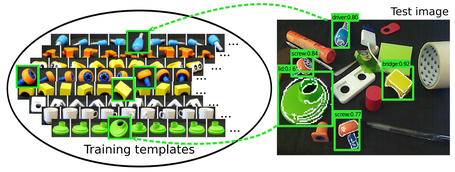
Fast Detection of Multiple Textureless 3-D Objects

Hongping Cai, Tomas Werner and Jiri Matas

Center for Machine Perception, Czech Technical University, Prague

Introduction

- **Task:** Detection+Recognition +Pose Estimation of Textureless objects

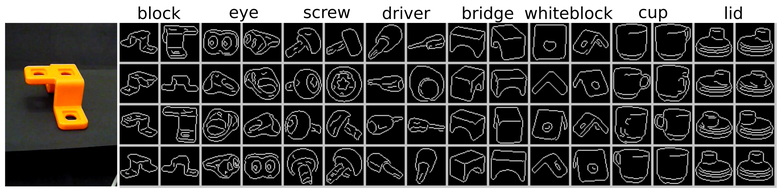


- We propose a fast edge-based approach for detection and approximate pose estimation of multiple textureless objects in a single image, by combining an efficient voting procedure with an effective verification stage.

Challenges

- **Texture-free objects:**
  - Appearing in both industrial manipulation and daily life.
  - Approaches with local detectors and descriptors fail. We use edges.
- **3-D pose:**
  - Large shape variety from different viewpoints.
- **Real-time requirement:**
  - A key challenge due to a large set of templates.

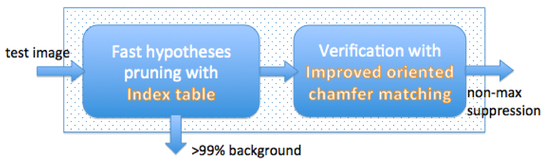
Acquiring Training Templates



Acquiring training set for our CMP-8objs dataset. Left: each object is placed on a turntable and captured by video camera, covering a viewing hemisphere. Right: examples of 12,960 training edge-based templates.

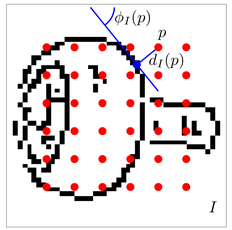
Our Approach

Two-stage cascade



Stage1: Fast Hypotheses Pruning

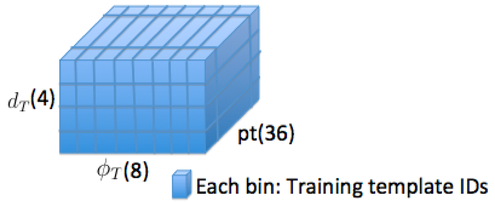
A novel sparse edge-based image descriptor



- m reference points (in red) are assigned two features: $(d_I(p_1), \ldots, d_I(p_m), \phi_I(p_1), \ldots, \phi_I(p_m))$.
- $d_I(p)$: distance to the nearest edge.
- $\phi_I(p)$: orientation of the nearest edge.

Fast voting with quantized features

- Quantize the descriptors
- Put all training template indices into a table
- Fast voting scheme
- Speed up by grouping triplets of reference points



Stage2: Improved Oriented Chamfer Matching (OCM)

Compensating the bias towards simpler shapes

- Our CompOCM score:

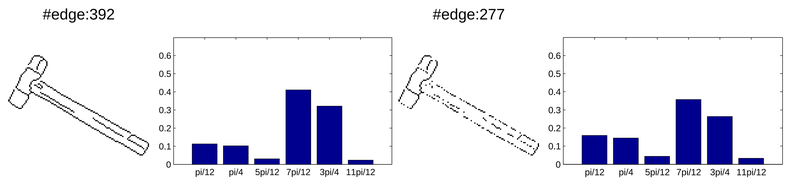
$$s_\lambda(T, I) = \frac{\left|\left\{\, e \in T \;\middle|\; d_I(e) \le \theta_d,\ |\phi_T(e) - \phi_I(e)|_\pi \le \theta_\phi \,\right\}\right|}{\lambda|T| + (1-\lambda)\overline{|T|}}, \qquad (1)$$

$\overline{|T|} = \frac{1}{n}\sum_{i=1}^{n}|T_i|$: the average number of edges over all training templates, $\lambda \in [0, 1]$.

- Standard OCM score: $s_1(T, I)$

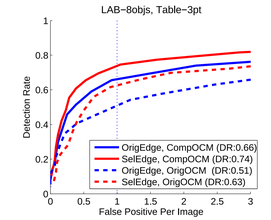
Selecting discriminative edges of templates

- Selection by stability to viewpoint.
  - Only the edges with higher repeatability after slight changes of viewpoint are kept.
- Selection by edge orientations.
  - 40% of edges corresponding to the two highest orientation bins are randomly removed.



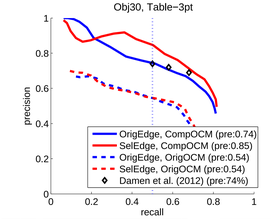
Left: original edges and its orientation histogram,
Right: selected edges and its orientation histogram.

Results



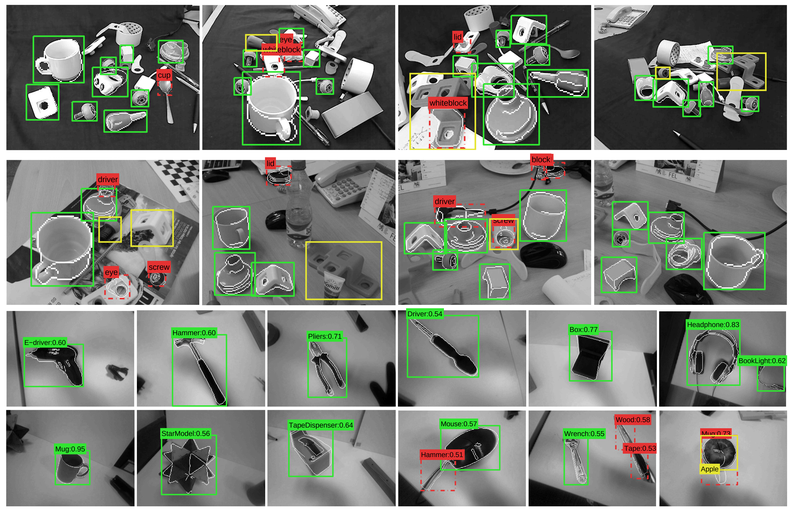
CMP-8objs*

- 8 objects
- 12,960 training templates
- 60 test images of size 640×480
- DR=0.74@FPPI=1.0
- 0.65 s/frame



Obj30 [Damen et al. (2012)]

- 30 objects
- 7,056 training templates
- 1,200 test images of size 640×480
- Prec=0.85@recall=0.5
- 0.26 s/frame



Example detections for CMP-8objs (top) and Obj30 (bottom) datasets. TP, FP and FN are shown in green, red and yellow, respectively.

Summary

- Fast detection of textureless 3D objects, better than state-of-the-art accuracy.
- Speed: 4fps on Obj30 and 1.5fps on CMP-8objs dataset.
- Purely edge-based method, no other clues, e.g., color, used.
- A new dataset CMP-8objs is generated for evaluation of 3-D textureless object detection and pose estimation.*

* http://cmp.felk.cvut.cz/data/textureless

Email: caihongp@cmp.felk.cvut.cz

Homepage: http://cmp.felk.cvut.cz/~caihongp/