ACTION RECOGNITION USING SPATIO-TEMPORAL DIFFERENTIAL MOTION

Gaurav Yadav, Amit Sethi
gaurav.05ec@gmail.com, amitsethi@gmail.com
Indian Institute of Technology Guwahati

Motivation and Contributions

Motivation for this work:

- Human action recognition has become a popular research topic mainly because of its application in video classification and surveillance.
- In this paper, we focus on small datasets where deep learning methods fails to achieve good performance.
- We have developed a method for human action recognition based on the observation that information about an action is contained in differential motion between objects in space and time.

Our Contribution:

- Differential motion: Proposed a method for computation of differential motion. It captures the motion of the moving objects with respect to a potentially non-stationary back-ground very effectively.
- Differential motion maps: Proposed a feature representation of video based on the differential motion maps for classification of actions.
- Differential motion vs. optical flow: We experimented with both differential motion and optical flow, that is, any motion with respect to the camera. Feature representation was computed for both and it was found that differential motion gives better performance than optical flow.

Challenges

- There are various challenges in human action recognition such as variations in environments, viewpoints and actor movement.
- Variations in environments are caused by moving background, occlusion and addition of noise while capturing the video.
- Environment and recording settings also causes various types of noise in different lighting conditions.
- Inter-class and intra class variations.

Proposed Method

The proposed method has two major parts: feature extraction and classification as shown in Fig. 1.

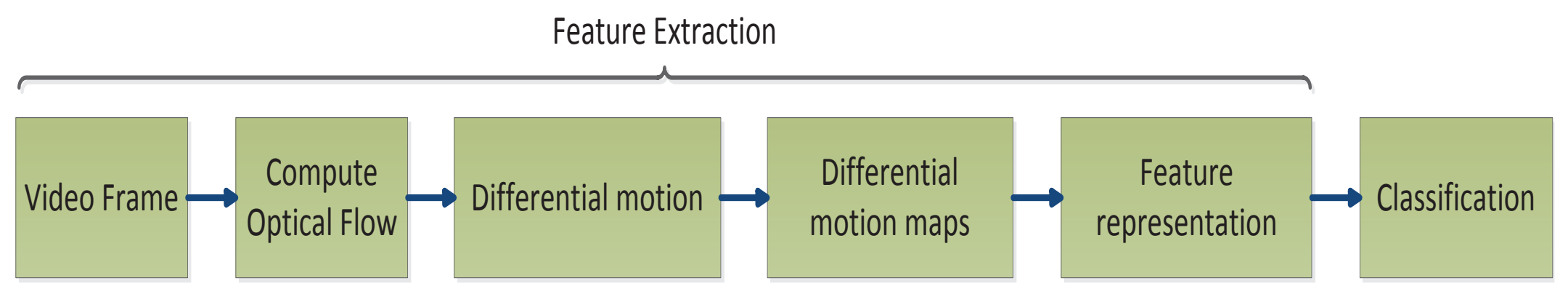


Figure 1: Block diagram of the proposed action recognition system.

Feature extraction:

1. To compute motion information we use optical flow based on Lucas-Kanade method.
2. Divergence is a differential operation on the vector field defined by flow map as follows:

$$R(x,y,t) = \nabla.\overrightarrow{V}(x,y,t) \tag{1}$$

$$\nabla.\overrightarrow{V}(x,y,t) = \frac{\partial V_1(x,y,t)}{\partial x} + \frac{\partial V_2(x,y,t)}{\partial y} \tag{2}$$

Differential Motion Maps

1. A point-wise absolute difference of divergence of pairs of consecutive frames is taken as shown in Equation 3 to capture the change in motion.

$$D_f(x,y,t) = |R(x,y,t+1) - R(x,y,t)| \tag{3}$$

2. To compress the information, we project the divergence map for each frame onto the three orthogonal Cartesian planes defined by (x,y,t) coordinates as shown in Fig. 2.
3. For the front view in which the dimension t is eliminated, this operation is defined as:

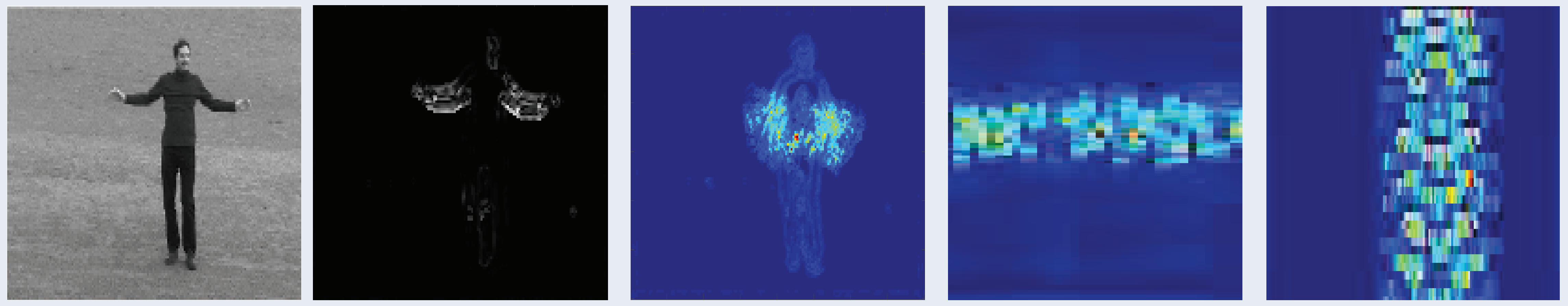
$$DMM_{front}(x,y) = \sum_{t=1}^{T-1} D_f(x,y,t) \tag{4}$$

Similarly for Side and top view:

$$DMM_{side}(y,t) = \sum_{x=1}^{m} D_f(x,y,t) \tag{5}$$

$$DMM_{top}(t,x) = \sum_{y=1}^{n} D_f(x,y,t) \tag{6}$$

Differential Motion Maps



(a) (b) (c) (d) (e)

Figure 2: (a) Example frames for hand-clapping video sequence, (b) divergence magnitude of optical-flow (c) Front view ($xy-$plane) differential motion map, (d) Side view ($yt-$plane) differential motion map, (e) Top view ($xt-$plane) differential motion map.

Classifier

For classification l_2-regularized collaborative classifier (LRCC) is used.

$$\overset{\wedge}{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \{\|\mathbf{g} - \mathbf{A}\boldsymbol{\alpha}\|_2^2 + \lambda \|\mathbf{L}\boldsymbol{\alpha}\|_2^2\} \tag{7}$$

where λ is regularization parameter and $\mathbf{L}$ is the Tikhonov regularization matrix. The class label of $\mathbf{g}$ can be obtained from equation 8 as follows:

$$class(\mathbf{g}) = \arg\min_{j}(\mathbf{e}_j) \tag{8}$$

where $\mathbf{e}_j = \left\|\mathbf{g} - \mathbf{A}_j \overset{\wedge}{\boldsymbol{\alpha}}_j\right\|_2$.

Experiment and Results

- KTH and UCF11 datasets were used for action recognition.
- Principal component analysis (PCA) was applied to reduce the dimensionality of the features.
- Table 1 shows the action recognition results compared to the state-of-the-art methods.
- Table 2 shows the effect of using differential motion over normal optical flow.

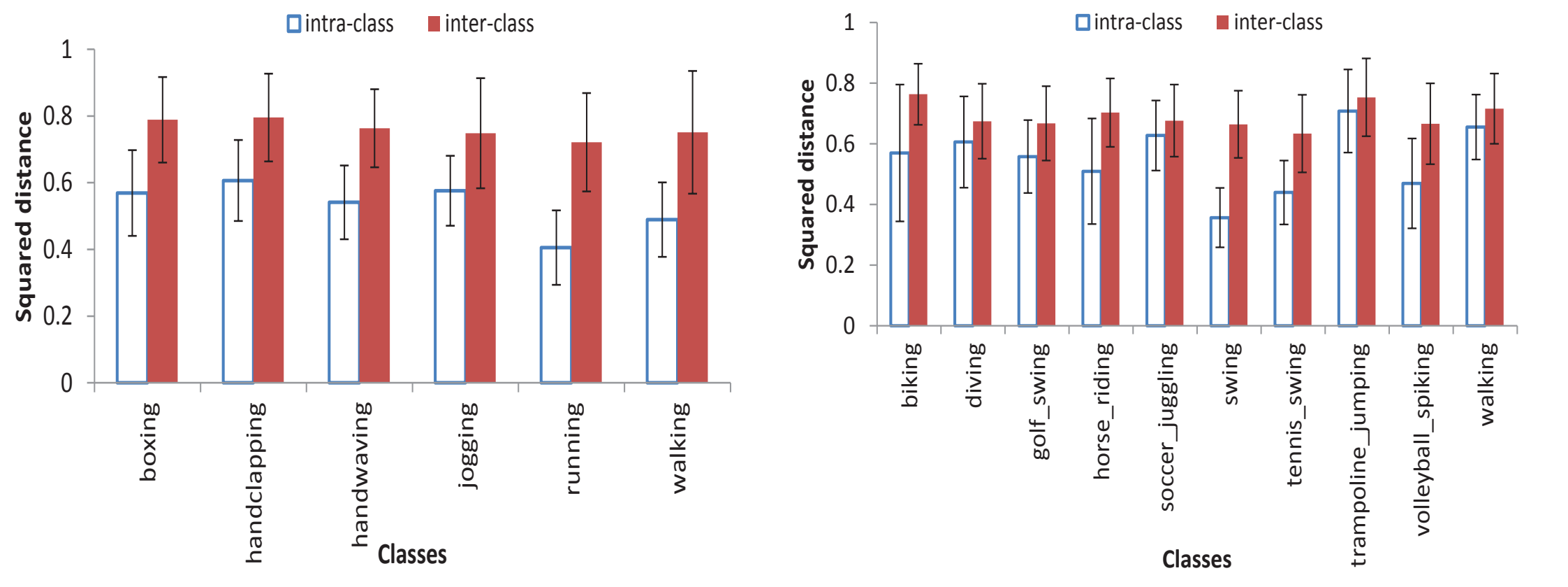
Table 1: Recognition accuracy for KTH and UCF11 datasets.

Method	KTH	UCF11
Proposed method	96.98%	90.24%
Yadav et al.	98.2%	91.3%
Kovashika et al.	94.53%	90.45%
Gilbert et al.	94.50%	–
Wang et al.	94.20%	84.20%
Laptev et al.	91.80%	–
Shuiwang et al. (CNN)	90.2%	–
Mahdyar et al. (CNN)	–	89.5%
kizler-Cinbis et al.	–	75.21%
Liu et al.	–	71.20%

Table 2: Comparison of optical flow and differential motion for KTH and UCF11 datasets.

Dataset	Optical flow	Differential motion
KTH	65.00%	96.98%
UCF11	44.10%	90.24%

- Class separation: To show that the proposed features are discriminative, mean-square distances between inter- and intra-class were computed as shown in Fig. 3.



(a) (b)

Figure 3: Inter- and intra-class mean squared distances (and their variances) for the proposed video representation for (a) KTH and (b) UCF11 datasets.

Conclusion

- We proposed a feature representation based on differential motion map for action recognition.
- Differential motion captures the motion information very effectively and shows better performance compared to state-of-the-art methods.
- Differential motion maps capture the action structure as well as motion.
- A comparison of differential motion and optical flow based feature was also done to show that differential motion gives better feature representation.