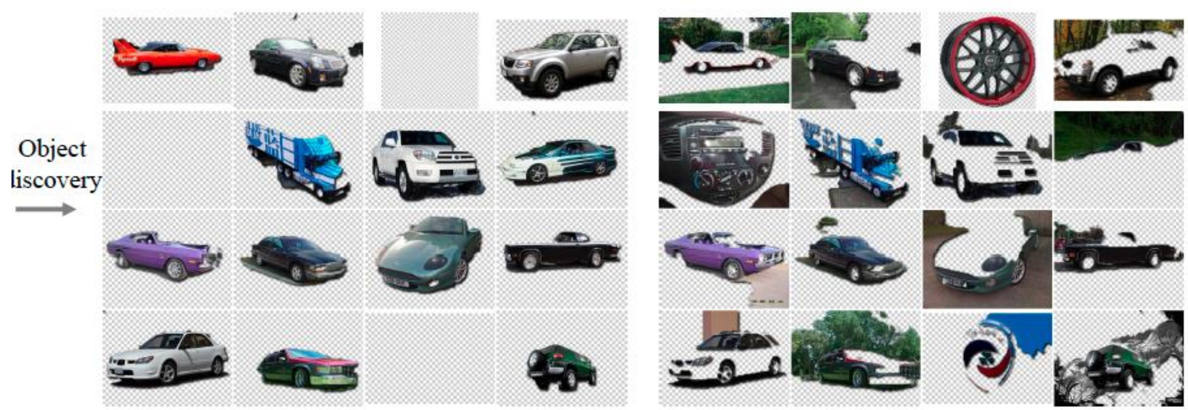
Unsupervised Joint Object Discovery and Segmentation in Internet Images
CVPR ‘13

M.Ilker SARAC
BIL722 Spring 2014
03.03.2014

(a) Images downloaded from the Internet



(b) Our automatic segmentation results

(c) State-of-the-art co-segmentation results [8]

Content

- Introduction
- Related Work
- Object Discovery and Segmentation
  - Saliency
  - Pixel correspondence
  - Foreground Likelihood
  - Regularization
  - Optimization
- Results

Introduction

- The goal is; label each pixel in a set of images if they belong to underlying common object.
- The task is; *co-segmentation*, jointly segment recurring objects that are common in multiple images.
- The environment is; *internet image collections*, lots of images and diverse results.
  - Challenging, many noise images
  - (not anymore? - Google)

Introduction

- In this paper;
  - A correspondence based *object discovery* and *co-segmentation* algorithm is proposed.
- The assumption is; pixels belonging to common object should be;
  - *Salient*
  - *Sparse*

Related Work

- Object Discovery
  - In a supervised setup – LDA (Russell *et al.*, Sivic *et al.*)
  - Generative model for the distribution of mask, edge and color. (Winn and Jojic)
  - VisualRank (Jing and Baluja)

Related Work

- Co-segmentation
  - Cosegmentation of image pairs by histogram matching, Rother *et al.*, CVPR '06
  - Discriminative clustering for image cosegmentation, Joulin *et al.*, CVPR '10
  - Segmentation propagation in imagenet, Kuettel *et al.*, ECCV '12

Object Discovery and Segmentation

- Dataset, N images
  - $\mathbf{I} = \{I_1, \ldots, I_N\}$
- Goal, compute binary masks for each image I_i
  - $\mathbf{B} = \{b_1, \ldots, b_N\}$
- Pixel $\mathbf{x} = (x, y)$
  - $b_i(x) = 1$, indicates foreground
  - $b_i(x) = 0$, indicates background at location $\mathbf{x}$.

Image saliency

- Contrast-based saliency(Cheng *et al.*)
  - Based on color contrast; how different a pixel is from other pixels
- Given a saliency map for each image they compute dataset-wide normalized saliency M_i and define the term below;

$$F^{i}_{saliency}(x) = -\log M_i(x)$$

Pixel Correspondence

- Establish reliable correspondence between pixels in different images.
  - SIFT flow
- A new objective function encourages matching foreground pixels between images.

$$E\left(\mathbf{w}_{ij};\mathbf{b}_i,\mathbf{b}_j\right)=$$

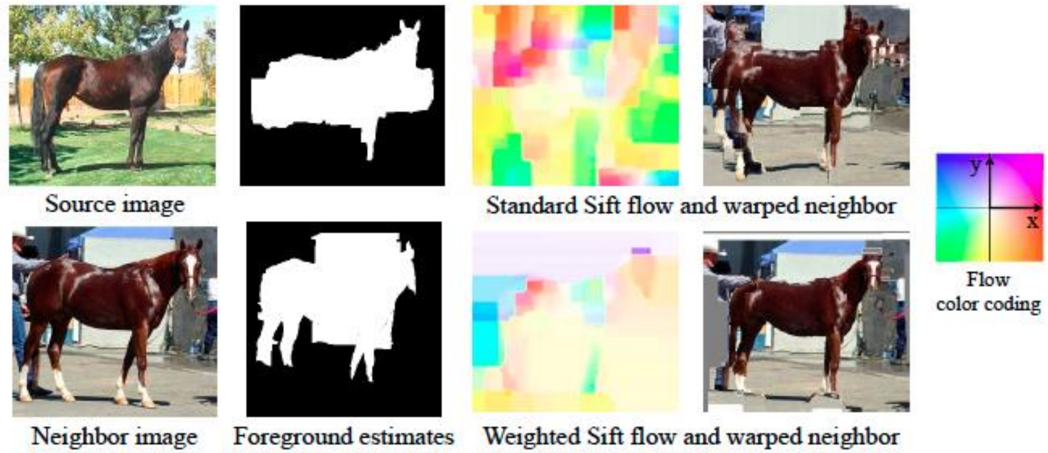
$$\sum_{\mathbf{x}\in\Lambda_i}\mathbf{b}_i(\mathbf{x})\Big(\mathbf{b}_j(\mathbf{x}+\mathbf{w}_{ij}(\mathbf{x}))\left\|S_i(\mathbf{x})-S_j(\mathbf{x}+\mathbf{w}_{ij}(\mathbf{x}))\right\|_1$$

$$+\left(1-\mathbf{b}_j(\mathbf{x}+\mathbf{w}_{ij}(\mathbf{x})\right)C_0+\sum_{\mathbf{y}\in\mathcal{N}_{\mathbf{x}}^i}\alpha\left\|\mathbf{w}_{ij}(\mathbf{x})-\mathbf{w}_{ij}(\mathbf{y})\right\|_2\Big),\quad(2)$$

Pixel Correspondence

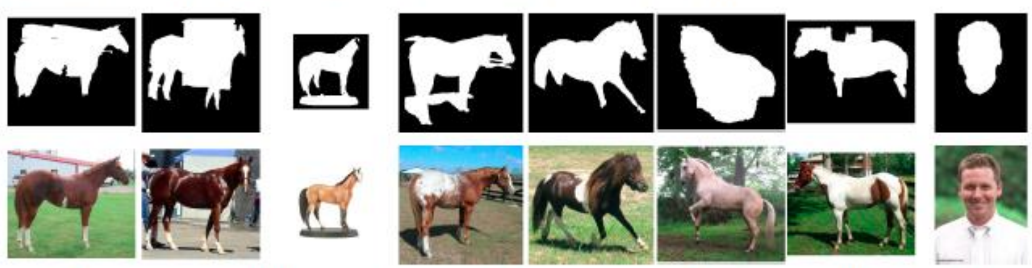
- For large datasets first find for each image I_i a set of similar images N_i.
  - Global image statistics
  - Weighted GIST
- Based on computed correspondence, the matching term is defined;

$$\widehat{\Phi^i_{match}}(\mathbf{x}) = \frac{1}{|N_i|} \sum_{j \in N_i} \|S_i(\mathbf{x}) - S_j(\mathbf{x} + \mathbf{w}_{ij}(\mathbf{x})\|_1$$



(a) Comparison between standard and weighted Sift flow.

(b) Nearest neighbor ordering (left to right) for the source image in (a), computed with the standard Gist descriptor.



(c) Nearest neighbor ordering (bottom row; left to right) for the source image in (a), computed with a weighted Gist descriptor using the foreground estimates (top row).

Foreground Likelihood

- Use the saliency and matching term to define the likelihood of a pixel label.

$$\Phi^i(\mathbf{x}) = \begin{cases} \Phi^i_{saliency}(\mathbf{x}) + \lambda_{match}\Phi^i_{match}(\mathbf{x}), & \mathbf{b}_i(\mathbf{x}) = 1, \\ \beta, & \mathbf{b}_i(\mathbf{x}) = 0, \end{cases} \quad (4)$$

Regularization

- *Intra-image compatibility*

$$\Psi^{i}_{int}(\mathbf{x}, \mathbf{y}) = [\mathbf{b}_i(\mathbf{x}) \neq \mathbf{b}_i(\mathbf{y})] \exp\left(-\|I_i(\mathbf{x}) - I_i(\mathbf{y})\|_2^2\right), \quad (5)$$

- *Inter-image compatibility*

$$\Psi^{ij}_{ext}(\mathbf{x}, \mathbf{y}) = [\mathbf{b}_i(\mathbf{x}) \neq \mathbf{b}_j(\mathbf{y})] \exp\left(-\|S_i(\mathbf{x}) - S_j(\mathbf{y})\|_1\right). \quad (6)$$

Regularization

- Learn the color histogram

$$\Phi^{i}_{color}(\mathbf{x}, \mathbf{h}_i) = -\log \mathbf{h}_i^{\mathbf{b}_i(\mathbf{x})}(\mathbf{x})$$

- Combining all these, cost function;

$$E(\mathbf{B}; \mathbf{W}, \mathbf{H}) = \sum_{i=1}^{N} \sum_{\mathbf{x} \in \Lambda_i} \Big(\Phi^{i}(\mathbf{x}) + \lambda_{color} \Phi^{i}_{color}(\mathbf{x}, \mathbf{h}_i)$$

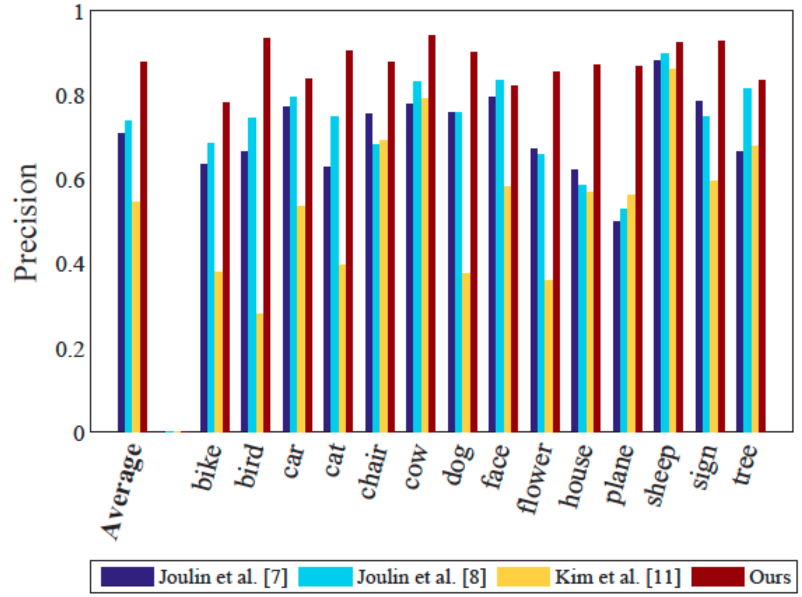
$$+ \sum_{\mathbf{y} \in \mathcal{N}^{i}_{\mathbf{x}}} \lambda_{int} \Psi^{i}_{int}(\mathbf{x}, \mathbf{y}) + \sum_{j \in \mathcal{N}_i} \lambda_{ext} \Psi^{ij}_{ext}(\mathbf{x}, \mathbf{x} + \mathbf{w}_{ij}(\mathbf{x})) \Big). \quad (8)$$

Optimization

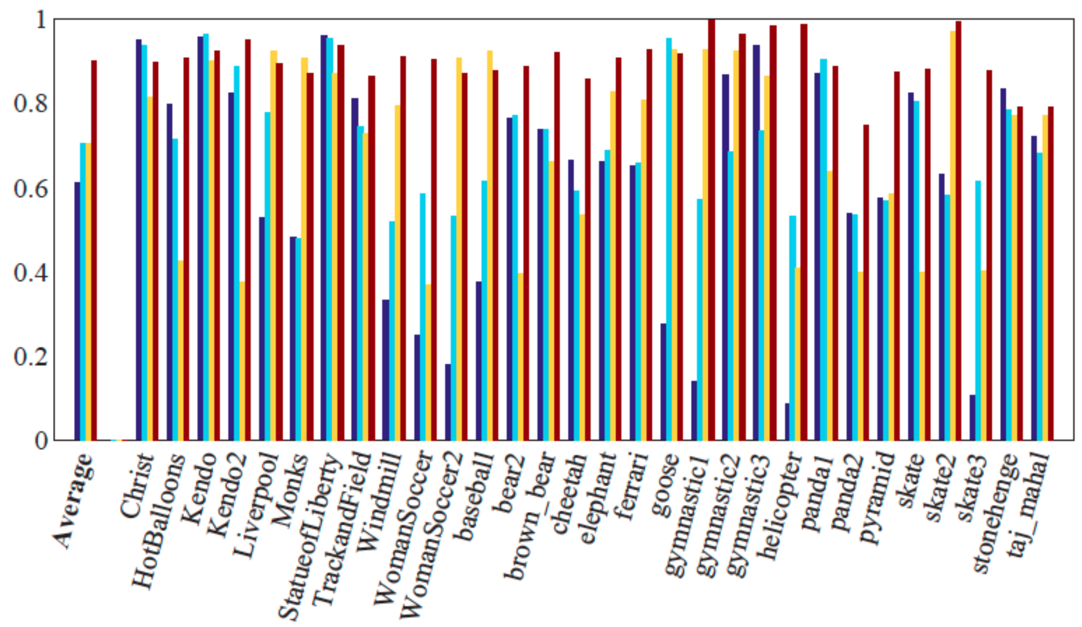
- Optimize the correspondences **W** and bin. masks **B.**
- Coordinate descent is used.
- Cost function is non-convex.

Results

- 4 algorithms are compared using co-seg. datasets;
  - MSRC (14 classes)
    - New precision 87.66%
    - Old precision 73.61% and 54.65%
  - iCoseg (30 classes)
    - New precision 89.84%
    - Old precision 70.21% and 70.41%



Precision
Average
bike
bird
car
cat
chair
cow
dog
face
flower
house
plane
sheep
sign
tree
Joulin et al. [7]
Joulin et al. [8]
Kim et al. [11]
Ours



Average
Christ
HotBalloons
Kendo
Kendo2
Liverpool
Monks
StatueofLiberty
TrackandField
Windmill
WomanSoccer
WomanSoccer2
baseball
bear2
brown_bear
cheetah
elephant
ferrari
goose
gymnastic1
gymnastic2
gymnastic3
helicopter
panda1
panda2
pyramid
skate
skate2
skate3
stonehenge
taj_mahal

Results

- Better than Vicente *et al.*'s state-of-the art object co-segmentation.

Method	MSRC		iCoseg	
	$\bar{P}$	$\bar{J}$	$\bar{P}$	$\bar{J}$
Vicente et al. [25]	90.2	70.6	85.34	62.04
Ours	**92.16**	**74.7**	**89.6**	**67.63**

Results

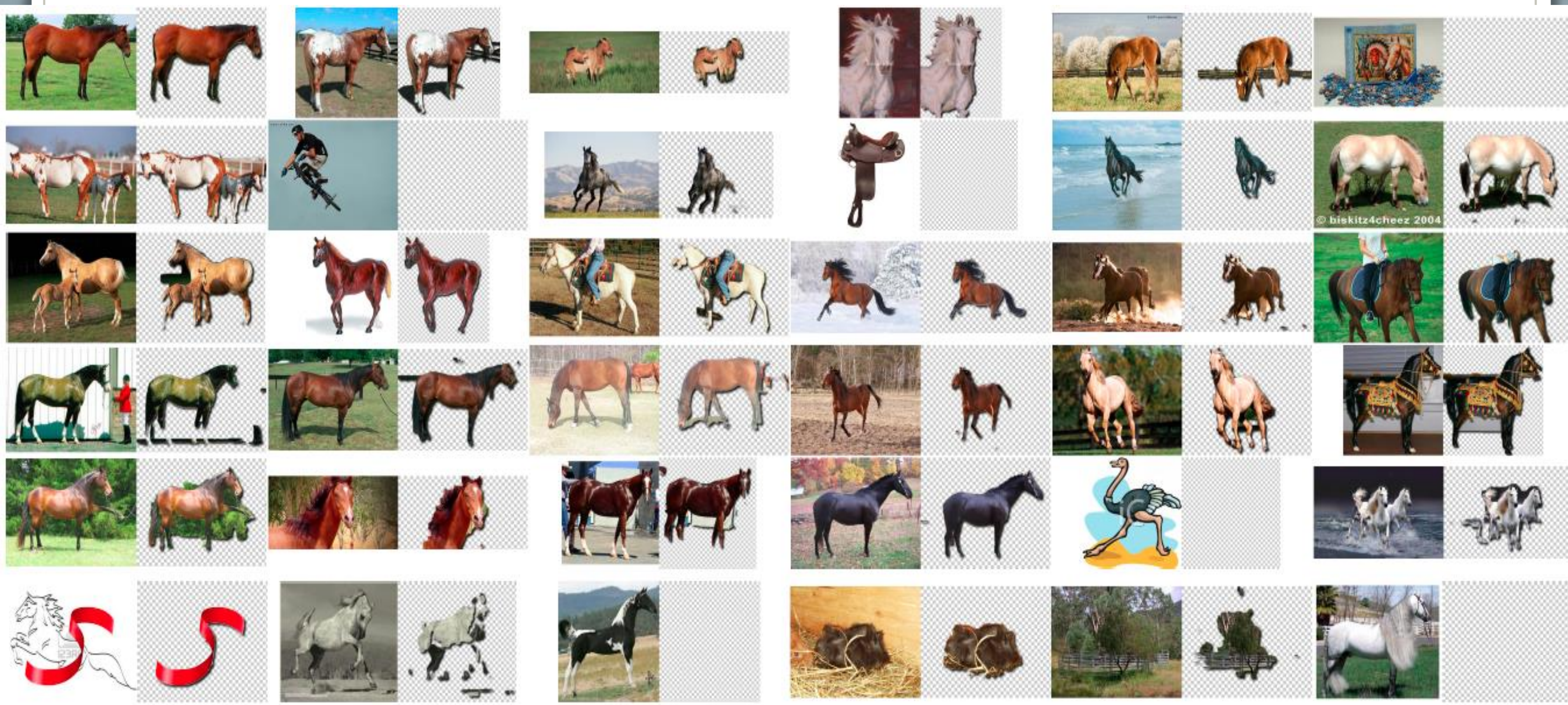
- On internet dataset
  - Car (~4K images), horse (~6K images), airplane (~4K images).
- LabelMe and Mechanical Turk
  - 1306 car, 879 horse, 561 airplane images are labeled.

Method	**Car** (7.5%)		**Horse** (7.8%)		**Airplane** (16%)	
	P	J	P	J	P	J
Without corr.	72.25	46.10	74.88	50.06	80.53	51.18
With corr.	**83.38**	**63.36**	**83.69**	**53.89**	**86.14**	**55.62**

Results

- Comparison with previous co-segmentation methods on the internet dataset.

Method	Car (11%)		Horse (7%)		Airplane (18%)	
	P	J	P	J	P	J
Baseline 1	68.91	0	81.54	0	87.48	0
Baseline 2	31.09	34.93	18.46	19.85	12.52	15.26
Joulin et al. [8]	58.7	37.15	63.84	30.16	49.25	15.36
Joulin et al. [9]	59.2	35.15	64.22	29.53	47.48	11.72
Kim et al. [12]	68.85	0.04	75.12	6.43	80.2	7.9
Ours	**85.38**	**64.42**	**82.81**	**51.65**	**88.04**	**55.81**


© biskitz4cheez 2004

Thank you!

- Questions?