Efficient Human Action Recognition by Cascaded Linear Classification

Mauthner T., Roth P. M., Bischof H., - Institute of Computer Graphics and Vision
{mauthner,pmroth,bischof}@icg.tugraz.at

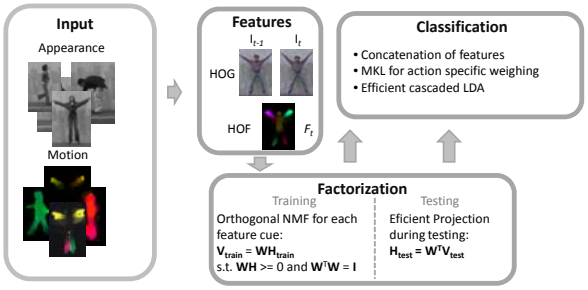
ICVSS 2010
Registration and Video Analysis

Abstract

We present a human action recognition system suitable for very short sequences. In particular, we estimate Histograms of Oriented Gradients (HOGs) for the current frame as well as the corresponding dense flow field estimated from two frames. The thus obtained descriptors are then efficiently represented by the coefficients of a Non-negative Matrix Factorization (NMF). To further speed up the overall process, we apply an efficient cascaded Linear Discriminant Analysis (CLDA) classifier. In the experimental results we show the benefits of the proposed approach on standard benchmark datasets as well as on more challenging and realistic videos. In addition, since other state-of-the-art methods apply weighting between different cues, we provide a detailed analysis of the importance of weighting for action recognition and show that weighting is not necessarily required for the given task.

Overview:

- Human action recognition on short sequences
- Various actions request different weighting
- Efficient classification for real-time performance



Classification on a two-frame-basis using appearance and motion information in parallel. Classical descriptors like Histogram of Oriented Gradients (HOG) and optical Flow (HOF) are combined with individual weighting during 1vsAll classification.

Orthogonal NMF (oNMF):

- Non-negative input features **V**
- Sparse basis vectors **W**, reduced dimensionality **H**

$$\mathbf{H} \leftarrow \mathbf{H}\frac{\mathbf{W}^T\mathbf{V}}{\mathbf{W}^T\mathbf{W}\mathbf{H}} \qquad \mathbf{W} \leftarrow \mathbf{W}\frac{\mathbf{V}\mathbf{H}^T}{\mathbf{W}\mathbf{H}\mathbf{V}^T\mathbf{W}} \qquad \mathbf{W}\mathbf{W}^T = \mathbf{I}$$

- Orthogonality constraint for basis **W** [Yoo08]
- Iterative solution **only** during training

Multiple Kernel Learning (MKL):

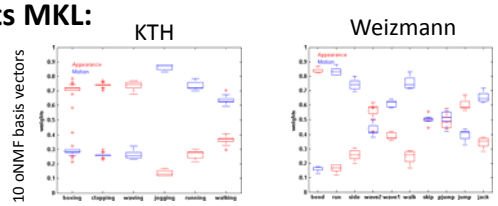
- Action specific weighting for appearance and motion
- Multiple Kernel Learning e.g. SimpleMKL by [Rako08]
-Convex combination of M basis kernels

$$\mathbf{K}(\mathbf{x},\mathbf{x}') = \sum_{j=1}^{M} d_j \mathbf{K}_j(\mathbf{x},\mathbf{x}')$$

- Update coefficients α_i and weights d_m during training

$$g(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i y_i \mathbf{K}(\mathbf{x}_i,\mathbf{x}) - b = \sum_{i=1}^{N} \alpha_i y_i \sum_{j=1}^{M} d_j \mathbf{K}_j(\mathbf{x}_i,\mathbf{x}') - b$$

Results MKL:

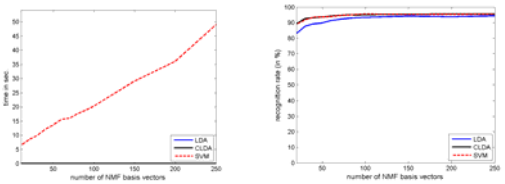


$$\mathbf{K}(\mathbf{x},\mathbf{x}') = d_{mot}\mathbf{K}_{mot}(\mathbf{x},\mathbf{x}') + d_{app}\mathbf{K}_{app}(\mathbf{x},\mathbf{x}')$$

- Trained weights for motion and appearance

Cascaded Linear Discriminant Analysis (cLDA):

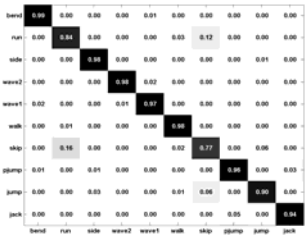
- First stage: projection matrix $\mathbf{P}_c$ containing all c classes
- Second stage: projection matrix $\mathbf{P}_{ij}$, nearest two classes i,j



Classification and timing results for the Weizmann human action dataset.

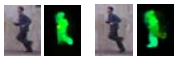
Results on Action Recognition Dataset:

Weizmann [Blank 05] : (*)9/10 actions performed by 9 persons

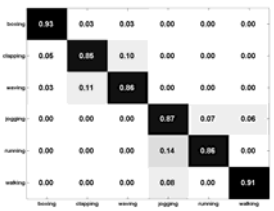


Method	rec.-rate	# frames
proposed *	97.0%	2
proposed	94.2%	2
[Schindler 08]*	93.5%	2
	99.6%	10
[Niebles 07]	55.0%	1
	72.8%	all
Blank et al. [ICCV'05]*	99.6%	all

run skip



KTH [Schueldt04] :



Subset d1

Limitations of classification on short sequences

Method	Dataset d1	d3	d4
proposed	88,1%	84,4%	88,4%
[Schindler 08]	90,9%	88,5%	92,2%

- 6 actions performed by 25 persons
- 4 different scenarios d1-d4

Referenzes:
[Rako08] A. Rakatomamonjy, F.R. Bach, S. Canu and Y. Grandvalet. SimpleMKL. JMLR,9:2491-2521,2008.
[Yoo08] J. Yooa and S. Choi. Orthogonal non-negative matrix factorization for clustering. In Proc. IDEAL, 140-147, 2008.
[Schindler08] K. Schindler and L v. Gool. Action snippets: How many frames are enough. In Proc. CVPR 2008.
[Blank 05] M. Blank, L. Gorelick, E. Shechtman, M. Irani and R. Basri. Actions as space-time volumes. In Proc. ICCV 2005.
[Schueldt04] C. Schüldt, I. Laptev and B. Caputo. Recognizing human actions: a local SVM approach, In Proc. ICPR 2004.