ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Severin Stalder, Helmut Grabner, and Luc Van Gool
{sstalder,grabner,vangool}@vision.ee.ethz.ch

SCOVIS
BIWI Computer Vision

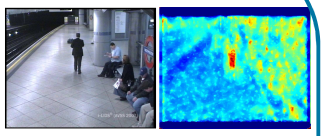
Cascaded Confidence Filtering for Improved Tracking-by-Detection

Abstract

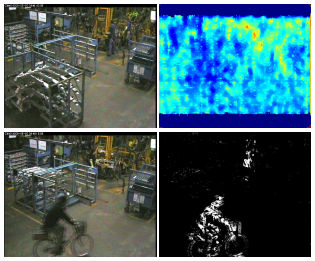
We propose a novel approach to increase the robustness of object detection algorithms in surveillance scenarios. The cascaded confidence filter successively incorporates constraints on the size of the objects, on the preponderance of the background and on the smoothness of trajectories.
In fact, the continuous detection confidence scores are analyzed locally to adapt the generic detector to the specific scene. The approach does not learn specific object models, reason about complete trajectories or scene structure, nor use multiple cameras. Therefore, it can serve as preprocessing step to robustify many tracking-by-detection algorithms.
Our real-world experiments show significant improvements, especially in the case of partial occlusions, changing backgrounds, and similar distractors

Introduction

i-LIDS dataset
- careful camera placement
- few occlusions
- person detector gives reasonable results



SCOVIS dataset
- industrial environment
- many partial occlusions
- detection confidence map is highly ambiguous



color background modeling: occluded persons vs. background structures?

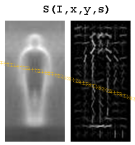
CCF: Cascaded Confidence Filtering

Detection confidence filter cascade for static cameras
Assumptions on background and smooth trajectories

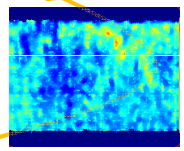
Input: **detection scores** of any sliding window detector
Output: **refined detections scores**
⇨ without learning person specific models
⇨ without modeling complete trajectories
⇨ without learning scene structure

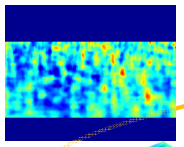
Pre-processing step for Tracking-by-Detection methods

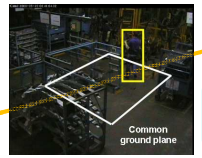
Filtering Steps



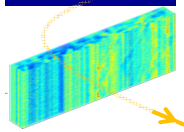
INPUT: object detector
- any sliding window detector can be given to CCF, we used [1]

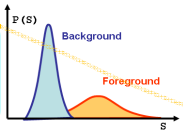




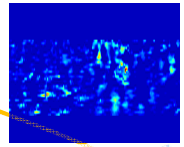


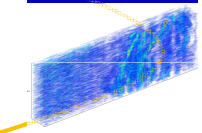
geometric filter
- common ground plane, similar to [2]

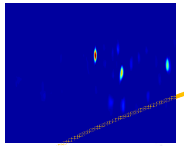


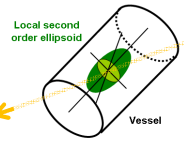


background filter
- long-term
- similar to traditional color background modeling [3]

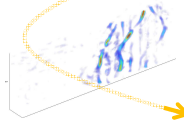


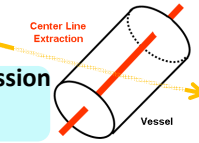






trajectory filter
- short-term
- similar to vessel enhancement in medical imaging [4]



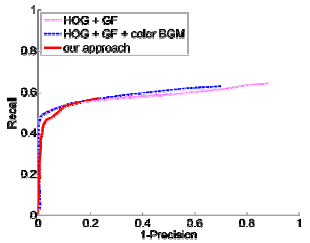


non-maxima suppression
- centerline extraction using particle filtering

OUTPUT: Refined detections scores

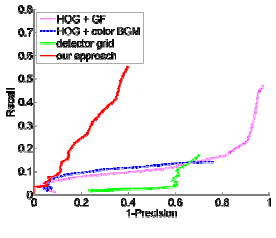
Evaluation

i-LIDS dataset



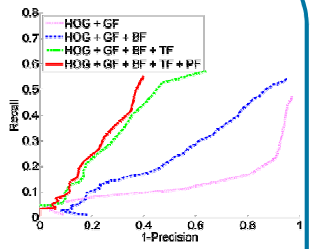
approach	recall	prec.	F-m.
HoG [2]+GF	54.7 %	86.6 %	67.0 %
simple fusion	54.3 %	87.5 %	67.0 %
our approach	53.3 %	89.4 %	66.8 %

SCOVIS dataset



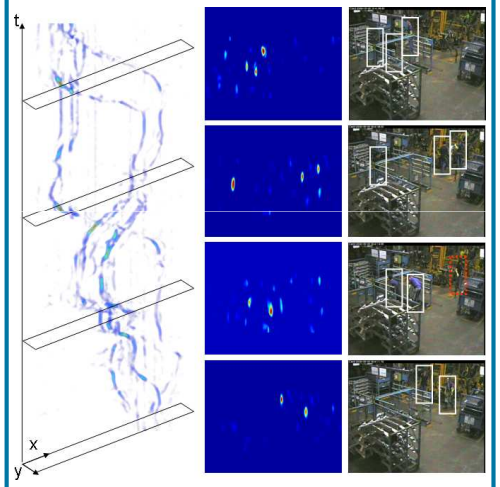
approach	recall	prec.	F-m.
HoG [2]+GF[5]	17.0 %	18.9 %	17.9 %
simple fusion	12.9 %	47.2 %	20.2 %
det. grid [6]	17.6 %	29.5 %	22.1 %
our approach	55.4 %	60.2 %	57.7 %

details



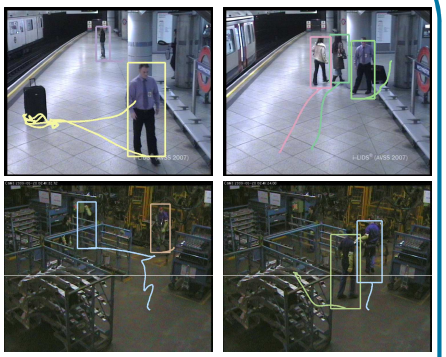
filter[5]	recall	prec.	F-m.
HoG [2] (input)	0.3 %	0.1 %	0.2 %
GF	17.0 %	18.9 %	17.9 %
GF+BF	30.1 %	38.6 %	33.8 %
GF+BF+TF	52.7 %	52.7 %	52.7 %
GF+BF+TF+PF	55.4 %	60.2 %	57.7 %

Detections using CCF

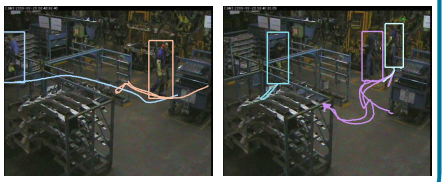


Improved Tracking-by-Detection

CCF output: tracklets



CCF results given to [5]



⇨ long trajectories, impossible to obtain only from raw detections!

References

[1] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proc. CVPR. Volume 1. (2005) 886–893
[2] Hoiem, D., Efros, A.A., Hebert, M.: Putting objects in perspective. In: Proc. CVPR. Volume 2. (2006) 2137–2144
[3] Stauffer, C., Grimson, W.: Adaptive background mixture models for real-time tracking. In: Proc. CVPR. Volume II. (1999) 246–252
[4] Frangi, A., Niessen,W., Vincken, K., Viergever, M.: Multiscale vessel enhancement filtering. (1998) 130–137
[5] Huang, C., Wu, B., Nevatia, R.: Robust object tracking by hierarchical association of detection responses. In: Proc. ECCV. (2008) 788–801