USING HIGH LEVEL INFORMATION FOR LOW LEVEL TRACKING

Horbert E., Mitzel D., Leibe B.
UMIC Research Centre RWTH Aachen University, Germany

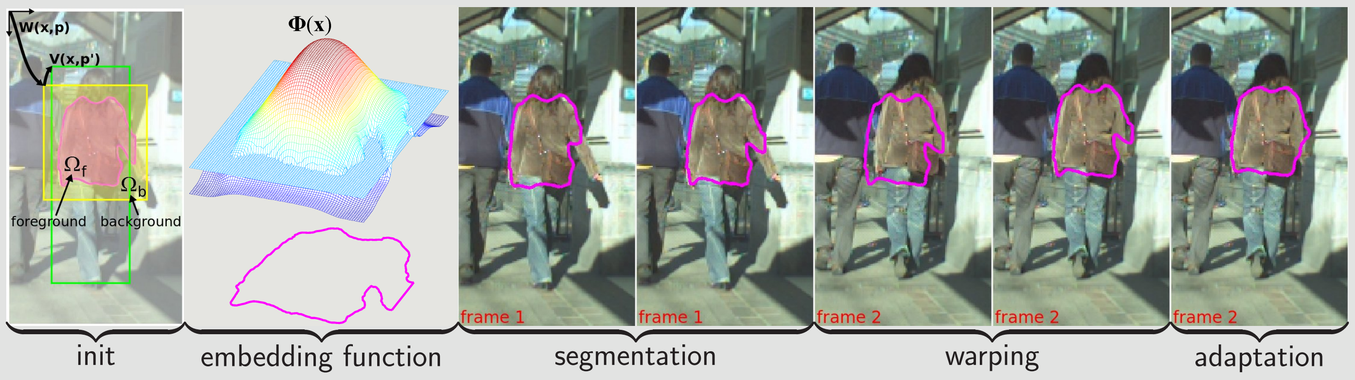
RWTH AACHEN UNIVERSITY
MMP
Mobile Multimedia Processing

ICVSS 2011
Sicily ~ 11-16 July
International Computer Vision Summer School

Abstract

Level Set Tracking is a low level tracking approach, where the object of interest is segmented in the first frame, then tracked through the subsequent frames, in which the contour is adapted. We integrate LS tracking with a tracking-by-detection framework and thereby show how LS tracking can benefit from high level information, e.g. a ground plane estimate. The LS tracker can also pass information to higher levels, such as object positions, a car's rotation angle or detailed appearance models.

Level Set Tracking



- Segmentation

$$\frac{\partial P(\Phi,\mathrm{p}|\Omega)}{\partial \Phi} = \underbrace{\frac{\delta_\epsilon(\Phi)(P_f - P_b)}{P(\mathrm{x}|\Phi,\mathrm{p},\mathrm{y})}}_{\text{deviation from fg/bg model}} + \underbrace{\frac{1}{\sigma^2}\left[\nabla^2\Phi - div\left(\frac{\nabla\Phi}{|\nabla\Phi|}\right)\right]}_{\text{deviation from signed dist. fct.}} + \underbrace{\lambda\delta_\epsilon(\Phi) div\left(\frac{\nabla\Phi}{|\nabla\Phi|}\right)}_{\text{contour's length}}$$

where $P(\mathrm{x}|\Phi,\mathrm{p},\mathrm{y}) = H_\epsilon(\Phi(\mathrm{x}))P_f + (1 - H_\epsilon(\Phi(\mathrm{x})))P_b$, ∇^2 is the Laplacian, H_ϵ is a smoothed Heaviside step function and δ_ϵ its derivative, a smoothed Dirac delta function.

- Warping

$$\Delta\mathrm{p} = \left[\sum_{i=1}^{N} \frac{1}{2P(\mathrm{x}_i|\Phi,\mathrm{p},\mathrm{y}_i)}\left[\frac{P_f}{H_\epsilon(\Phi(\mathrm{x}_i))} - \frac{P_b}{(1-H_\epsilon(\Phi(\mathrm{x}_i)))}\right]\mathrm{J}^\mathsf{T}\mathrm{J}\right]^{-1} \times \sum_{i=1}^{N} \frac{(P_f - P_b)\mathrm{J}^\mathsf{T}}{P(\mathrm{x}_i|\Phi,\mathrm{p},\mathrm{y}_i)}$$

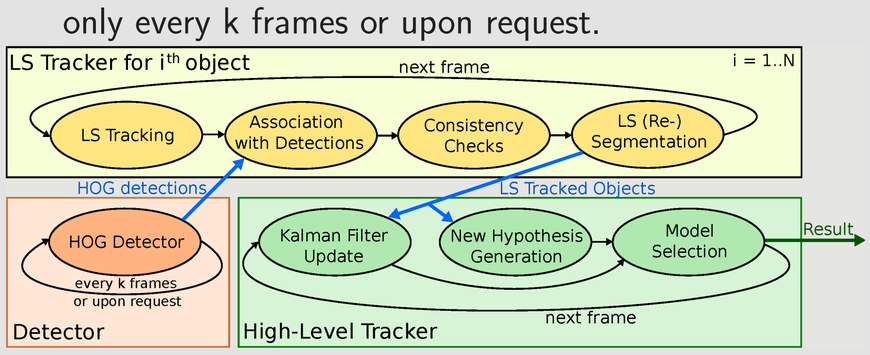
Extension of C. Bibby, I. Reid: Robust Real-Time Visual Tracking using Pixel-Wise Posteriors. ECCV (2008)

Multi-Person Tracking with Sparse Detection and Continuous Segmentation [ECCV'10]

Integration of LS tracking and tracking-by-detection

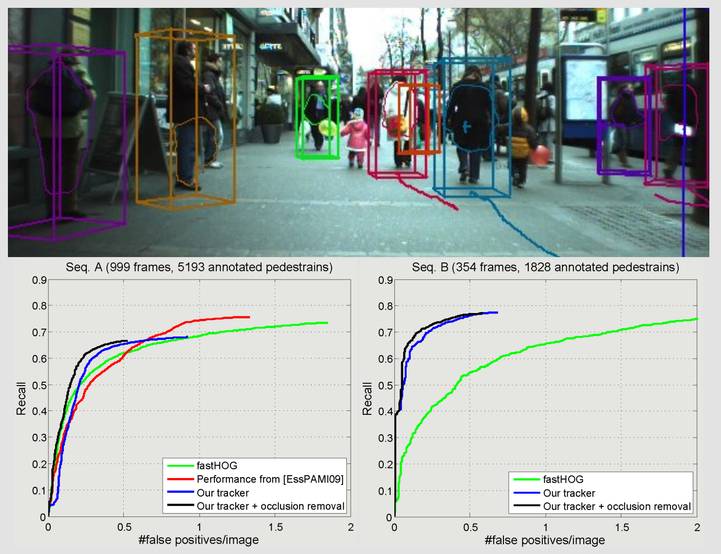
- Start LS tracker for each newly detected person
- LS tracker performs short-term data association
- High-level tracker resolves multi-person associations

⇒ Run the computationally expensive object detector only every k frames or upon request.



- Foreground/Background models based on appearance and depth:
  $P_i = (1-\alpha)P_{i,color} + \alpha P_{i,depth}, \quad i \in \{f, b\}$
- LS tracker performs consistency checks and can terminate tracklets or request new detections.

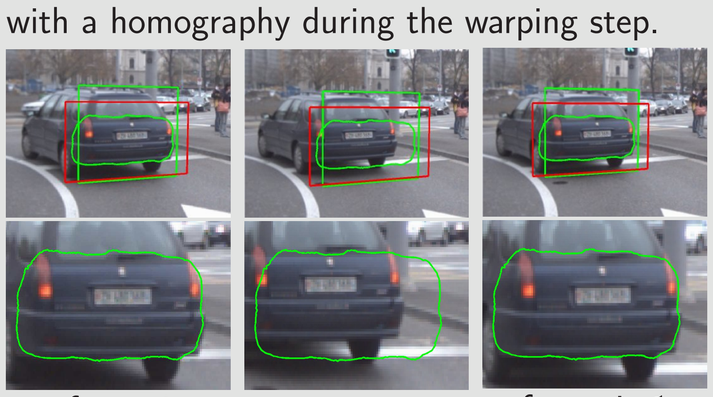
+ The level set tracker can continue tracking persons that are partially occluded or near the camera, for which no new detections are available.



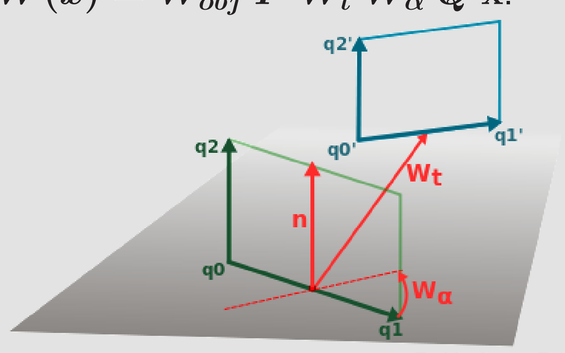
D. Mitzel, E. Horbert, A. Ess, B. Leibe: Multi-Person Tracking with Sparse Detection and Continuous Segmentation. ECCV (2010)

Geometrically Constrained Level Set Tracking for Automotive Applications [DAGM'10]

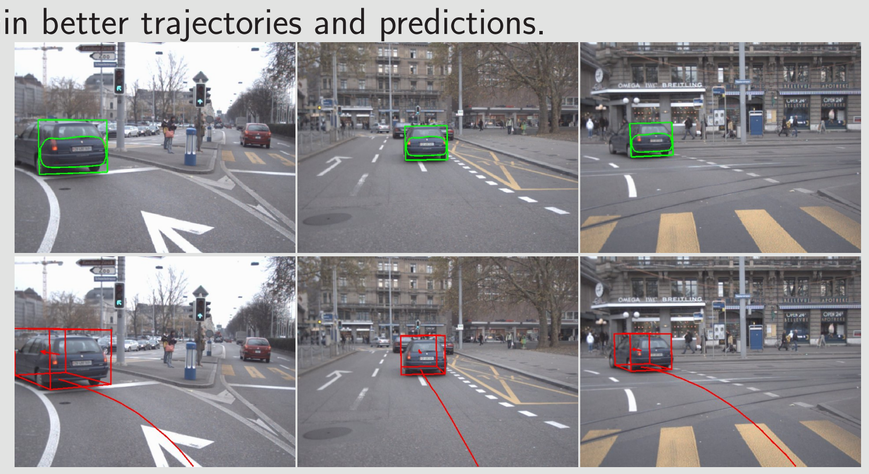
Idea: Track cars in *"3D"* (capture the perspective distortions) by transforming the contour with a homography during the warping step.



- Constrain the homography to rigid motion on the groundplane for greater robustness.
- Gauss Newton along the gradient $\frac{\partial W}{\partial \mathrm{p}}$ with $W(x) = W_{obj}\ P\ W_t\ W_\alpha\ Q\ \mathrm{x}$.



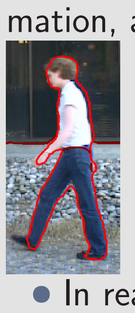
A high-level tracker can use the rotation angle in a motion model to obtain better trajectories and predictions.



E. Horbert, D. Mitzel, B. Leibe: Geometrically Constrained Level Set Tracking for Automotive Applications. DAGM (2010)

Person Segmentation and Tracking with Multi-Region Appearance Models and Top-Down Shape Information

Goal: Accurate segmentation of persons, e.g. for pose estimation, appearance models for recognition.



- Appearance models consisting of one color histogram can be too weak, e.g. if trousers have same color as background of head.
  Idea: Subdivide into multiple regions with one color histogram each.
- In realistic scenes it is often difficult to distinguish persons from background.
  Idea: Use top-down segmentations from a Hough Forest Detector (Implicit Shape Model) as shape prior.

- Regions $k \in \{f1, f2, b1, b2\}$
- Shape prior $P(M_k|\mathrm{h})$

$$P(\mathrm{x}|\Phi,\mathrm{p},\mathrm{y},\mathrm{h}) = H H_f P_{f1} + H\tilde{H}_f P_{f2} + \tilde{H} H_b P_{b1} + \tilde{H}\tilde{H}_b P_{b2}$$

where $P_k = \frac{P(\mathrm{y}|M_k)P(M_k|\mathrm{h})}{\sum_l \eta_l P(\mathrm{y}|M_l)}$

submitted to ICCV (2011)