Object Co-detection

Sid Yingze Bao Yu Xiang Silvio Savarese
University of Michigan at Ann Arbor, USA

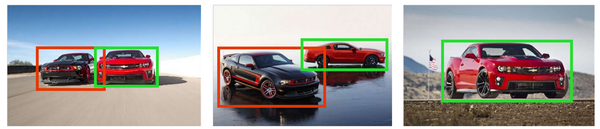
MICHIGAN VISION LAB

12th European Conference on Computer Vision

Object Co-detection Problem

Goals

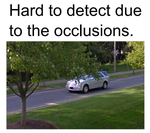
(1) Detect objects in all images.
(2) Recognize same object instance in different images (matching objects).
(3) Estimate the viewpoint transformation between matching objects.



Motivations

(1) Better detection accuracy than single-image methods

Hard to detect due to the occlusions.



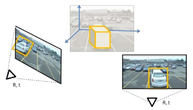
Transfering information across images makes detection easier.

Pose variation makes low-level feature matching unreliable.

(2) Better matching accuracy than methods based on low-level features.

(3) Helpful in solving other problems.

Tracking by detection



Semantic Structure From Motion [1]

Challenges

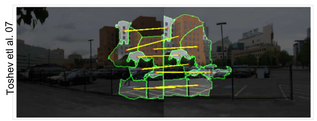
(1) Object appearance variations (pose changes and self-occlusions).
(2) Input images may contain different backgrounds.

Related Problems

Single image object detection

Co-segmentation

Single instance detection



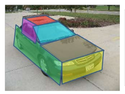
Region Matching

Our Framework

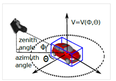
Represent an object by a part-based model.

Object $O = \{r, V, p_1, p_2, ..., p_n\}$

r : root filter



p : parts

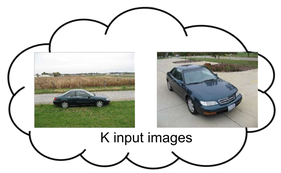


V : view points

Energy Formulation

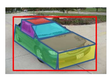
Object detection are obtained by maximizing:

$$E(\mathcal{O}, \mathcal{I}) = \sum_{k=1}^{K} E_{\text{unit}}(O^k, I^k) + \sum_{i=1}^{n} E_{\text{match}}(\{p_i^k\}_{k=1}^{K}, \{V^k\}_{k=1}^{K}, \mathcal{I})$$



Unitary potential E_{unit}
- Leverage part-based object detectors
- E.g. Xiang & Savarese [2], Felzenszwalb et al. [3].

$$E_{\text{unit}}(O^k, I^k) = E_{\text{root}}(r^k, V^k, I^k) + \sum_{i=1}^{n} E_{\text{part}}(p_i^k, V^k, I^k) + \sum_{i=1}^{n} E_{\text{rp}}(r^k, p_i^k, V^k, I^k)$$



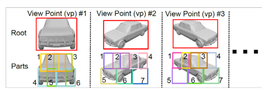
Matching potential E_{match}
- Rectifying parts based on estimated poses
- Matching by using combination of features

Advantages of the Co-detection Framework

(1) Handle object pose variations and self-occlusions.
(2) Reject false alarms produced by a single-image detector

Using 2D Object Representation

- Treat different poses as different categories.
  - Cannot match objects with large pose variation.
- More choices to compute unitary potential.
  - e.g. Leibe et al. 04, Felzenszwalb et al. 09, Bourdev et al. 09, Yang & Ramanan 11



References

[1] Y. Bao, S. Savarese, Semantic Structure From Motion, CVPR 2011.
[2] Y. Xiang, S. Savarese. Estimating the aspect layout of object categories. CVPR 2012.
[3] P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan. Object detection with discriminatively trained part-based models. PAMI 2010.
[4] D. Lowe. Object recognition from local scale-invariant features. ICCV 1999.
[5] S. Lazebnik, C. Schmid, J. Ponce. Spatial pyramid matching for recognizing natural scene categories. CVPR 2006
[6] A. Ess, B. Leibe, L. van Gool, Depth and Appearance for Mobile Scene Analysis, ICCV 2007.
[7] S. Savarese, L. Fei-Fei, 3d Generic Object Categorization, Localization and Pose Estimation, ICCV 2007

Acknowledgement

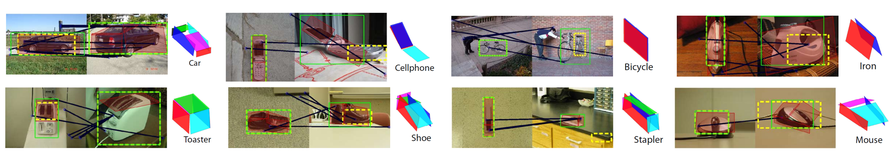
We acknowledge the support of NSF CAREER #1054127 and NSF CPS grant #0931474.

Experiments

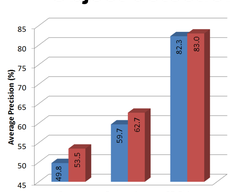
Car dataset (300 image pairs) [1]

Pedestrian dataset (200 image pairs) [6]

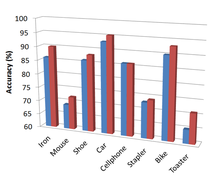
3D Object dataset (500 image pairs for each category) [7]



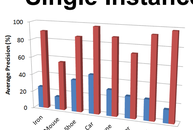
Object detection Accuracy



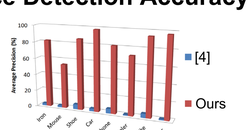
Pose Estimation Accuracy



Single Instance Detection Accuracy

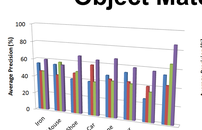


same pose

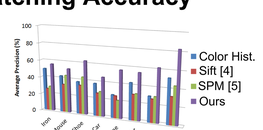


different poses

Object Matching Accuracy



same pose



different poses