Reading Notes

ORIGINAL

Orit Kliper-Gross, Yaron Gurovich, Tal Hassner and Lior Wolf. Motion Interchange Patterns for Action Recognition in Unconstrained Videos.

CATEGORIES

[**Computer Vision**]: Unconstrained Videos, MIP.

KEYWORDS

Motion Interchange Patterns, Action Recognition.

AUTHOR

Nino Lau, School of Data and Computer Science, Sun-Yet-Sen University.

1. BACKGROUND

Local Descriptors

First, the method seeks coordinates of space-time interest points. The local information around each such point in one of the frame is then represented using one of several feature point descriptors such as HOF, HOG. A video is then represented using, for example, a bag-of-features representation. Videos supplying either too few (e.g., videos of subtle motion) or too much motion (e.g., back- ground, textured motion such as waves in a swimming pool) may be inaccurate.

Optical-flow based methods

The method estimate the optical-flow between successive frames, sub-volumes of the whole video. Optical-flow, filtered or otherwise, provides a computationally efficient means of capturing the local dynamics in the scene. Optical-flow based methods commit early-on to a particular motion estimate at each pixel. Unreliable or wrong flow estimates would therefore provide misleading information to any subsequent processing.

Dynamic-texture representations

These methods evolved from techniques originally designed for recognizing textures in 2D images, by extending them to time-varying “dynamic textures”. This representations are always construct in a histogram form. And then, the whole 2D image is represented by the frequencies of these binary strings.

2. APPROACH

A video clip is encoded as a single vector. Given an input video, we are able to encode every pixel on every frame by eight strings of eight trinary digits, where each digit compares the compatibility of two motions with the local patch similarity pattern: one motion in a specific direction from the previous frame to the current frame, and one motion in a different direction from the current frame to the next one. Going in multiple directions, the methods provide a complete characterization of the change from one motion to the next. Another source of ambiguity stems from the camera motion which introduces image motion even in motionless parts of the scene. Attempts to employ independent video stabilization showed only little improvement for our descriptors. Notably, combination between conventional bag-of-words techniques and suitable learning techniques to obtain discriminative action models are made.

3. EVALUATION

Dataset and Metric

The effectiveness of such method is demonstrated on a variety of Action Recognition benchmarks, focusing on recently published benchmarks of challenging, “in-the-wild” videos. As mentioned earlier, the various parameters are fixed and no real effort was made to optimize them; however, to demonstrate the contribution of the various parts, they compare with partial variants of it.

Experiment

The experiments lists the resulting performance measures for several descriptors. We are surprised to find that the proposed Motion Interchange Patterns method considerably outperforms all single feature methods, and even outperforms the application of the elaborate OSSML algorithm to the three methods HOG, HOF, and HNF combined. Furthermore, MIP seems to be complimentary to HOG, HOF, and HNF, and combining MIP with these descriptors improves performance even further.

Comparison to previous results on the HMDB51 database and UCF50 database shows that our method significantly outperforms all results obtained by previous work.

4. COMMENTS

The modular structure of our system enables further improvements, such as the addition of learning-based hierarchical encoding layers, using recently proposed methods.

Also there exist some other descriptors which haven't been tested according to the experiments. So there performance and the hybrid performance still need to be explored.