Carried Object Detection based on an Ensemble of Contour Exemplars

Farnoosh Ghadiri[1], Robert Bergevin[1], Guillaume-Alexandre Bilodeau[2]

[1]LVSN-REPARTI, Université Laval, [2]LITIV lab., Polytechnique Montréal

UNIVERSITÉ LAVAL

POLYTECHNIQUE MONTRÉAL WORLD-CLASS ENGINEERING

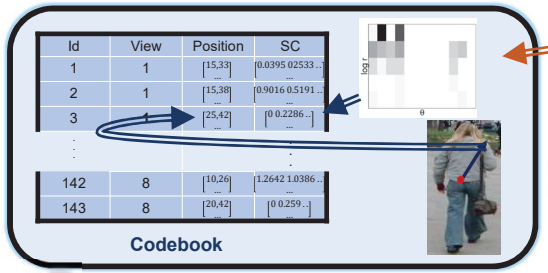
ECCV'16 EUROPEAN CONFERENCE ON COMPUTER VISION

https://sites.google.com/site/cosdetector/home

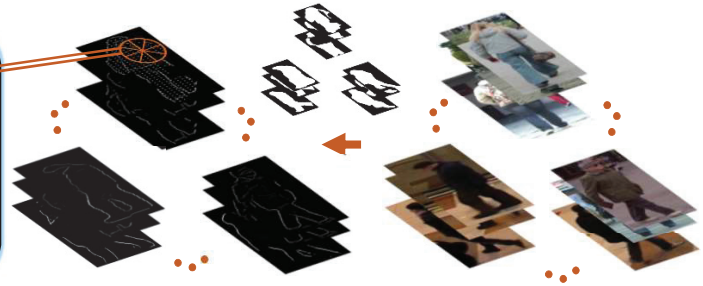
CHALLENGE

People can carry **a variety of objects** such as handbag, a musical instrument, or even an unusual /dangerous item like an improvised explosive device.

Can we automatically detect all kind of objects carried by people without defining specific model for them ?



Codebook



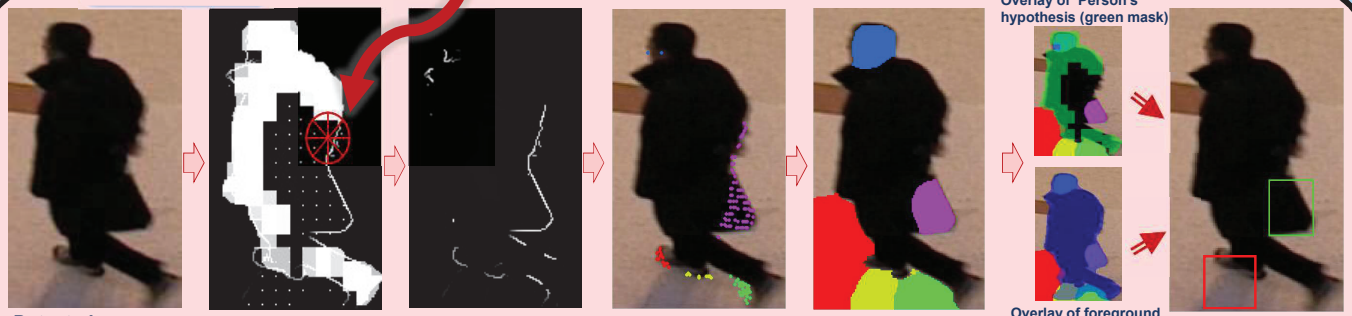
Ensemble of Contour Exemplars (ECE)

Exemplars in 8 viewing directions and various poses

Learning an ECE

- Collecting exemplars in different viewing directions and various poses.
- Extracting Shape Contexts as local image features.
- Building a codebook of local features from exemplars together with their foreground mask and their viewing direction information.

Method



Detected pedestrian | Person's contour hypothesis mask | Candidate carried object (CO) contours | Generated seed points for candidate CO's contours | Assigned regions to seed points | Detected CO

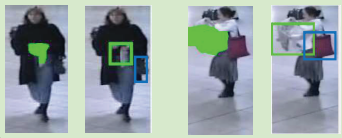
Carried Object Detection

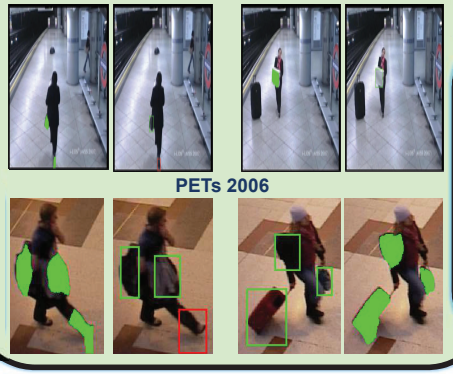
- Extracting edge map from the moving object and scale the image.
- **Person's view Classification**: Each local image feature is compared with a codebook entry at the same relative position. If a match is found , the corresponding codebook entry will cast a vote to the class, which it belongs to. The class for which maximum number of matching features is found, is selected as person's view class.
- **Hypothesis generation**: A hypothesis mask of a person's contour is built by backtracking the matching results of the person's view class.
- **Assigning a region to a probable carried object contours:** Contours that are less probable to be a person's contour are selected and their endpoints are connected to each other to create a closed regions. Then some pixels from each closed region are used as an input for biased normalized cut to generate a region.
- **Non Maximal Suppression** : A probability is assigned to each region based on its overlap with the foreground mask and the person's hypothesis mask. Then, between two overlapped regions, the one with the highest probability of being carried object is chosen.

CONTRIBUTIONS

- Generating a person's contour hypothesis combined with low-level information cues to detect COs.
- Analysing irregularity of a person's contours instead of human silhouettes.
- No prior knowledge of CO shape, location, and motion is assumed

Success & Failure





FN
FP
TP

Results

Comparison of Damen *et al.* & Tavanai *et al.* with the proposed method over PETS 2006.

	Prec.	Rec.	TP	FP	FN	F1 Score
Proposed Method (ECE)	57%	71%	59	44	24	63%
Damen *et al.*	50%	55%	46	45	37	52%
Tavanai et al.	-	-	-	-	-	~53%

Comparison of Damen *et al.* method with the proposed method over i-Lids AVSS

	Prec.	Rec.	TP	FP	FN
Proposed Method (ECE)	67%	60%	41	25	27
Damen *et al.*	52%	47%	32	45	36

PETS 2006

- Bounding box ground truth
- 83 carried objects
- 75 people

i-Lids AVSS

- Bounding box ground truth
- 68 carried objects
- 59 people

Reasons behind FPs & FNs

Poor person's hypothesis

- Due to the inherent variability of pose and cloth appearances.
- Wrongly detected person's view

Poor extracted foreground

Conclusion

- Our experiments indicate that learning human model from human's contour makes the system more robust to the factors that may give rise to irregularities such as clothing, than methods that model humans based on silhouettes.
- Using biased normalized cut to segment object combined with the high-level information of human model, provides us with a rough estimation of the carried object shape.