MOMI-Cosegmentation: Simultaneous Segmentation of Multiple Objects among Multiple Images

Wen-Sheng Chu[1], Chia-Ping Chen[1,2] and Chu-Song Chen[1,3]
{wschu,cpchen,song}@iis.sinica.edu.tw

[1]Institute of Information Science, Academia Sinica
[2]Dept. of Computer Science and Information Engineering, National Taiwan University
[3]Graduate Institute of Networking and Multimedia, National Taiwan University
Taipei, Taiwan

ACCV2010

July 15, 2010

Co-segmentation

Segmenting **shared visual patterns** in an image set draws increasing interest and becomes more important for various applications.

Applications

- Image-Based Applications
  - clustering
  - indexing
  - summarizing
  - biomedical imaging
  - categorization
  - similarity measure
  - content-based image retrieval
  - ...

Outline

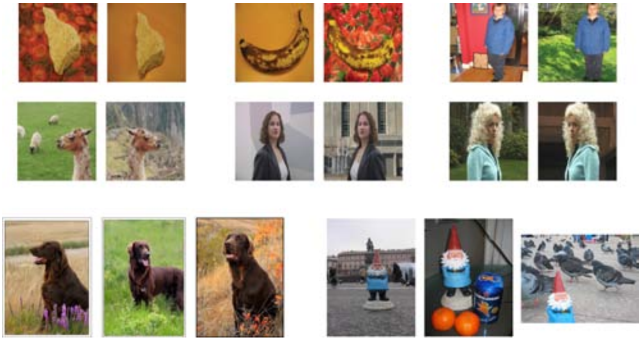
Outline

Related Works

- Cosegmentation
  - Simultaneously segment the same object from two (or more) images
  - Make segmentation of an object easier without user interaction

- Rother et al. (2006)
  C. Rother, T. Minka, A. Blake, and V. Kolmogorov. Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs. In *CVPR*, 2006
- Cao and Fei-Fei (2007)
  L. Cao and L. Fei-Fei. Spatially coherent latent topic model for concurrent object segmentation and classification. In *ICCV*, 2007
- Cheng and Figueiredo (2007)
  D. S. Cheng and M. Figueiredo. Cosegmentation for image sequences. In *International Conference on Image Analysis and Processing*, pages 635–640, 2007
- Gallagher and Chen (2008)
  A. C. Gallagher and T.-H. Chen. Clothing cosegmentation for recognizing people. In *CVPR*, pages 1–8, 2008
- Mukherjee et al. (2009)
  L. Mukherjee, V. Singh, and C. R. Dyer. Half-integrality based algorithms for cosegmentation of images. In *CVPR*, 2009
- Hochbaum and Singh (2009)
  D. S. Hochbaum and V. Singh. An efficient algorithm for Co-segmentation. In *ICCV*, 2009
- Batra et al. (2010)
  D. Batra, A. Kowdle, D. Parikh, J. Luo, and T.-H. Chen. icoseg: Interactive co-segmentation with intelligent scribble guidance. In *CVPR*, 2010
- Joulin et al. (2010)
  A. Joulin, F. Bach, and J. Ponce. Discriminative clustering for image co-segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010

Cosegmentation Assumptions



- Each image contains only one instance of the same object.
- Supply only *one* additional image to achieve completely automatic segmentation.
- Experimental images are somewhat unrealistic.

Daily Pictures

Images[1] may share more than one common object.

An object may appear more than one time in an image.

[1] The images were collect from Flickr and Google Image.

Outline

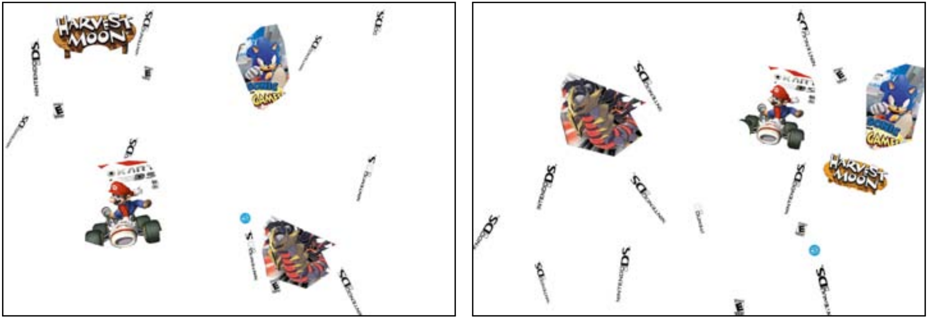
Outline

Common Pattern Discovery

What are the **common patterns** in the two images?

- Goal: unsupervisedly detect visual patterns that repeatedly appear in an image set.

Difficulties



Do you believe that “*finding common patterns is difficult even for humans*”?

- No prior knowledge is provided for the common patterns
  - What are the common patterns?
  - How many common patterns are there in a set of images?
  - How many times does each common pattern appear in an image?

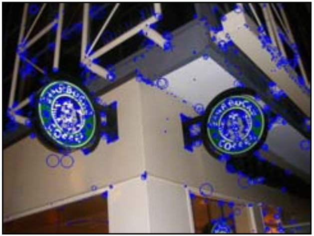
Review Common Pattern Discovery

- An intuitive way to find the common patterns is to exhaustively compare all sub-images at all possible positions and scales.
  - Search domain is extremely huge.
  - The computational cost increases exponentially with the number of input images.
- In this work, we used the common pattern discovery algorithm proposed by Chen et al. (2010)[2], which is mainly composed of four steps:
  1. Image representation & candidate matches
  2. Incompatibility matrix
  3. Correspondence graph
  4. Density-based clustering

[2] C.-P. Chen, W.-S. Chu, C.-S. Chen, and Y.-P. Hung. Common pattern discovery with high-order constraints by density-based cluster discovery. *Submitted to IEEE Transaction on Systems, Man, and Cybernetics, Part B*, 2010

① Image Representation

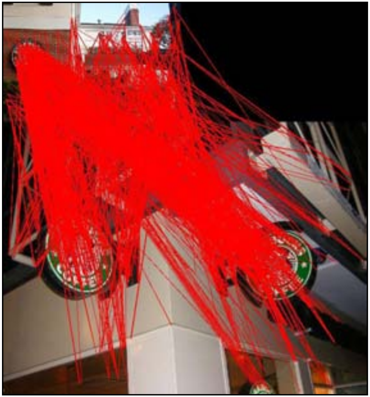
- Given an image I_n, $n = 1, \ldots, N$, a set of local appearance features is extracted as $\mathcal{F}_n = \left\{(\mathbf{p}_n^i, s_n^i, \mathbf{d}_n^i) | i = 1, \ldots, |\mathcal{F}_n|\right\}$, where $\mathbf{p}_n^i$ and s_n^i are the position and the scale of I_n.
- Here, we use the OpponentSIFT descriptor[3] to extract features.



[3] K. van de Sande, T. Gevers, and C. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Trans. on PAMI*, (in press), 2010

① Candidate Matches (cont.)

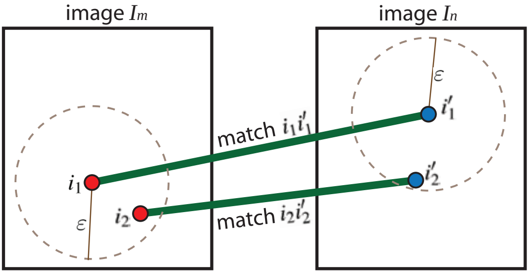
- Then, given two sets of local features of images I_m and I_n, we filter out the candidate set $\mathcal{M} = \{ii' | \ \|\mathbf{d}_m^i - \mathbf{d}_n^{i'}\| < \lambda\}$ from all possible correspondences across each pair of local features.



Candidate matches

Correct matches

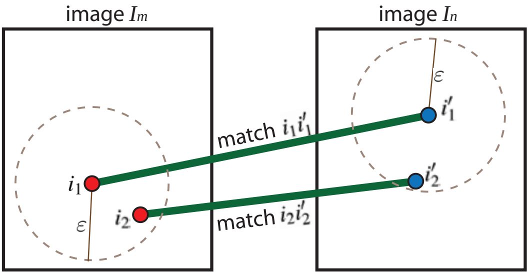
② Incompatibility Matrix



- Consider those candidate matches lying within the spatial ε-neighborhood, i.e., $sd^m(i_1, i_2) < \varepsilon$ and $sd^n(i'_1, i'_2) < \varepsilon$.
- An incompatibility matrix $\mathbf{D}$ is constructed to represent the incoherence between a pair of candidate matches.

$$\mathbf{D}(i_1 i'_1, i_2 i'_2) = \alpha_1 \times unary(i_1 i'_1, i_2 i'_2) + \alpha_2 \times binary(i_1 i'_1, i_2 i'_2),$$

② Incompatibility Matrix (cont.)



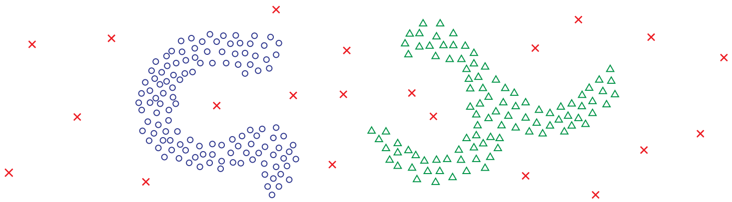
- $unary(i_1 i'_1, i_2 i'_2)$ and $binary(i_1 i'_1, i_2 i'_2)$ capture the **appearance dissimilarity** and **geometric inconsistency** for each pair of candidate matches $i_2 i'_2$ and $i_1 i'_1$:

$$unary(i_1 i'_1, i_2 i'_2) = \frac{\|\mathbf{d}_m^{i_1} - \mathbf{d}_n^{i'_1}\| + \|\mathbf{d}_m^{i_2} - \mathbf{d}_n^{i'_2}\|}{2},$$

$$binary(i_1 i'_1, i_2 i'_2) = \frac{|sd^m(i_1, i_2) - sd^n(i'_1, i'_2)|}{\sqrt{sd^m(i_1, i_2) sd^n(i'_1, i'_2)}}.$$

③ Correspondence Graph

- Small values in $\mathbf{D}$ reflects potential correct matches of a shared object in the image pair, because appearance difference and geometric inconsistency between correct matches shall be small.
- Incorrect matches tend to be inconsistent with each other with large incompatibilities.
- We can see the candidate matches $\mathcal{M}$ as the nodes that forms the **correspondence graph** with corresponding linkage weights specified by $\mathbf{D}$.



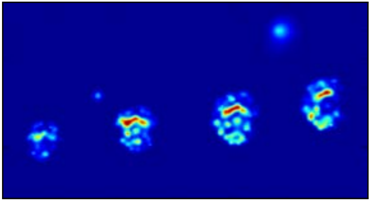
④ Density-Based Clustering

- Given the correspondence graph, the problem of finding common patterns in an image set is reduced to a dense cluster discovery problem.
- We do not know in advance the cluster shapes.
  - Clustering methods that assume each cluster has a globular shape is not suitable, such as K-means or affinity propagation.
- The number of dense clusters in the correspondence graph is also unknown.
- We utilize the density-based algorithm[4] to discover clusters with arbitrary shapes in the presence of a large number of outlier matches.

[4] M. Ester, H. P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Knowledge Discovery and Datamining*, pages 226–231, 1996

Confidence Map

- We derive $(N-1)$ feature masks for each image in the unannotated image set.
- Each feature mask records the confidence of each local feature and indicates that how likely a local feature is a part of a common pattern.
- A confidence map for each image can be obtained by fusing the $(N-1)$ feature masks.



Preliminary Results

- Preliminary results can be obtained by performing a simple thresholding on confidence maps.

Outline

MOMI-Cosegmentation with Common Pattern Discovery

- Conventional cosegmentation methods are restrictive in two assumptions
  - The input is an image pair.
  - Each image contains the same object in different backgrounds.
- Goal: detect *multiple objects* that may appear *multiple times* in one image.
- We considering the cosegmentation problem as an individual foreground/background segmentation by incorporating the **confidence maps** and **preliminary segmentation results**.

Formulation

- The segmentation problem of image I_n is interpreted as a binary labelling problem by minimizing the following cost function:

$$E(X) = \lambda_{\text{color}} \sum_{p \in \mathcal{V}} E_{\text{color}}(x_p) + \lambda_{\text{smoothness}} \sum_{(p,q) \in \mathcal{E}} E_{\text{smoothness}}(x_p, x_q) +$$
$$\lambda_{\text{confidence}} \sum_{p \in \mathcal{V}} E_{\text{confidence}}(x_p) + \lambda_{\text{locality}} \sum_{p \in \mathcal{V}} E_{\text{locality}}(x_p).$$

  where $\mathcal{E}$ is the set of all adjacent pixels pairs in I_n, $\mathcal{V}$ is the set of all pixels in I_n and $X = \{x_p | p \in \mathcal{V}\}$ is the set of labels.
- The parameters λ_{color}, $\lambda_{\text{smoothness}}$, $\lambda_{\text{confidence}}$ and $\lambda_{\text{locality}}$ balance the contribution of each cost term.

Color Term & Smoothness Term

- We explain the color and smoothness terms as the **fundamental model**, which are frequently used in segmentation problems[5,6,7].
- **Color terms** use the fact that different groups of fg/bg segments tend to follow different color distributions.

$$E_{\text{color}}(x_p) = -\log G(p|x_p),$$

$$G(p|x_p) = \sum_{k=1}^{K} \pi_k \frac{1}{\sqrt{\det \Sigma_k}} \exp\left(-\frac{1}{2}(p-\mu_k)^T \Sigma_k^{-1} (p-\mu_k)\right),$$

  where $G(p|x_p)$ is the Gaussian mixture model indicating the probability that pixel p belongs to the label x_p.
- The color term encourages the pixels to follow the labels of the most similar color model.

---

[5] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM SIGGRAPH*, page 314. ACM, 2004

[6] J. Sun, W. Zhang, X. Tang, and H.-Y. Shum. Background cut. *ECCV*, 3952:628–641, 2006

[7] J. Y. Guillemaut, J. Kilner, and A. Hilton. Robust graphcut scene segmentation and reconstruction for free-viewpoint video of complex dynamic scenes. In *ICCV*, 2009

Color Term & Smoothness Term (cont.)

- **Smoothness terms** preserve the coherence between two neighboring pixels of similar pixel values and imply a tendency to solidity of objects:

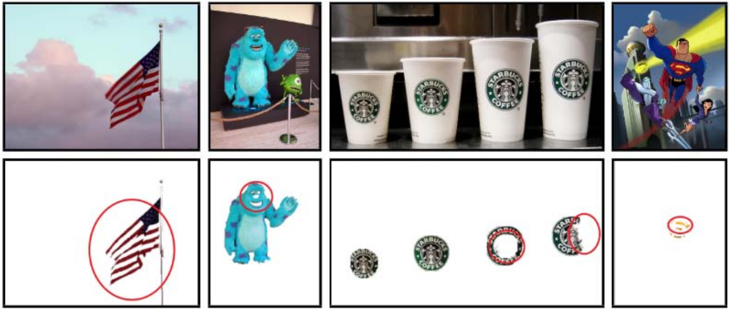
$$E_{\text{smoothness}}(x_p, x_q) = [x_p \neq x_q] \exp\left(-\beta \|p - q\|^2\right),$$

  where $[expr]$ denotes the indicator function taking value $0, 1$ for the predicate $expr$.

- The minimization problem using only E_{color} and $E_{\text{smoothness}}$ is similar to the GrabCut method proposed by Rother et al. (2004).

Segmentation using the Fundamental Model

- Similar colors between the foreground and the background models distract the labelling of foreground pixels.



- The segmentation domain is expanded from the initial user-defined rectangle trimap to the entire image.
- We extend the fundamental model by the confidence term $E_{\text{confidence}}$ and the locality term E_{locality} to recover correct foreground pixels as well as remove false background artifacts.

Confidence Term

- Given $c(p)$ as the original value of a confidence map, the confidence term is defined as

$$E_{\text{confidence}}(x_p) = \begin{cases} (2x_p - 1)\tilde{c}(p), & \tilde{c}(p) > 0 \\ (1 - 2x_p)\tilde{c}(p), & otherwise \end{cases}$$

  where $\tilde{c}(p)$ is the normalized confidence cost of pixel p in $[-1, 1]$ by the sigmoid function:

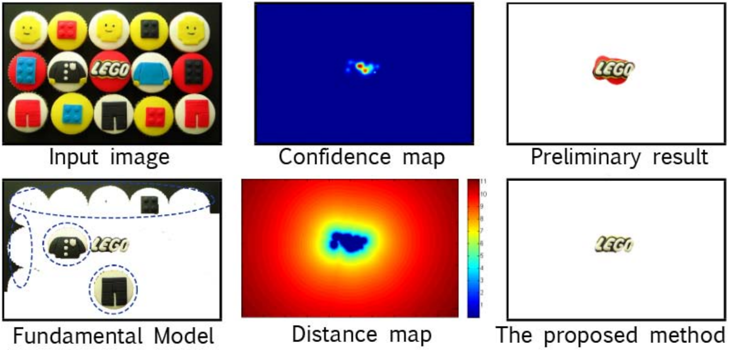
$$\tilde{c}(p) = 4\left(\frac{1}{1 + \exp\left(-c(p)\right)} - \frac{3}{4}\right).$$

- When $\tilde{c}(p) > 0$, p has high possibility of belonging to the foreground, and thus the confidence term encourages the foreground $(x_p = 1)$ likelihood by adding $\tilde{c}(p)$ and penalizes the background $(x_p = 0)$ by subtracting $\tilde{c}(p)$.
- On the other hand, when $\tilde{c}(p) \leq 0$, we subtract $\tilde{c}(p)$ from $x_p = 1$ and add $\tilde{c}(p)$ to $x_p = 0$.

Confidence Term (cont.)

- Most neglected foreground pixels could be recovered by incorporating the confidence term.

Locality Term



- Background artifacts (blue dashed circles) occur when they have similar colors as foreground pixels.

Locality Term (cont.)

- The further a pixel p is away from a reference pixel q, the less possible p belongs to the foreground.
- To remove background clutters, we impose the distance penalty on pixels away from those with high enough confidence values:

$$E_{\textsf{locality}}(x_p) = \log\left(\max_{q\in\mathcal{V},c(q)>\delta} dist(p,q)\right),$$

where $dist(p,q) = \|\mathbf{p}_p - \mathbf{p}_q\|^2$ is the spatial distance between pixel pairs (p,q), δ controls the threshold for candidates of pixel q.

Outline

Experimental Results

- We used min-cut algorithm[8] to minimize the cost function $E(X)$.
- $\epsilon = 2000$ and $d = 20$ are fixed across all the experiments for the density-based clustering algorithm.
- $K = 5$ is set for the fg/bg color models.
- $\lambda_{\text{color}} = 1$ and $\lambda_{\text{smoothness}} = 40$ are set for the proposed cost function, while $\lambda_{\text{confidence}}$ and $\lambda_{\text{locality}}$ are user-specified.
- Qualitative and quantitative analysis of the proposed method are evaluated on 12 image sets collected from Flickr with moderate variations in illumination and scale.

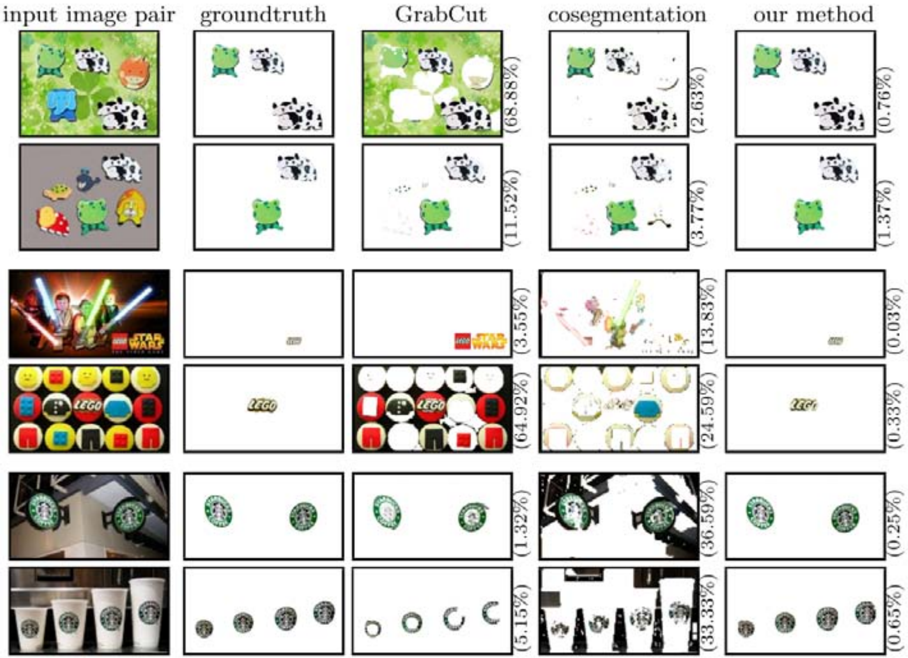
---

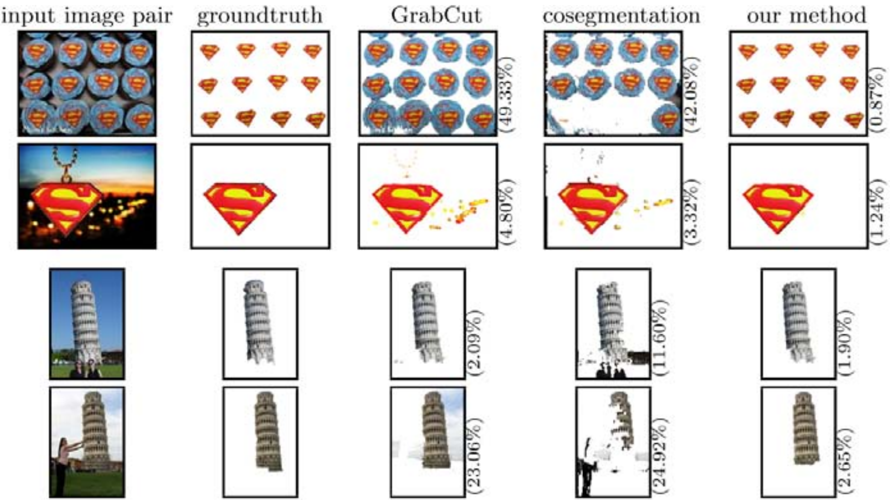
[8] Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in ND images. In *ICCV*, volume 1, pages 105–112, 2001

Comparison with Cosegmentation

- We compare MOMI-cosegmentation with the state-of-art cosegmentation proposed by Hochbaum and Singh (2009)[9].
- Some shortcomings of the implementation of Hochbaum and Singh (2009):
  1. considers only two input images
  2. takes a large memory storage of additional nodes, i.e., segmentation errors were reported for lower-resolution images
  3. requires manually labelling of RGB intensities for foreground and background
- In contrast, the proposed method:
  1. considers a small image set
  2. handles full-resolution images
  3. preliminary labelling is completely automatic

[9] D. S. Hochbaum and V. Singh. An efficient algorithm for Co-segmentation. In *ICCV*, 2009


input image pair
groundtruth
GrabCut
cosegmentation
our method
(68.88%)
(2.63%)
(0.76%)
(11.52%)
(3.77%)
(1.37%)
(3.55%)
(13.83%)
(0.03%)
(64.92%)
(24.59%)
(0.33%)
(1.32%)
(36.59%)
(0.25%)
(5.15%)
(33.33%)
(0.65%)


input image pair
groundtruth
GrabCut
cosegmentation
our method
(49.33%)
(42.08%)
(0.87%)
(4.80%)
(3.32%)
(1.24%)
(2.09%)
(11.60%)
(1.90%)
(23.06%)
(24.92%)
(2.65%)

Comparison with the Fundamental Model

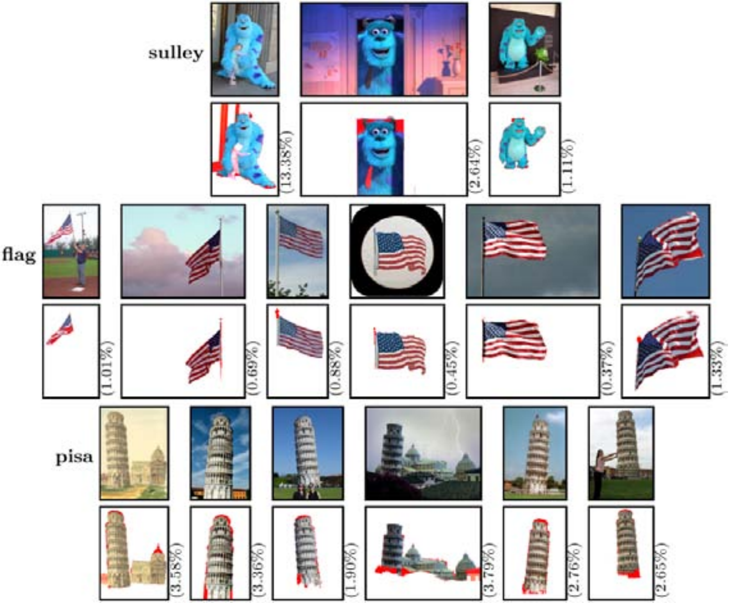
- We experimented MOMI-cosegmentation ($\mathtt{MOMI-CS}$) on more than two input images and compared the performance with the fundamental model ($\mathtt{FM}$) used by Rother et al. (2004)[10].

set(#img)	sulley(3)	starbucks(3)	magnet(4)	flag(6)	pisa(6)	superman(7)
FM	20.50	2.68	22.56	7.71	17.88	18.62
MOMI-CS	5.71	0.41	1.20	0.79	3.01	1.38
set(#img)	domino(6)	heineken(8)	warcraft(6)	kfc(6)	lego(4)	pringles(8)
FM	26.65	18.47	26.65	35.21	43.52	15.24
MOMI-CS	2.46	1.25	2.63	6.78	1.08	4.17

- The proposed approach achieved an average of 2.57$\%$ segmentation errors across the 12 image sets.
  - some objects of the same class appear in heterogeneous circumstances
  - most images contain cluttered backgrounds
  - each image has mega-pixel resolution

[10] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM SIGGRAPH*, page 314. ACM, 2004

Some Results on Deformable Objects



Outline

Conclusion

- We have proposed a new cosegmentation approach called MOMI-cosegmentation, which is more general and scalable.
- The proposed can deal with more than two input images, and allow multiple objects to appear more than one time in an image.
- We incorporated a common pattern discovery algorithm with color, smoothness, confidence and locality cues to achieve satisfactory segmentation.
- Label initialization and segmentation process are completely automatic in the proposed methods.
- Experiments have demonstrated that the performance of the proposed method outperforms the state-of-art cosegmentation method.

References

D. Batra, A. Kowdle, D. Parikh, J. Luo, and T.-H. Chen. icoseg: Interactive co-segmentation with intelligent scribble guidance. In *CVPR*, 2010.

Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in ND images. In *ICCV*, volume 1, pages 105–112, 2001.

L. Cao and L. Fei-Fei. Spatially coherent latent topic model for concurrent object segmentation and classification. In *ICCV*, 2007.

C.-P. Chen, W.-S. Chu, C.-S. Chen, and Y.-P. Hung. Common pattern discovery with high-order constraints by density-based cluster discovery. *Submitted to IEEE Transaction on Systems, Man, and Cybernetics, Part B*, 2010.

D. S. Cheng and M. Figueiredo. Cosegmentation for image sequences. In *International Conference on Image Analysis and Processing*, pages 635–640, 2007.

M. Ester, H. P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Knowledge Discovery and Datamining*, pages 226–231, 1996.

A. C. Gallagher and T.-H. Chen. Clothing cosegmentation for recognizing people. In *CVPR*, pages 1–8, 2008.

J. Y. Guillemaut, J. Kilner, and A. Hilton. Robust graphcut scene segmentation and reconstruction for free-viewpoint video of complex dynamic scenes. In *ICCV*, 2009.

D. S. Hochbaum and V. Singh. An efficient algorithm for Co-segmentation. In *ICCV*, 2009.

A. Joulin, F. Bach, and J. Ponce. Discriminative clustering for image co-segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

L. Mukherjee, V. Singh, and C. R. Dyer. Half-integrality based algorithms for cosegmentation of images. In *CVPR*, 2009.

C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM SIGGRAPH*, page 314. ACM, 2004.

C. Rother, T. Minka, A. Blake, and V. Kolmogorov. Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs. In *CVPR*, 2006.

J. Sun, W. Zhang, X. Tang, and H.-Y. Shum. Background cut. *ECCV*, 3952:628–641, 2006.

K. van de Sande, T. Gevers, and C. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Trans. on PAMI*, (in press), 2010.

Thank you for your attendance!

Questions?