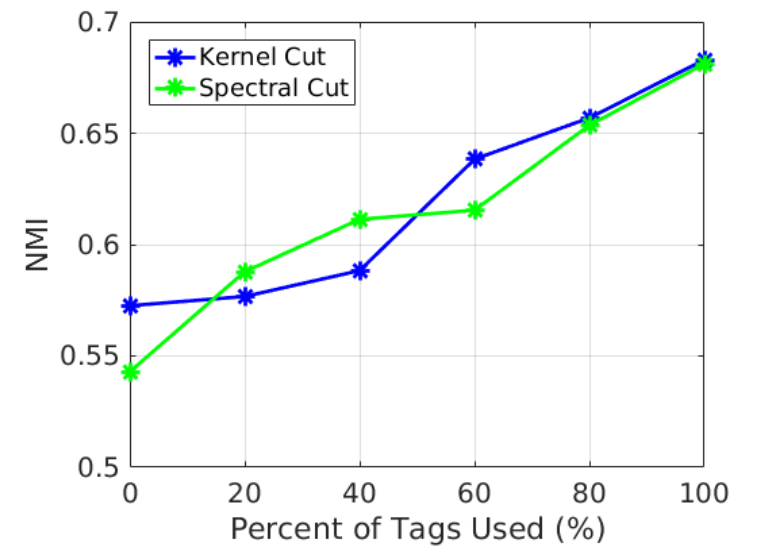


Experiments: MRF helps Normalized Cut

using image tags (e.g. **beach**, **car**) to help image clustering

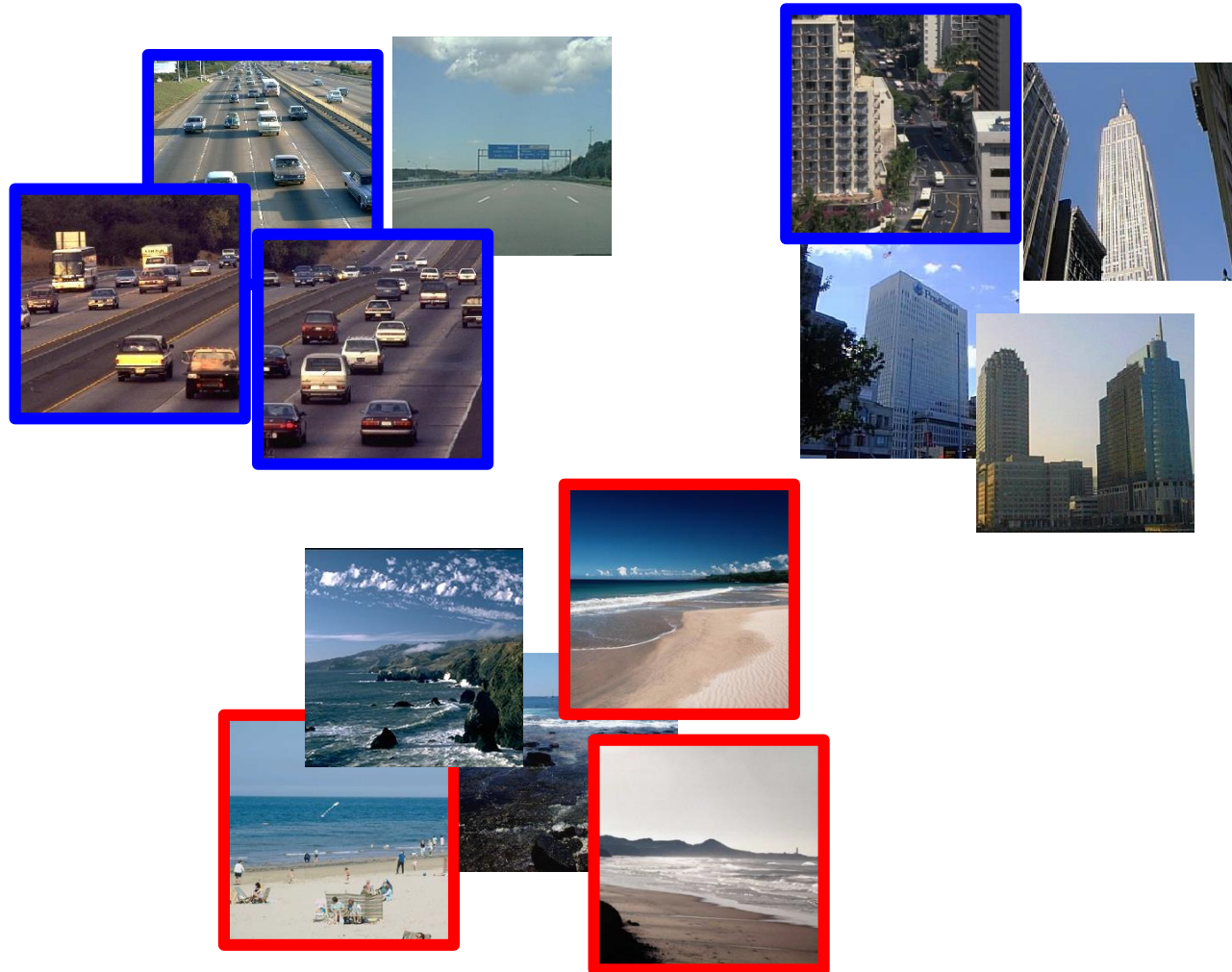


NC + robust P^n Potts

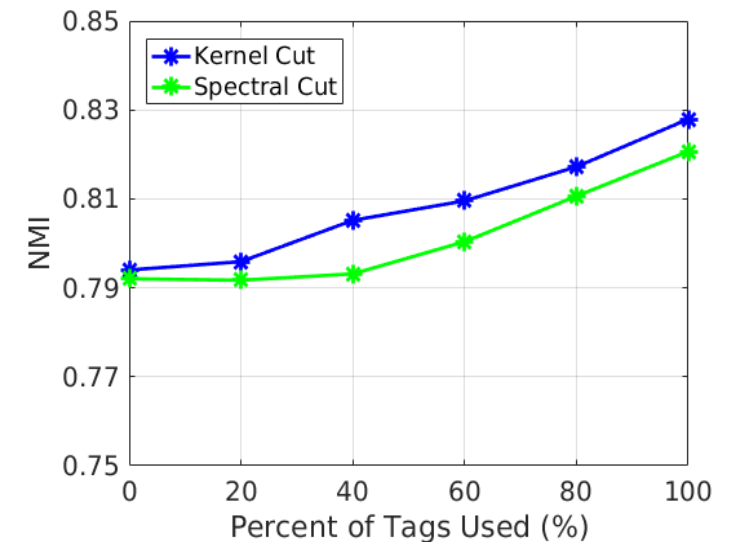


Experiments: MRF helps Normalized Cut

using image tags (e.g. **beach**, **car**) to help image clustering



NC + robust P^n Potts



⁺with *knn* kernel on deep features

Experiments: Normalized Cut helps MRF

(a) Video frames

(b) Optical flow

(c) Our Kernel Cut (NC + Potts)

Fig. motion segmentation using RGB, location (XY) and motion (M). “+xy” means with MRF

Experiments: Normalized Cut helps MRF

(a) Video frames

(b) Optical flow

(c) Our Kernel Cut (NC + Potts)

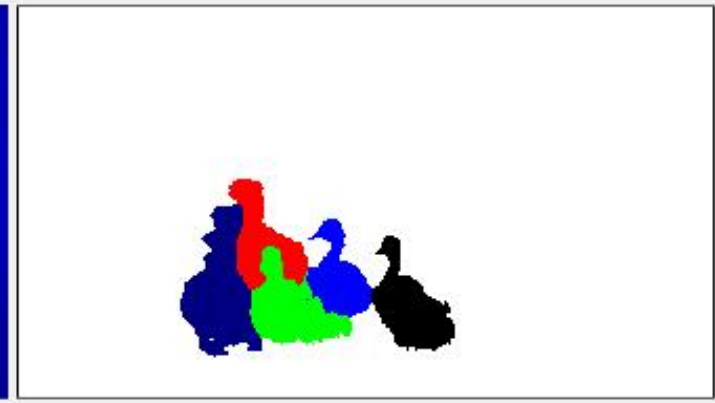
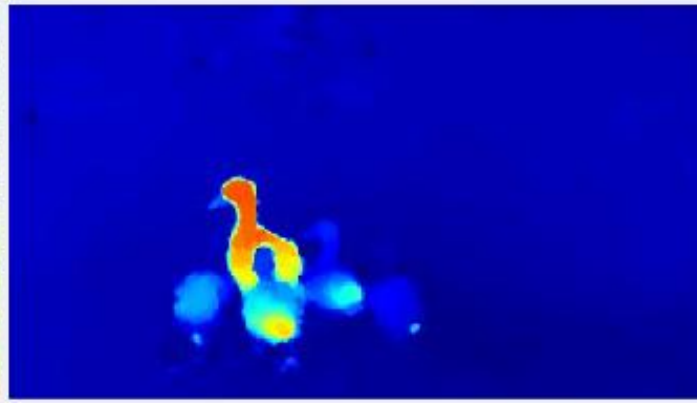
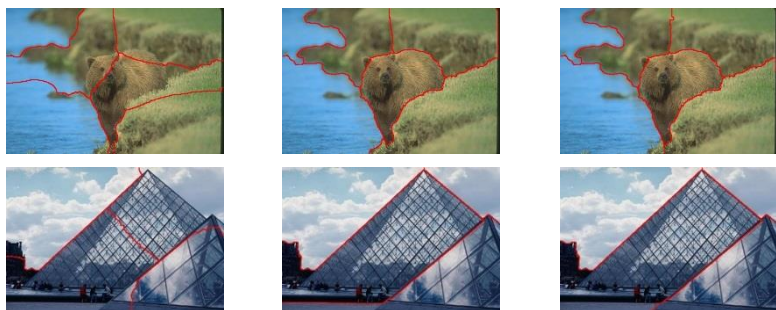


Fig. motion segmentation using RGB, location (XY) and motion (M). “+xy” means with MRF

More Experiments

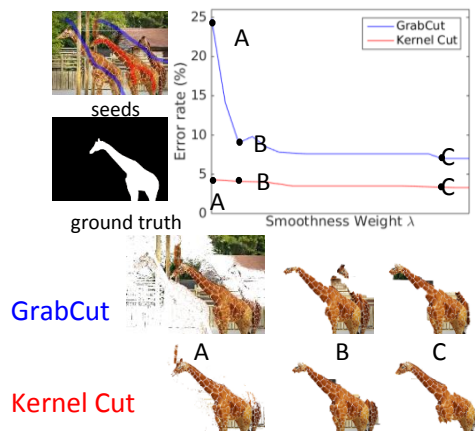
Potts model improves edge alignment



NC with increasing **label cost** →

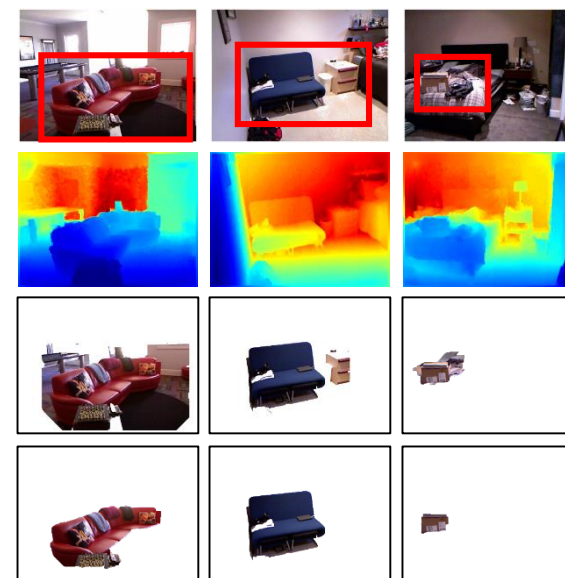


Spectral Clustering (no Potts) Our Kernel Cut (NC with Potts) Our Spectral Cut (NC with Potts)

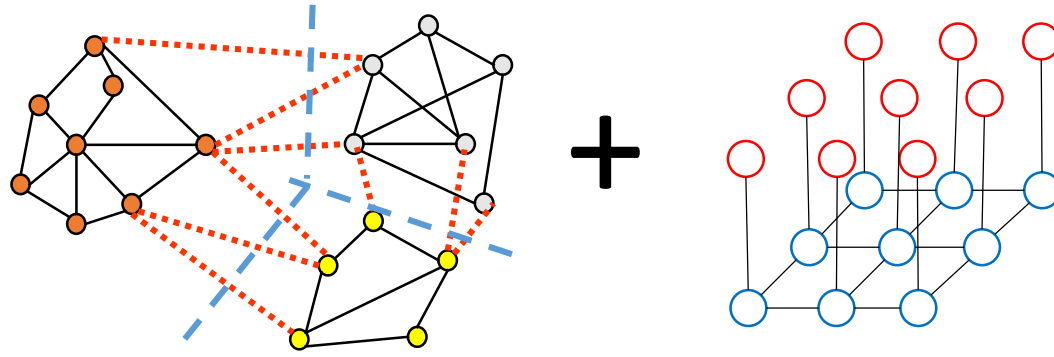


separating similar objects

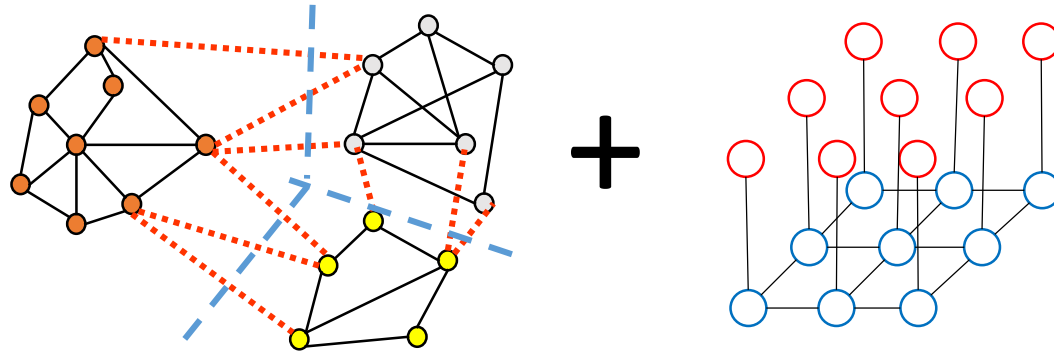
Fig. 1. RGBD segmentation



Conclusion

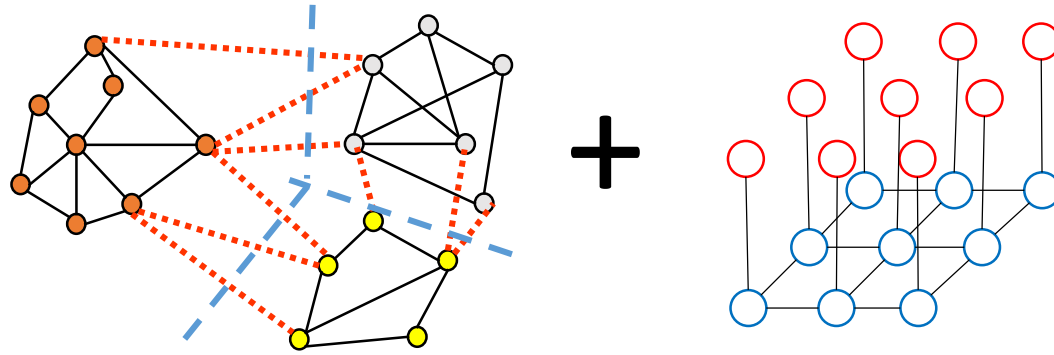


Conclusion



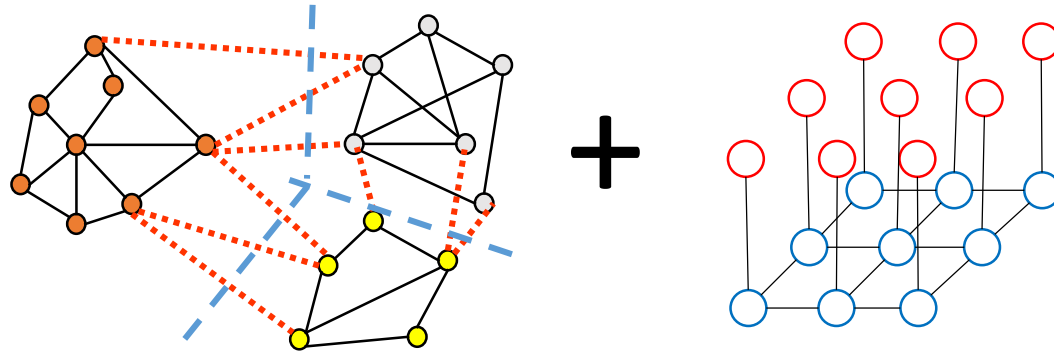
□ new unary kernel and spectral bounds for NC

Conclusion



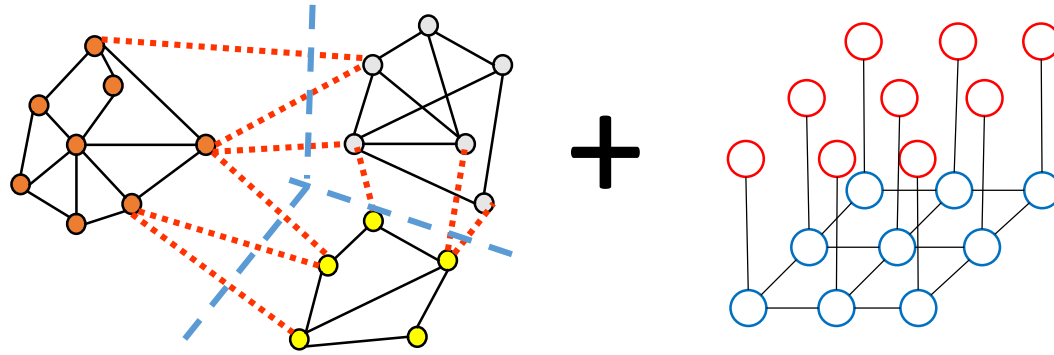
- ❑ new unary kernel and spectral bounds for NC
- ❑ can combine NC with any MRF constraints

Conclusion



- ❑ new unary kernel and spectral bounds for NC
- ❑ can combine NC with any MRF constraints
- ❑ can combine MRF with balanced clustering

Conclusion



- ❑ new unary kernel and spectral bounds for NC
- ❑ can combine NC with any MRF constraints
- ❑ can combine MRF with balanced clustering
- ❑ MRF with features of any dimension (RGBD, RGBM, RGBXYM, deep,...)