TOWARDS THE ROBUST VISUAL TRACKER

- Kalal et al. [2] presented Median-Flow tracker, where robustness is achieved by:
  1. Robust motion estimator — median
  2. Covering object of interest by local trackers [3] placed on a regular grid and reset to its default position each frame (Grid FoT)
  3. Local trackers are filtered by normalized cross-correlation and Forward-Backward procedure — slows down tracking by factor of two!

- Similar as Kalal et. al, we are using median statistics for robust translation and scale estimation, for tolerance of 50% data corruption (outliers) for translation and 100*(1-v0.5)=29% for scale.

- We redesign part (2) and present novel filters which explicitly model local trackers state (inlier/outlier) and take in to account group behaviour

LOCAL TRACKERS FILTERS IN CELL FOT

- We combine three types of information:
  a) Local appearance (normalized cross-correlation)
  b) Spatial consistency (similar to smoothness assumption used in optical flow estimation, denoted as Neighbourhood constraint)
  c) Temporal consistency (Markov Model)

- Together they form a very strong filter with negligible computational cost (less than 10%)

- Markov Model represents each local tracker as a two-state (i.e. Inlier, outlier) probabilistic automaton with transition probabilities \( p(i \mid s) \)

- Learning is done on-the-fly (the transition probabilities between states are learned, the “true” state of local trackers is obtained from the motion estimator)

- Markov Model can be extended by “forgetting”, which allows faster response to object appearance change

- For each frame local tracker probability of being an inlier is computed (local trackers are remove from estimation based on thresholding)

- Neighbourhood constant score is computed for each local tracker \( i \) as follows:

\[
A^N = \sum_{i \neq i} \Delta_{2D}(i) \quad \text{where} \quad \Delta_{2D}(i) = \begin{cases} 0 & \text{if expression is true} \\ 1 & \text{else} \end{cases}
\]

- Local trackers with score less than threshold (in experiments set to 1) are removed from motion estimation

RESULTS ON PUBLIC SEQUENCES

- Table shows number of “correctly tracked” frames in sequences (i.e. number of frames, where overlap with GT > 50%)

- Results of other algorithms were obtained from [2]

- Best achieved results for sequences are marked bold

Algorithms:


FUTURE WORK

- Tracker initialization by a single point instead of by a bounding box (by exploiting current object neighbourhood)

- Replace median translation-scale estimator by robust homography estimator (e.g. RANSAC)

- Address the problem of out-of-plane rotation

- Incorporate novel tracker to TLD framework

REFERENCES