ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Multi-target Tracking by Continuous Energy Minimization

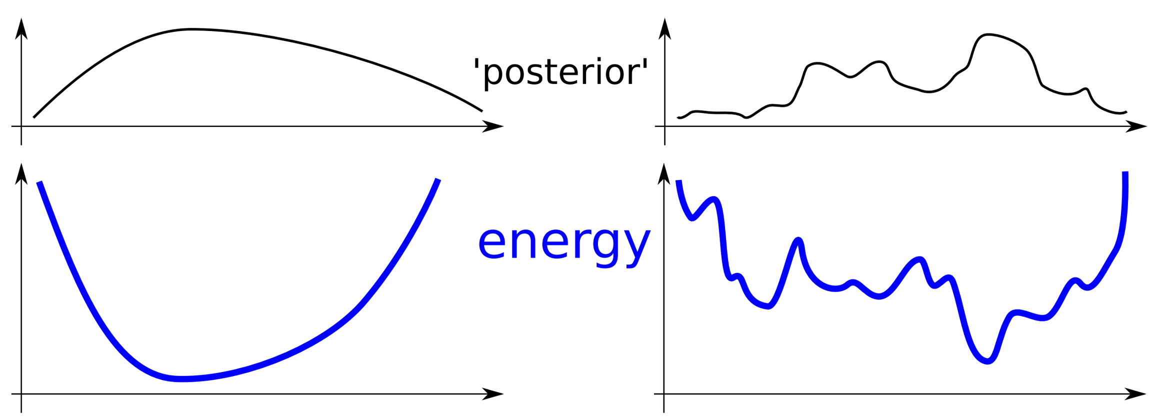
Anton Andriyenko[1] and Konrad Schindler[2]
[1] Computer Science Department, TU Darmstadt, Germany
[2] Photogrammetry and Remote Sensing Group, ETH Zürich

TECHNISCHE UNIVERSITÄT DARMSTADT

Objectives

- an energy function which represents the actual situation as faithfully as possible
- no a-priori restrictions of the state space
- a suitable (local) optimization scheme for the resulting energy

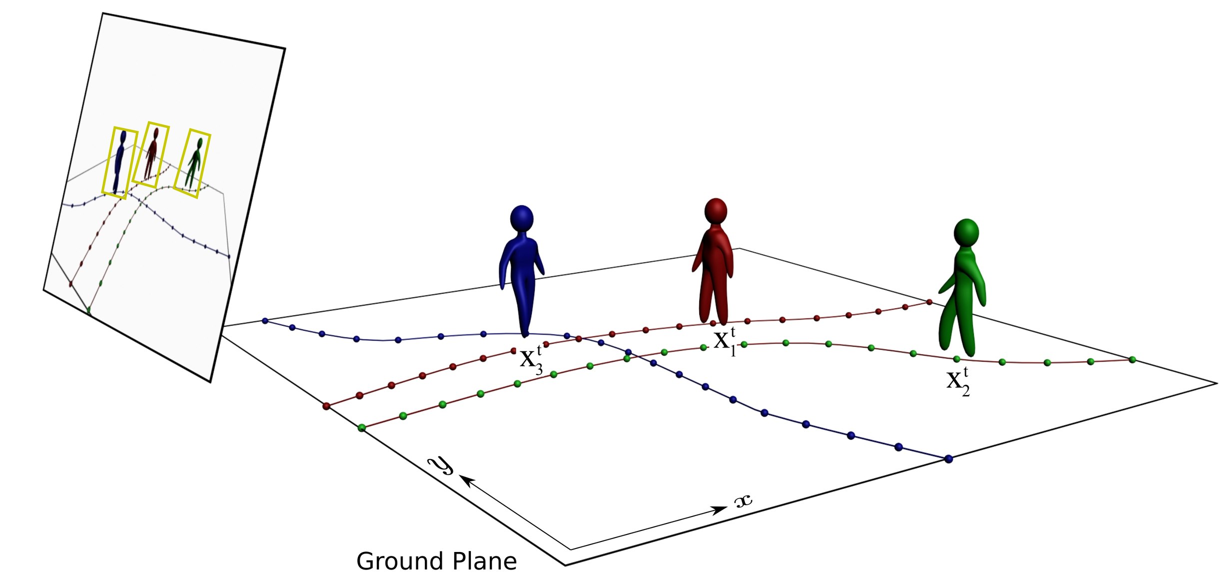
Choice of an energy function



convex (submodular)	non-convex
globally optimizable	locally optimizable
ILP [1, 2]	**gradient descent + jumps**
Network Flows [3]	
crude approximation	**better approximation**

Setting

- Tracking-by-detection: sliding window HOG, HOF [4]
- Tracking is performed in 3D (on the ground plane)
- State vector X consists of continuous ground plane coordinates of *all* targets in *all* frames



References

[1] H. Jiang, S. Fels, and J. J. Little. A linear programming approach for multiple object tracking. In *CVPR*, 2007.

[2] J. Berclaz, F. Fleuret, and P. Fua. Multiple object tracking using flow linear programming. In *Winter-PETS*, 2009.

[3] L. Zhang, Y. Li, and R. Nevatia. Global data association for multi-object tracking using network flows. In *CVPR*, 2008.

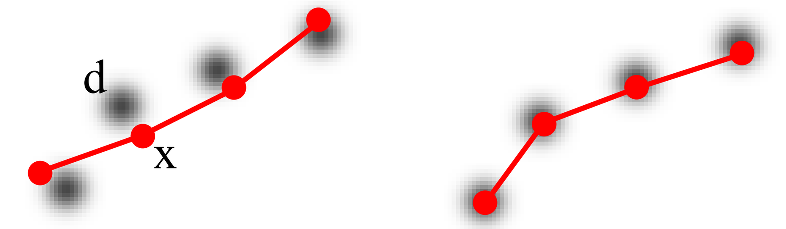
[4] Stefan Walk, Nikodem Majer, Konrad Schindler, and Bernt Schiele. New features and insights for pedestrian detection. In *CVPR*, 2010.

Energy

Global, continuous, differentiable

$$E = E_{\text{obs}} + \alpha E_{\text{dyn}} + \beta E_{\text{exc}} + \gamma E_{\text{per}} + \delta E_{\text{reg}} \qquad (1)$$

Observation Model keeps trajectories close to detections

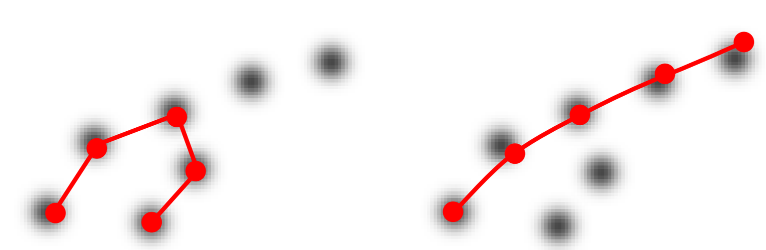


Targets locations are not restricted to a discrete grid or non-maxima suppressed detections.

$$E_{\text{obs}}(\mathbf{X}) = \sum_{t=1}^{F}\sum_{i=1}^{N}\left(\lambda - \sum_{g=1}^{D(t)} \frac{c}{\|\mathbf{x}_i^t - \mathbf{d}_g^t\|^2 + c}\right)$$

where $\mathbf{x}_i^t$ = location of target i and $\mathbf{d}_g^t$ = location of detection g at frame t

Dynamic Model favors constant velocity, performs 'intelligent smoothing'

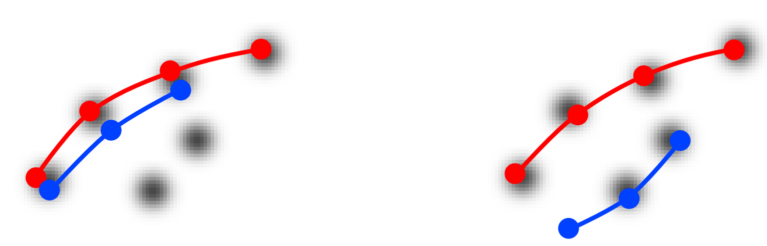


In continuous state space the dynamic model does not suffer from aliasing.

$$E_{\text{dyn}}(\mathbf{X}) = \sum_{t=1}^{F-2}\sum_{i=1}^{N}\|\mathbf{v}_i^t - \mathbf{v}_i^{t+1}\|^2$$

where $\mathbf{v}_i^t = \mathbf{x}_i^{t+1} - \mathbf{x}_i^t$

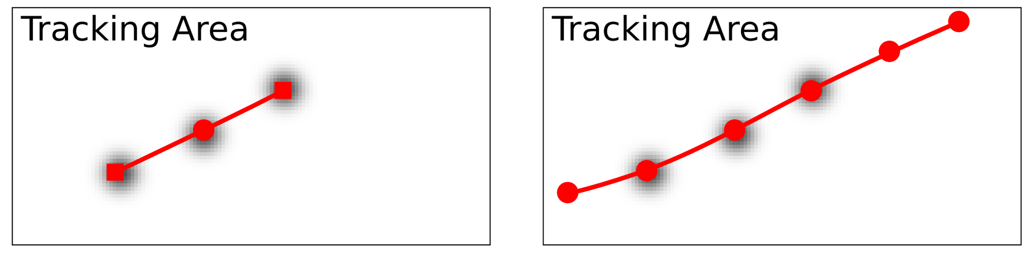
Mutual Exclusion avoids collisions



This continuous constraint accurately captures the actual overlap between target volumes.

$$E_{\text{exc}}(\mathbf{X}) = \sum_{t=1}^{F}\sum_{i \neq j} \frac{s_g^2}{\|\mathbf{x}_i^t - \mathbf{x}_j^t\|^2}$$

Persistence avoids abrupt termination of trajectories



$$E_{\text{per}}(\mathbf{X}) = \sum_{i=1}^{N}\sum_{t \in \{1,F\}} \frac{1}{1 + \exp(1 - q \cdot b(\mathbf{x}_i^t))}$$

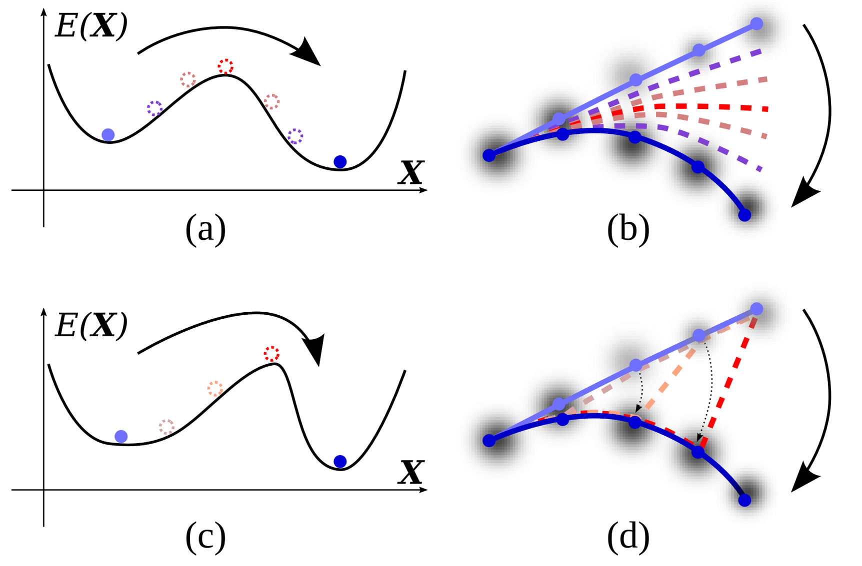
where $b(\cdot)$ denotes the distance to the border of the tracking area

Regularization keeps the solution simple with fewer targets and longer trajectories

$$E_{\text{reg}}(\mathbf{X}) = N + \sum_{i=1}^{N}\frac{1}{F(i)}$$

Non-convexity

The energy (1) is clearly not convex:



The ridge of high energy between two local minima is caused either by E_{obs} (a-b) or by E_{dyn} (c-d).

In general, non-convexity is due to the high-order dependence between variables caused by physical constraints.

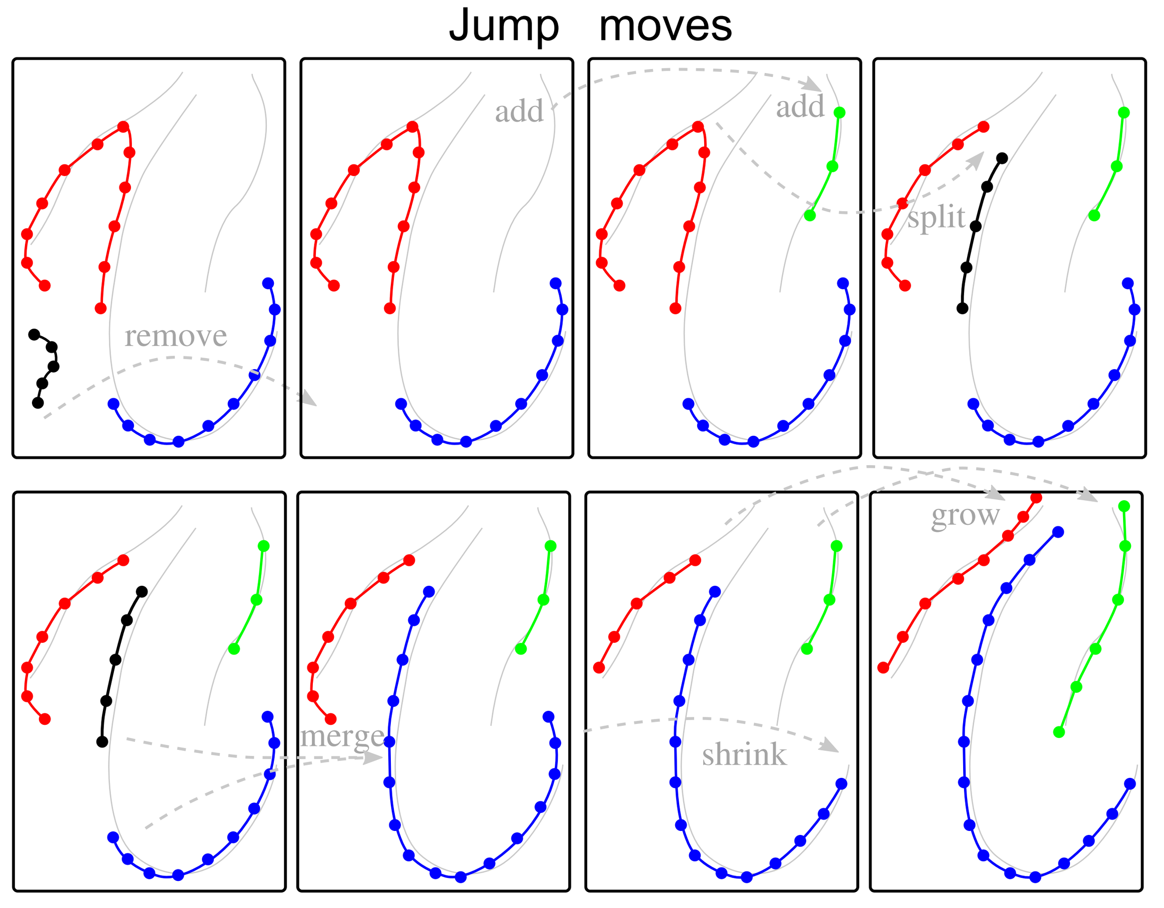
Minimization

Conjugate Gradient + 6 trans-dimensional jumps

Input: S initial solutions
Output: Best of $\leq S$ solutions

```
for s = 1 → S
  for m = 1 → 6
    try jump move m (greedy)
    if E_new < E_old
      perform jump move
      perform conjugate gradient descent
    end if
  end for
end for
```

Return: $\arg\min_{\mathbf{X}_s} E(\mathbf{X}_s)$

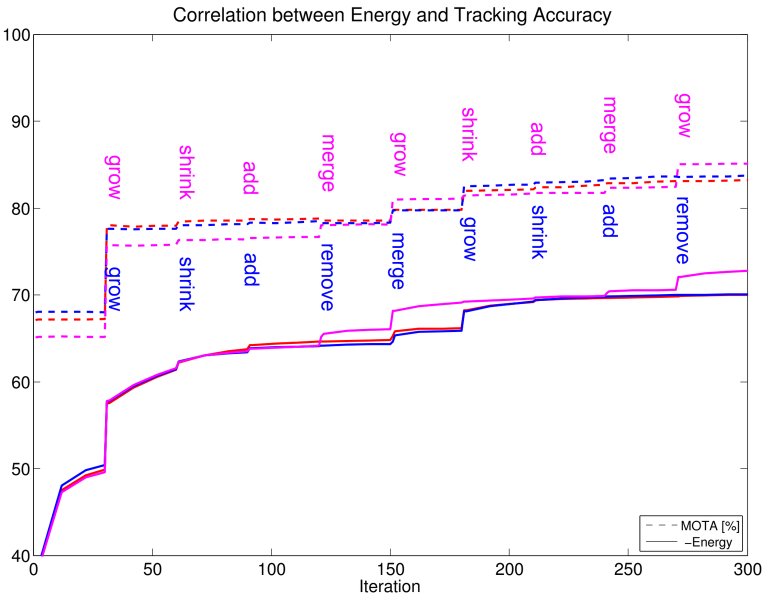


Initialization.

- Any tracker or zero-solution possible
- In our experiments: Extended Kalman Filter or Integer Linear Programming

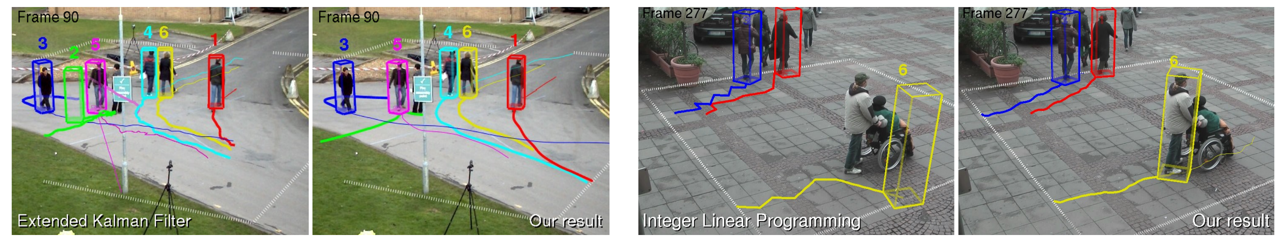
Experiments

Analyzing the objective.



Our energy function correlates well with tracking performance w.r.t. ground truth.

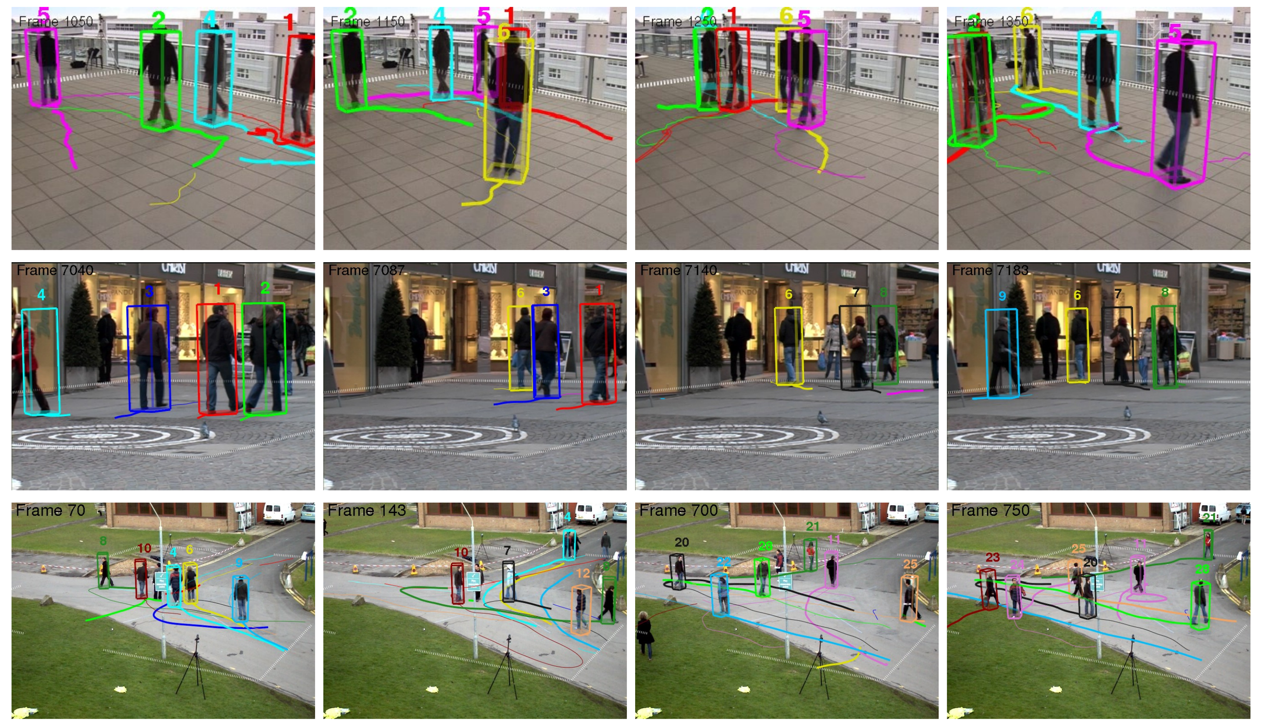
Comparison to baselines.



Quantitative Evaluation. Initial and final values for multiple optimization runs.

Sequence	MOTA [%]			MOTP [%]			MT			ID Switches		
	initial	*final*	**diff**	initial	*final*	**diff**	init	*fin*	**diff**	initial	*final*	**diff**
terrace1	82.8	*84.9*	**+2.1**	74.3	*79.6*	**+5.3**	7	*7*	**-0**	14.4	*19.6*	**+5.1**
terrace2	75.1	*83.8*	**+8.7**	71.7	*76.7*	**+5.1**	9	*7*	**-2**	11.7	*17.1*	**+5.4**
TUD-Stadtmitte	53.3	*60.9*	**+7.6**	57.4	*65.9*	**+8.4**	5	*6*	**+1**	3.8	*6.0*	**+2.2**
PETS'09 S2L1	64.7	*78.7*	**+14.0**	75.4	*76.7*	**+1.4**	9	*16*	**+7**	17.7	*14.2*	**-3.4**
mean	69.0	*77.1*	**+8.1**	69.7	*74.7*	**+5.1**	8	*9*	**+2**	11.9	*14.2*	**+2.3**

Example Results.



Conclusion

- Multi-target tracking can be formulated with continuous objective functions
- Meaningful (although local) energy minima can be achieved with gradient based optimization and trans-dimensional jump moves
- A hybrid optimization scheme yields state-of-the-art results on public datasets

Acknowledgements We thank Stefan Walk and Christian Wojek for providing their detection and tracking (EKF) implementation.