Cascaded Confidence Filtering for Improved Tracking-by-Detection

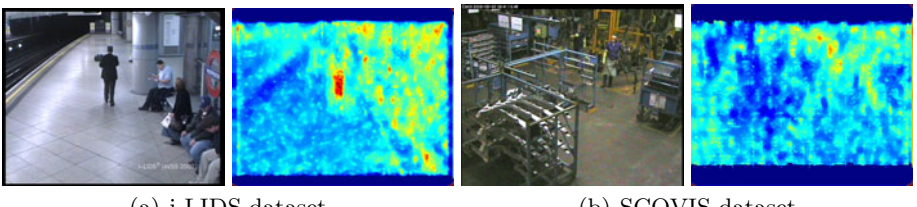
Severin Stalder[1], Helmut Grabner[1], and Luc Van Gool[1,2]

[1] Computer Vision Laboratory, ETH Zurich, Switzerland
{sstalder,grabner,vangool}@vision.ee.ethz.ch
[2] ESAT - PSI / IBBT, K.U. Leuven, Belgium
luc.vangool@esat.kuleuven.be

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.

1 Introduction

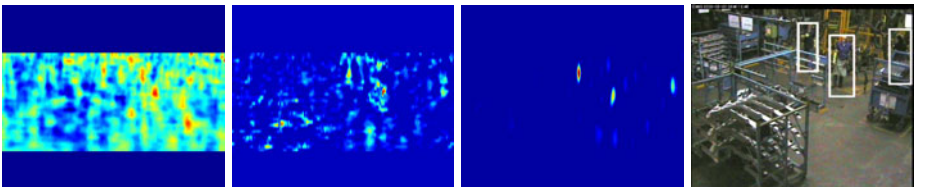
Monocular multi-object detection and tracking with static cameras is a challenging, but practically important problem. The task is inherently difficult due to the variability of object and background appearances. Most current methods depend on careful camera placement, as to avoid occlusions, see Fig. 1(a). However, in many settings this is virtually impossible, with deteriorating detection results as a consequence as illustrated in Fig. 1(b).



(a) i-LIDS dataset (b) SCOVIS dataset

Fig. 1. (a) Typical state-of-the art methods for object detection perform quite well when applied to current datasets, i.e. the maximum of the detection confidence map clearly corresponds to the fully visible person. (b) However, the detection confidence map is highly ambiguous in more cluttered scenes such as the SCOVIS dataset due to partial occlusions and similar structures in the background.

K. Daniilidis, P. Maragos, N. Paragios (Eds.): ECCV 2010, Part I, LNCS 6311, pp. 369–382, 2010.



(a) geometric filter (b) background filter (c) trajectory filter (d) final detections

Fig. 2. The proposed *Cascaded Confidence Filter* (CCF) refines the noisy confidence map to robustify tracking-by-detection methods. It combines prior information about the object size (a), predominant background (b) and smooth trajectories (c) in order to detect objects which move smoothly on the ground plane (d). The confidences correspond to the foot-points on the ground plane.

Currently, most object detectors consist of a learned appearance model, e.g., [1,2]. Despite significant, recent improvements, e.g., [3], their accuracy is still far from perfect. Further improvements increasingly come from the analysis of the detection context rather than merely analyzing the object patch, e.g., considering perspective [4]. In fact, context is also beneficial in pre-processing [5] or during learning of object detectors to better match training data to test data. In particular, classifier grids [6] learn separate object detectors at each image location in an on-line manner. However, unsupervised learning of such models might lead to label noise and deterioration of results [7].

Tracking-by-detection methods, e.g., [8,9], apply object detection independently in each frame and associate detections across frames to bridge gaps and to remove false positives. The use of continuous detection confidence scores in combination with thresholded detections and specific object models has been shown to facilitate target association [10]. Furthermore, scene specific knowledge, like entry/ exit zones to initialize trackers, helps to improve tracking results [9]. Yet, in tracking-by-detection with a static camera, one might also benefit from long-term observations of the scene, similar to detection and tracking in early surveillance scenarios which involved background models and change detection, e.g., [11].

We introduce a cascaded filtering of the detector confidence map, coined *Cascaded Confidence Filtering* or CCF. CCF incorporates constraints on the size of the object, on the preponderance of the background and on the smoothness of trajectories. In fact, the results of any sliding window detection algorithm can be improved with CCF in case of a static camera. As an example, Fig. 2 shows the successively refined detector confidence map from Fig. 1(b), finally allowing for detection of all three persons in that scene.

Contribution. CCF adapts a generic person detector to a specific scene. In particular, CCF combines a number of the reviewed approaches, trying to inherit their strengths while avoiding their weaknesses.

- CCF refrains from globally thresholding the detections. Instead, the continuous confidence scores of the detector are modeled at each position separately, similar to classifier grid approaches. In contrast to the latter, we rather use a fixed object model and analyze the detection confidences similar to color background modeling. The embedded use of a discriminative object model in background modeling allows to circumvent difficulties with noisy blobs, partial occlusions and different objects with similar shape.
- Detection responses are put into their spatial and temporal context as with tracking-by-detection. But these confidence levels are first filtered on a small spatial and a short temporal scale before reasoning about longer trajectories or scene structure. In fact, the smoothness of trajectories is ensured through a process analogous to vessel filtering in medical imaging. Those additional constraints permit to keep the advantages of an object detector while suppressing detections on background structures.

The refined detection confidence map can then be given to various existing tracking-by-detection frameworks modeling the scene structure and learning specific object models. It is shown that the filtered detections can be associated to long trajectories employing a typical tracking-by-detection method. Conversely, the same tracking-by-detection method performs dismally using the unfiltered detections.

The remainder of the paper is organized as follows. In Sec. 2 we present our CCF approach. Detailed experiments as well as improved object detection and tracking results are shown in Sec. 3. Finally, Sec. 4 suggests further work and concludes the paper.

2 Cascaded Confidence Filtering Approach

The input for our algorithm are confidence scores S, which are proportional to the likelihood that the object of interest appears at a certain position, i.e., $P(obj) \propto S$. More formally, let $S(I, x, y, s) \in \mathbb{R}$ be the continuous confidence score of the object detector at the center position (x, y) and scale s in an image I. Then, the goal of our method is to successively filter the detection confidence responses by including spatial and temporal context as depicted in Fig. 2 and more detailed in Fig. 3. Summarizing, any object of the object class of interest moving smoothly on a ground plane fulfills the assumptions of the filter steps and their detection scores are enhanced accordingly. Each individual filtering step is described in one of the following subsections.

2.1 Geometric Filter

The geometric filter incorporates the assumption that the objects move on a common ground plane restricting their possible size in the image, see second row of Fig. 3. This constraint has already been used either as post-processing in order to suppress inconsistent detections (e.g., [4]) or more recently as preprocessing to fit the image better to the training data of the detector, e.g., [5].

Following the latter approach, we only evaluate the object detection scores S' within appropriate candidate windows

$$P(obj|geometry) \propto S' = S(I, x, y, s) \tag{1}$$

where (x, y, s) satisfy the geometric constraints.

2.2 Background Filter

The second filter benefits from long-term observations as typically provided in surveillance situations. We follow a similar approach as in traditional background modeling. However, instead of modeling the pixel color values as in [11], we model the geometrically filtered confidence scores S'. Those are modeled using mixture of Gaussians ensuring robustness against environmental changes, e.g., of illumination. In fact, the detector confidence scores S' are modeled separately at each location as

$$P(S') = \sum_{k=1}^{K} w_k \eta(S', \mu_k, \sigma_k^2) \tag{2}$$

$$\text{where } \eta(S', \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(S'-\mu)^2}{2\sigma^2}}. \tag{3}$$

Updating the K mixture components and their weights w_k is done in the same on-line manner as described in [11]. We also use the heuristic assumptions of the latter approach to identify the mixture components belonging to background activities, i.e. by selecting the Gaussian distributions which have the most supporting evidence (mixture weights w_k) and the least variance σ_k^2. In fact, the mixture components are sorted according to w/σ^2 and the first B components are selected to model a defined portion T of the data. Finally, the probability of the object belonging to the foreground class is given by

$$P(obj|geometry, background) \propto S'' = \begin{cases} 1 - \sum_{b=1}^{B} w_b \eta(S', \mu_b, \sigma_b^2) \text{ if } \forall b : S' > \mu_b \\ 0 \qquad\qquad\qquad\qquad\qquad \text{otherwise} \end{cases}$$

$$\text{where } B = \arg\min_b \Big(\sum_{i=1}^{b} w_i > T \Big). \tag{4}$$

The intuition behind this heuristic is that the variance in the detection confidence of a background activity is smaller than the variance of moving objects of interest. Moreover, the background is generally expected to be present more often than moving objects. In contrast to [11], we only consider confidences higher than the mean of the assumed background distribution, since object presence should always lead to a higher detection confidence even when the background confidence is already quite high.

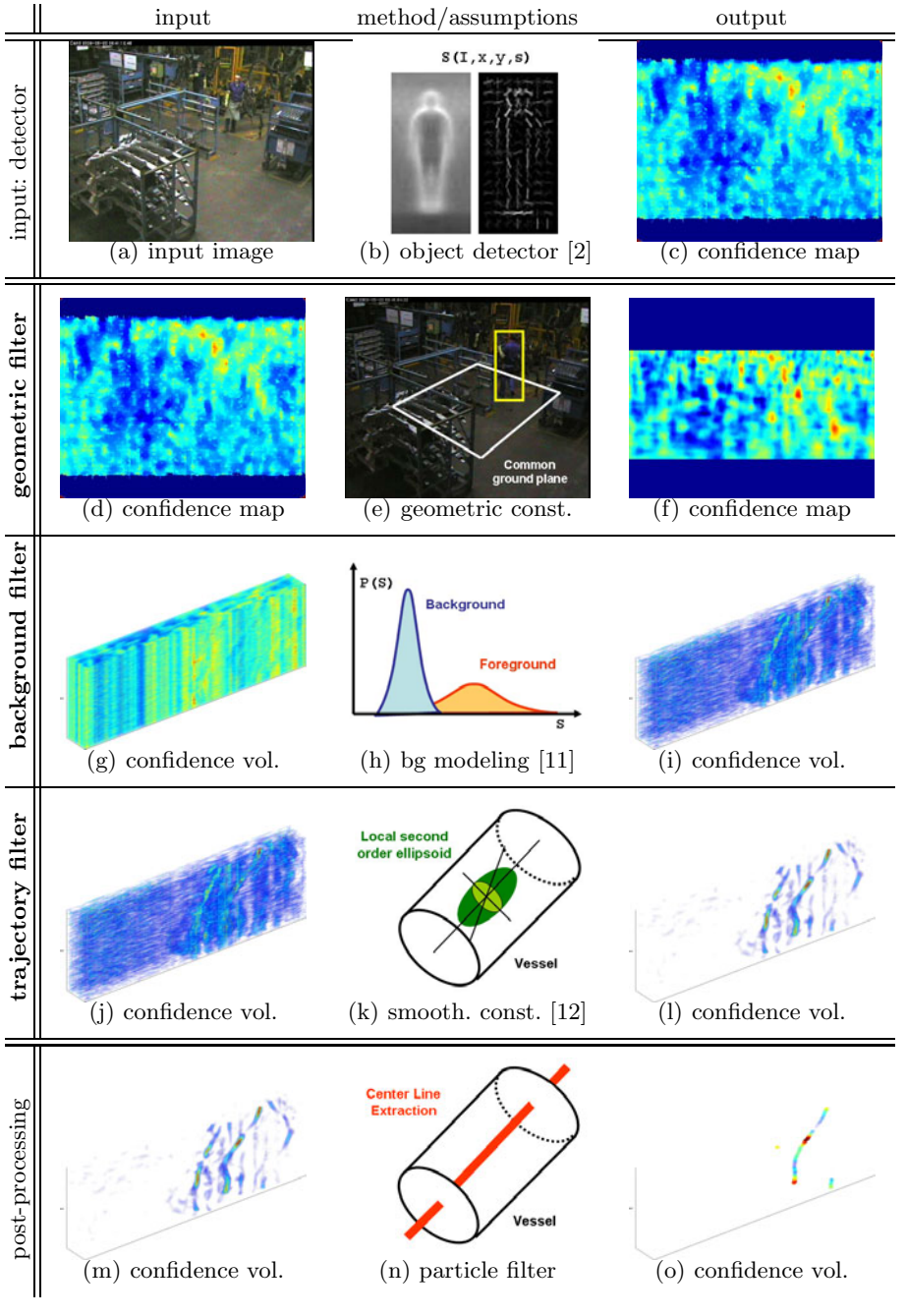


Fig. 3. The detection confidences are successively refined by our cascaded filtering approach incorporating a-priori constraints on scene, objects and trajectories

The proposed background filter can also be seen as *location specific threshold adaptation* for detection. Consequently, (i) systematically occurring false positives are removed, which increases precision, whereas (ii) the sensitivity in other regions (e.g., containing occluders) is increased, which possibly increases recall. An example is depicted in the third row of Fig. 3.

2.3 Trajectory Filter

The background filtered confidence map is supposed to suppress static background structures. The trajectory filter is designed to exclude other moving objects which might pass the previous filter step. In fact, there are two observations which are not likely to be correct for other moving objects. Firstly, an object appearance usually causes multiple (similar) responses within a local region in the image (see Fig. 1(a) or the related discussion in [1]). Secondly, the confidences should also be continuously changing over time, i.e. the object is not disappearing completely from one frame to the next.

We propose to analyze the volume spanned by the temporal aggregation of the confidence maps. In that volume, we enhance geometrical structures which resemble tubes while suppressing spurious or multiple responses. Therefore we apply the vessel filter approach proposed by Frangi et al. [12] in our context of trajectory filtering. In fact, the approach was originally designed to enhance vascular or bronchial vessel structures in medical data like magnetic resonance images. So a fitted ellipsoid in the defined spatial/ temporal volume Θ is used to extract the direction of elongation of high confidences. The approach does analyze the eigenvectors of the Hessian matrix to calculate the principal directions in which the local second order structure of the image can be decomposed. Therefore the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ (sorted ascend with respect to their absolute value) directly describe the curvature along the vessel. An ideal vessel has corresponding eigenvalues which fulfill the constraints

$$|\lambda_1| \approx 0 \text{ and } |\lambda_1| \ll |\lambda_2| \text{ and } \lambda_2 \approx \lambda_3, \tag{5}$$

,i.e., more or less no curvature along the first principle axis and similar (higher) curvature along the other two axes. The sign of the eigenvalues is negative in our case as locations on the trajectories are indicated through higher $P(obj|geometry, background)$. The "vesselness" measure is obtained on the basis of all eigenvalues of the Hessian matrix as

$$V(\Theta) = \begin{cases} 0 & \text{if } \lambda_2 > 0 \vee \lambda_3 > 0 \\ \left(1 - e^{-\frac{\lambda_1^2}{2|\lambda_2\lambda_3|\alpha^2}}\right) e^{-\frac{\lambda_2^2}{2\lambda_3^2\beta^2}} \left(1 - e^{-\frac{\lambda_1^2+\lambda_2^2+\lambda_3^2}{2c^2}}\right) & \text{otherwise} \end{cases} \tag{6}$$

where α, β and c are thresholds which control the sensitivity (please refer to [12] for more details). This function provides the probabilistic output

$$P(obj|geometry, background, trajectory) \propto V(\Theta). \tag{7}$$

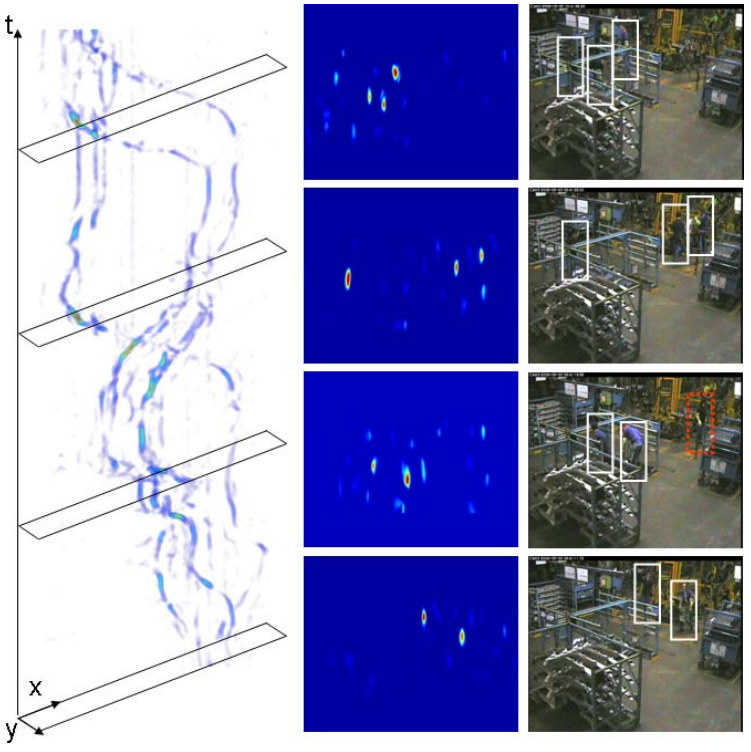


Fig. 4. Left: filtered detector confidence volume by our CCF approach with four transparent planes at selected time instances. Middle: confidence maps for the 4 instances projected onto the ground plane. Right: the corresponding frames overlaid with the detection results. As it can be seen, all the detections at this stage (white) correspond to true positives. There are no false positives among them in this example. Only one person was left undetected (dotted red).

An example is shown in the fourth row of Fig. 3. Additionally, a longer sequence (about half a minute) with corresponding images overlaid with the detections is depicted in Fig. 4.

2.4 Post-Processing

After applying the trajectory filter, possible trajectory-like structures are enhanced, but not yet segmented. Inspired by particle filtering methods of [13] for segmentation of coronaries and [10] for multi-object tracking, we use a simple particle filter to resolve ambiguities and output detections, see last row of Fig. 3. This last filtering step has a similar purpose as the common non-maxima suppression, used for instance in [1] to get non-overlapping detections. However, in the case of video streams, one should benefit from the temporal information.

We use a bootstrap filter, where the state of a particle $\mathbf{x} = (x, y, u, v)$ consists of the 2D position (x, y) on the ground plane and the velocity components (u, v). For prediction, we assume a constant velocity motion model, i.e.,

$$(x, y)_t = (x, y)_{t-1} + (u, v)_{t-1} \cdot \Delta t + \epsilon_{(x,y)} \tag{8}$$

$$(u, v)_t = (u, v)_{t-1} + \epsilon_{(u,v)} \tag{9}$$

where the process noise $\epsilon_{(x,y)}$, $\epsilon_{(u,v)}$ for each state variable is independently drawn from zero-mean normal distributions.

As observations we use the values from the filtered spatial/ temporal volume. The importance weight $w^{(i)}$ for each particle i at time step t is then described by:

$$w_t^{(i)} \propto w_{t-1}^{(i)} \cdot P(obj|(x, y)_t, geometry, background, trajectory) \tag{10}$$

From that weight distribution, re-sampling in each time step is performed using a fixed number of N particles.

For tracklet initialization, the filtered confidence must be above a certain user defined threshold θ_{init}.

Furthermore, the particles are gated in order to remove the multi-modality provided by particle filtering if the filtered confidence exceeds another user-defined threshold $\theta_{gating} \geq \theta_{init}$.

For each image, the particle with the highest filtered confidence corresponds to the assumed position of the object in this frame. The object can then be detected in the image by mapping it back to the corresponding image coordinates.

3 Experimental Results

In principle, the proposed CCF approach can be used with any sliding windows object detector, e.g., to locate cars or faces, as long as the camera remains static. However, our experiments are focused on the task of human detection.

Datasets. We use two video sequences for evaluation: (i) the public i-LIDS AB Easy dataset[1] and (ii) our recorded SCOVIS dataset.

The SCOVIS dataset was captured at a workstation in a car manufacturing site. It is challenging due to the industrial working conditions (e.g., sparks and vibrations), difficult structured background (e.g., upright racks, and heavy occlusions of the workers in most parts of the image), and other moving objects (e.g., welding machines and forklifts). Of the total 8 hours of video we evaluated a sequence of 5,000 frames after 5,000 frames used to initialize the background statistics. We manually annotated every 10^{th} frame for evaluation purposes.

3.1 Implementation Details

In this section, we shortly report the details of our implementation. Most subparts are publicly available and should make our experiments easily repeatable.

[1] Available at `http://www.elec.qmul.ac.uk/staffinfo/andrea/avss2007_d.html`, 2010/03/10

Person detector (HoG). We use the OpenCV 2.0 implementation of the histogram of oriented gradients (HoG) person detector [2][2]. Please note that more sophisticated part-based approaches, e.g., [3], could also be used.

Geometric filter (GF). We assume one common ground plane in the scene with a flat horizon. The ground plane calibration, i.e., determining the scale of a typical person at each position on the ground plane is done manually. The detector is evaluated on a virtual grid in the image plane with 32×297 resolution in the SCOVIS and 32×375 resolution in the i-LIDS dataset.

Background filter (BF). We use the on-line mixture of Gaussians implementation of Seth Benton[3] with $K = 2$ mixture components. The time constant for updating the distributions, see [11], is set to $\alpha = 10^{-4}$ in both datasets. The background model is expected to capture at least 80% of the mixture model, i.e., we set $T = 0.8$ in Eq.(4). Both parameters are set to ensure that standing persons are not easily integrated in the background model.

Trajectory filter (TF). We use Dirk-Jan Kroon's implementation[4] of the Frangi vessel filter [12] with default parameters and scale parameter set to 3.

Particle filter (PF). We use N=150 particles, the variance for the position noise is set to σ_x^2=0.8 and σ_y^2=0.2 and the variance for the velocity noise is set to $\sigma_{u,v}^2 = 0.1$, all variances are set with respect to the grid in the image plane.

Person detection takes approximately 600 ms per frame, BF about 6 ms, TF about 14 ms and PF about 10 ms using a single core of a 3 GHz processor. The computational bottleneck is clearly the detection part, the proposed filter steps are essentially real-time.

3.2 Improving Object Detection

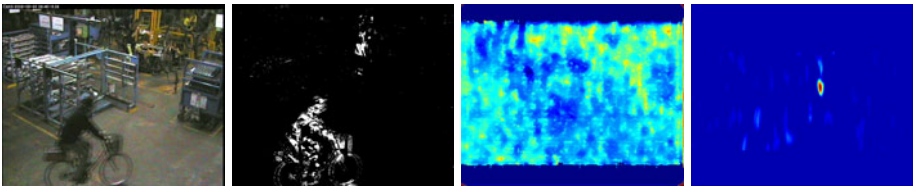
Our CCF approach combines the strength of background and appearance based approaches, whereas the individual ones are going to fail, as depicted in Fig. 5. For instance, pixel-wise color background subtraction will simply focus on all moving objects. Besides, occluded persons are hard to discern in the noisy foreground mask. Furthermore, the pre-trained detector ignores temporal information completely and a global threshold does not allow to detect occluded persons in the presence of difficult background. Our approach overcomes those limitations since the location specific detection confidence values are modeled temporally.

Details. Two typical detector confidence distributions and the fitted Gaussian mixture models are visualized in Fig. 6. At the position marked in green in Fig. 6(a), one Gaussian is considered to represent the background activities,

[2] Available at `http://sourceforge.net/projects/opencvlibrary`, 2010/02/24

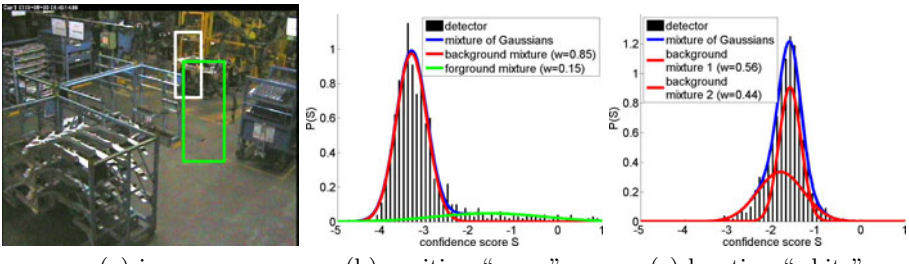
[3] Available at `http://www.sethbenton.com/mixture_of_gaussians.html`, 2010/03/07

[4] Available at `http://www.mathworks.com/matlabcentral/fileexchange/24409-hessian-based-frangi-vesselness-filter` ,2010/02/24



(a) image (b) change detection (c) person detector (d) our approach

Fig. 5. Confidence scores of different methods. (b) simple background modeling with change detection also reports other moving objects while partially occluded persons are difficult to retrieve, (c) appearance based object detectors have difficulties in detecting occluded persons and discriminating the background, (d) our approach successfully refines the confidence map allowing for better detection results.

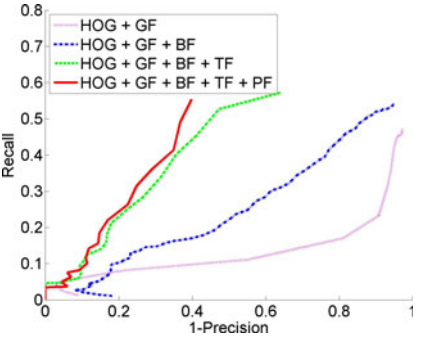


(a) image (b) position "green" (c) location "white"

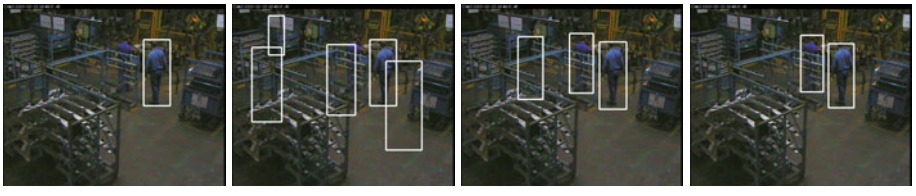
Fig. 6. Two typical mixture of Gaussians obtained by background modeling of [11] applied on detector confidence values. Mixture components which are considered to belong to the background model are shown in red, with $T = 0.8$.

whereas the second Gaussian (with a larger variance) is considered to model the activity of the object of interest at this location. At the second location depicted in 6(b), no persons are present (the welding machine is there) only parts of persons or sparks. Hence, both Gaussians with relatively high mean confidences belong to the background model. This also explains the drawbacks of using one global threshold for the whole image. Please note that these confidence statistics are computed over a relatively long time interval to not quickly integrate standing persons into the background model.

Evaluation. For a quantitative evaluation, we use recall-precision curves (RPCs). The recall corresponds to the detection rate whereas the precision relates to the trustfulness of a detection. In particular, a detection is accepted as true positive if it fulfills the overlap criterion of [14], i.e., a minimal overlap $a_0 = (\text{area}(B_p) \cap \text{area}(B_{gt}))/(\text{area}(B_p) \cup \text{area}(B_{gt}))$ of 50 % is needed between the predicted bounding box B_p and the ground truth bounding box B_{gt} to count as true positive. Additionally, we also report recall and precision at maximized F-measure which is the harmonic mean between recall and precision.



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 %



GF GF+BF GF+BF+TF GF+BF+TF+PF

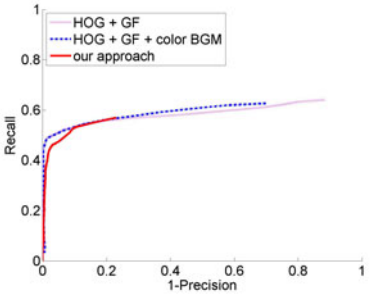
Fig. 7. Recall-precision curves for each individual filter step in the SCOVIS dataset. Bottom: Example detections at maximized f-Measure

The RPC of each individual filter step in the SCOVIS dataset is depicted in Fig. 7 in conjunction with example detections. Improved results are manifested after each individual filter step of CCF in terms of increased recall and precision. Since the filtering steps are complementary the performance can be boosted by roughly 15-20 % in terms of f-Measure each. In particular, the geometric filter does not consider detections at inappropriate scales, the background filter adjust locally the detection thresholds (increasing the sensitivity) while the trajectory filter enforces temporal consistency. The post-processing to better align the detections increases precision and recall only slightly.

Comparisons. Quantitative and qualitative comparisons are shown in Fig. 8 and Fig. 9 for the i-LIDS AB Easy dataset and the challenging SCOVIS dataset, respectively. As baseline, we consider the detection results after the geometric filter since similar assumptions, e.g., about the ground plane, are often found in literature, e.g., [4]. The i-LIDS AB Easy datasets was chosen intentionally to illustrate the formidable performance of the person detector which is not lowered by CCF. In particular, the datasets contains several standing persons whose detection is shown to not be affected by the filter cascade.

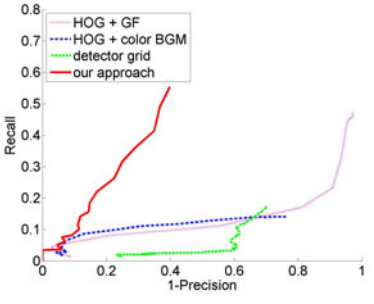
However, in the more challenging SCOVIS dataset with poor detection results, CCF significantly improves the results by about 40 % in terms of f-Measure. Detailed results of applying our algorithm on both datasets are depicted in Fig. 10.

[5] HoG: Histogram of Gradients person detector [2]; GF: geometric filter; BF: background filter; TF: trajectory filter; PF: particle filter.



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 %

Fig. 8. Results for the i-LIDS dataset in comparison to other approaches. Please note that the ground truth includes fully occluded persons which are impossible to detect. This said, the input is nearly perfect and all approaches perform similarly.



approach	recall	prec.	F-m.
HoG [2]+GF[6]	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 %

HoG + GF | simple fusion | det. grid | our approach

Fig. 9. Results for the SCOVIS dataset in comparison to other approaches

Additionally, we also evaluated a simple combination of traditional background modeling and human detection, i.e., a detection is only taken into account if at least 20 % of the bounding box is not modeled as foreground. However, this combination does not improve the results significantly as it is just a verification step a posteriori. We also compared our improved detections to the recently proposed approach of learned location specific classifiers [6][7] which aims to include background information during learning. Whereas they are able to show improved results on common datasets, the results clearly shows that such an approach can not cope well with the challenges of the SCOVIS dataset.

[6] The part-based person detector of [3] achieved a recall of 2.3 % at a precision of 63.3 % with geometric filtering and default parameters.

Fig. 10. Short tracklets obtained by using the proposed CCF approach to improve on the HoG detection confidence. More examples are given on the authors' web page.

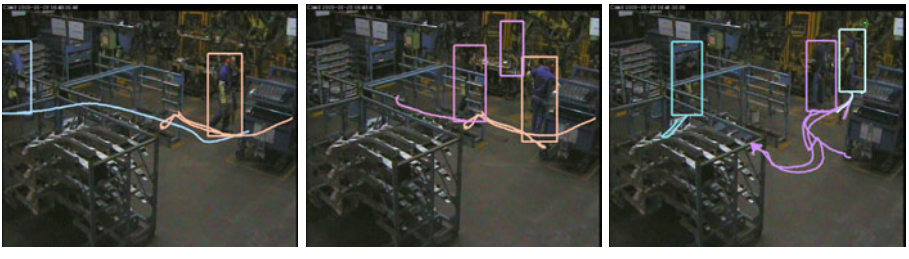


Fig. 11. Complete trajectories found by [9] using the improved detection results of our approach. The approach performs dismally if the HoG detections are directly passed to their detection association method.

3.3 Improving Tracking-by-Detection

Recent work explores tracking-by-detection [8,9], i.e. applying an object detector in each frame and then associating the detections across frames. The post-processing of CCF links the detections similarly, but at a lower level without extrapolation. To indicate improved tracking-by-detection results, we employ the approach of [9] which performs global trajectory association in an hierarchical manner[7]. The experiment was run twice, (i) providing the raw HoG detections and (ii) the improved ones obtained by CCF. Whereas the approach is performing dismally using the raw detections (not shown here), long trajectories are output when using CCF as pre-processing. Tracking results are depicted in Fig. 11.

[7] We gratefully thank the authors for applying their code on the SCOVIS dataset.

4 Conclusion and Future Work

We presented a novel method for filtering the detection confidences in surveillance scenarios. Our approach reamins very general, only requiring a static camera and a sliding-windows object detector. CCF involves geometric constraints, long-term temporal constraints to suppress the background confidence distribution and short-term smoothness constraints on possible trajectories. The experimental evaluation on the task of person detection shows significant improvement over the input object detection results, especially in the case of occlusions and cluttered background. The approach does not learn specific object models, incorporate scene specific constraints, reason about complete trajectories, or use multiple cameras. All those extensions remain to be explored to further robustify tracking-by-detection methods.

Acknowledgments. This research was supported by the European Communitys Seventh Framework Programme under grant agreement no FP7-ICT-216465 SCOVIS. We further thank Lee Middleton and Christine Tanner for inspiring discussions.

References

1. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. CVPR, vol. I, pp. 511–518 (2001)
2. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proc. CVPR, vol. 1, pp. 886–893 (2005)
3. Felzenszwalb, P., McAllester, D., Ramanan, D.: A discriminatively trained, multiscale, deformable part model. In: Proc. CVPR (2008)
4. Hoiem, D., Efros, A.A., Hebert, M.: Putting objects in perspective. In: Proc. CVPR., vol. 2, pp. 2137–2144 (2006)
5. Li, Y., Wu, B., Nevatia, R.: Human detection by searching in 3d space using camera and scene knowledge. In: Proc. ICPR (2008)
6. Roth, P., Sternig, S., Grabner, H., Bischof, H.: Classifier grids for robust adaptive object detection. In: Proc. CVPR (2009)
7. Stalder, S., Grabner, H., Gool, L.V.: Exploring context to learn scene specifc object detectors. In: Proc. PETS (2009)
8. Leibe, B., Schindler, K., Gool, L.V.: Coupled detection and trajectory estimation for multi-object tracking. In: Proc. ICCV (2007)
9. Huang, C., Wu, B., Nevatia, R.: Robust object tracking by hierarchical association of detection responses. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 788–801. Springer, Heidelberg (2008)
10. Breitenstein, M., Reichlin, F., Leibe, B., Koller-Meier, E., Gool, L.V.: Robust tracking-by-detection using a detector confidence particle filter. In: Proc. ICCV (2009)
11. Stauffer, C., Grimson, W.: Adaptive background mixture models for real-time tracking. In: Proc. CVPR, vol. II, pp. 246–252 (1999)
12. Frangi, A., Niessen, W., Vincken, K., Viergever, M.: Multiscale vessel enhancement filtering, pp. 130–137 (1998)
13. Florin, C., Paragios, N., Williams, J.: Particle filters, a quasi-monte carlo solution for segmentation of coronaries. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 246–253. Springer, Heidelberg (2005)
14. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes Challenge (VOC 2009) Results (2009)