Review of Dense Trajectory Features

Sourabh Daptardar

Action Recognition by Dense Trajectories [Weng et al CVPR 2011]

Earlier approaches to action recognition were based on the extension of 2D object recognition techniques to 3D. However, as two spatial dimensions and the temporal dimension have different properties – tracking features over temporal dimension i.e. trajectory features were found to be useful. This paper introduces a method of efficiently computing dense trajectories and making use of them to compute four descriptors: trajectory, HOG, HOF, MBH.

Dense Trajectories

- A dense grid of W pixel step size is sampled and each point is tracked to next frame using dense optical flow and median filtering for rounding off position (instead of bilinear interpolation). This is done independently for 8 scales, separated by a factor of $1/\sqrt{2}$.
- Drifting: problem of drifting is handled by restricting trajectory lengths to small sizes like 15
- Shi and Tomasi criterion for tracking features.
- Static trajectories are pruned.

Descriptors

1. **Trajectory:** Normalized displacements of the points along the trajectory. Encodes local motion patterns.
2. **Histogram of gradients:** HoG at spatio temporal grid of 32x2x3 at points along the trajectory. Encodes static appearance information.
3. **Histogram of flow :** Similar to 2, but encodes local motion information
4. **Motion boundary histogram:** Computes the gradient of the flow field. Constant motion is suppressed and changes in the flow field i.e. motion boundaries are captured.

Results

1. KTH : simplistic : on par with then state-of-the-art
2. Hollywood2 : high variability : outperforms then state-of-the-art
3. Youtube : camera motions : outperforms then state-of-the-art
4. UCF : camera motions, fast movements : comparable to then state-of-the-art

Action recognition with Improved Trajectories [Weng et al, ICCV 2013]

This work is continuation and improvement over the dense trajectory features. The major problem in action recognition on realistic videos is camera motions. Dense trajectory paper tried to address this by adding motion boundary histogram features, which are somewhat invariant to camera motion. The invariance to camera motion can be improved if we introduce camera motion estimation in our algorithm.

The paper does the following things to correct trajectories affected by camera motion:

1. Warp optical flow with robustly estimated homography

- Assume consecutive frames are related by a homography
- For purposes of homography estimation, use SURF features as they robust to motion blur
- Interest points at Corners (Shi-Tomasi) as well as blobs (SURF)
- Warp optical flow using the estimated homograpy
- HOF and MBF features should benefit from this
- Trajectories generated by camer motion can be removed, by thresholding displacement vectors in a warped flow field.

2. Remove inconsistent matches due to humans

- Apply state-of-the-art part based human detector by Prest et al
- To tackle misses by human detector track all bounding boxes obtained by human detector

Results

Fisher vector encoding is used instead of bag of words to improve classification. Beats state-of-the-art on Hollywood2, HMDB51, Olympic Sports, UCF50. The margin is high on Olympic sports and UCF50, where this approach seems to help.