Improving Automatic Pedestrian Detection by means of Human Perception

Quoc C. Vuong[1] & Modesto Castrillón-Santana[2]

[1]MPI, Tübingen; [2]Universidad de Las Palmas de Gran Canaria, Spain

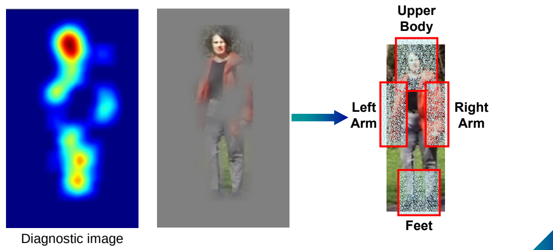
Problem

Automatic detection systems do not perform as well as human observers, even on simple detection tasks. A potential solution to this problem is training vision systems on appropriate regions of interests (ROIs), in contrast to training on predefined regions. Here we focus on detecting pedestrians in static scenes.

Can automatic vision systems for pedestrian detection be improved by training them on *perceptually-defined* ROIs?

Perceptual ROIs

[Bubbles [1] technique]



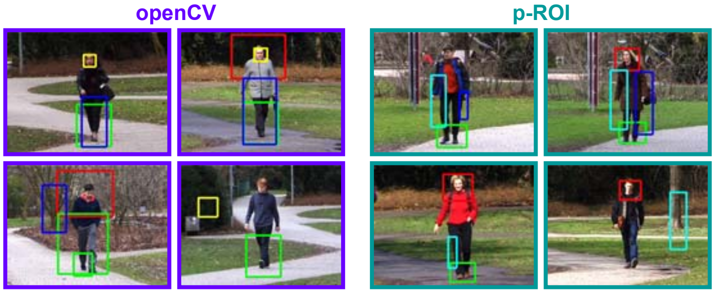
Diagnostic image

Results I: Comparison of default detectors and p-ROI detectors

	Total Detections	Double Detections	True Detections	False Detections	d'
default openCV Detectors [773/1488]					
Face	248	1	170	77	0.42
Upper Body	136	0	134	2	1.66
Lower Body	711	132	505	74	1.23
Full Body	213	3	69	41	0.24
			14.5%	**3.5%**	**0.89**
p-ROI Detectors [1301/1488]					
Upper Body	1166	56	984	126	1.79
Feet	909	89	636	184	0.97
Left Arm	616	81	407	128	0.76
Right Arm	601	70	353	178	0.46
			39.2%	**10.2%**	**1.00**

Results II: Sample detections

openCV **p-ROI**

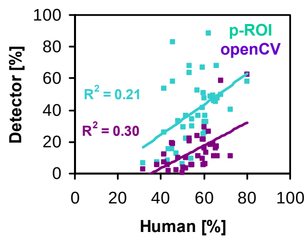


Viola and Jones Classifier

Framework: The Viola-Jones [2] general object detection framework achieves fast and robust performance by means of a cascade of weak classifiers trained using boosting. For example, given a 20-stage cascade of weak classifiers designed to reject 50% of non-object patterns (false positive rate) while accepting 99.9% of object patterns (detection rate) at each stage, the overall detection rate will be 0.999^{20} with a false positive rate of 0.5^{20}.

Training: Both detectors trained using openCV. The training set consisted of the CBCL pedestrian images [3] augmented with additional positive and negative samples of upper bodies and faces for the default openCV detectors and (a different set of) upper bodies for the p-ROI detectors.

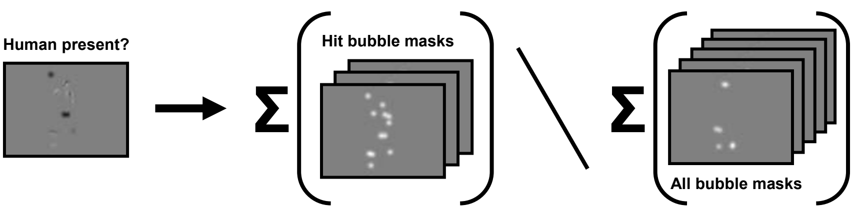
Testing: Stimuli from Vuong, Hof, Bülthoff, and Thornton [4] were used as test images for both detectors. These stimuli consisted of image sequences of 41 pedestrians walking through the park. The graph plots the correlation of detector hits with human performance (hits) for each pedestrian stimulus (dots).



Bubbles technique [1] to derive p-ROIs

Eight subjects were shown stimuli masked by Gaussian bubbles & had to judge if a human was present. Half the trials contained a human.

For each subject, bubble masks leading to hits (correct present trials) were summed and normalized by all bubble masks shown. The result is a *diagnostic image*. These normalized images were then averaged across subjects (see above).



Conclusions

Using a different set of features defined perceptually improved hits for automatic detection of pedestrian. The method provides a mean to select training features to improve automatic vision systems.

References

[1] Gosselin, F. and Schyns, P.G. (2001): *Vis. Res.* 41,2261-2271.

[2] Viola, P. and Jones, M. (2001): In *IEEE Conf. on Comp. Vis. and Pat. Recog*.

[3] http://cbcl.mit.edu/cbcl/software-datasets/]

[4] Vuong, Q. C., Hof, A., Bülthoff, H. H., and Thornton, I. M. (2004): *Vision Science Society Annual Meeting*.

Thanks to M. Kleiner for help with the training & A. Hof for help with data analysis.