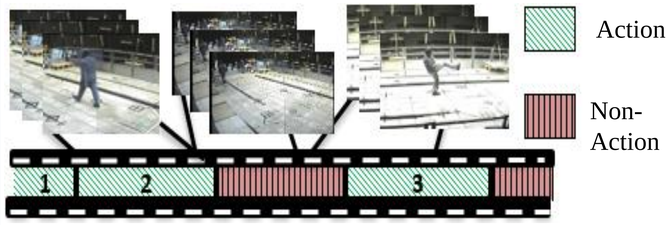
PMHI: Proposals from Motion History Images for Temporal Segmentation of Long Uncut Videos

Fiza Murtaza[1] Muhammad Haroon Yousaf MIEEE[1] Sergio A. Velastin SMIEEE[2,3,4]

[1] Univ. of Engg. & Tech. Taxila, Pakistan [2] Univ. Carlos III de Madrid, Spain [3] Cortexica Vision Systems Ltd., UK [4] Queen Mary Univ. of London, UK

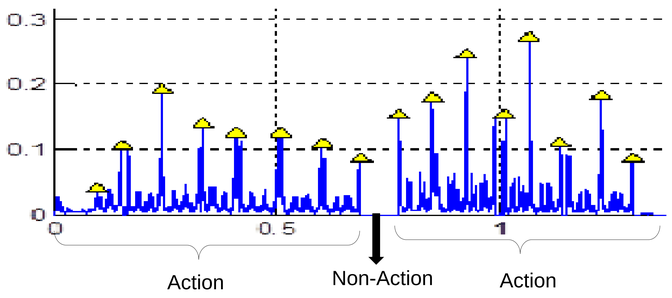
Goal

- To generate temporal action proposals for the segmentation of long uncut videos.



Overview

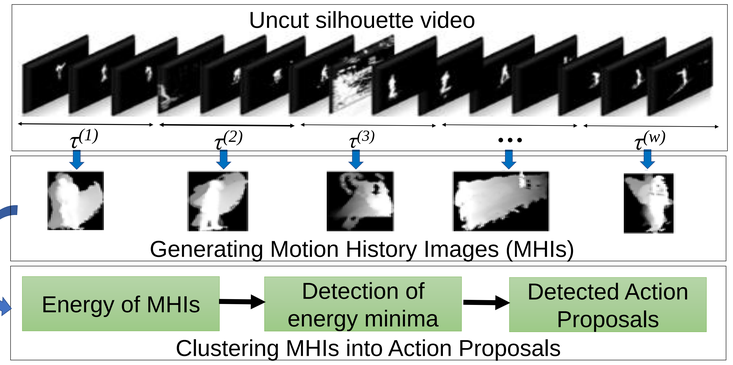
- Energy of consecutive nonoverlapping Motion History Images (MHIs), provides spatiotemporal information of motion.
- Non-action segments can be detected by finding minima from the energy of MHIs as shown below:



Energy of consecutive MHIs

Contribution

- Propose a method, Proposals from Motion History Images (PMHI), which generates the temporal action proposals in long duration uncut videos.
- Propose a clustering algorithm to segment the MHIs into actions and nonaction segments.
- PMHI is unsupervised; hence, it does not require prior training.
- PMHI outperforms the recall rate of recent methods on the MuHAVi-uncut dataset as well as the CVPR 2012 Change Detection dataset (CCD).



Algorithm to find MHI_k

Input: Silhouette frames $I(x,y,t)$, τ is the size of window
Output: MHI_k for all temporal windows
Procedure:

1. **for** $k = 1{:}w$ **do** % w is the total no. of temporal windows
2. **for** $t = 1{:}\tau$ **do**
3. $M(x,y,t) = \begin{cases} \tau & if\ I(x,y,t) = 1 \\ \max(0, M(x,y,t-1) - 1) & otherwise \end{cases}$
4. **end for**
5. $MHI_k = M(x,y,\tau)$ % after above loop
6. **end for**

Approach

Energy of MHI_k

$$E_k = \sum_{x,y} MHI_k(xy)$$

Algorithm to find the temporal locations A of action regions

Input: $E_{min} = 0,\ E'_k, r, w, G = [\,]$
Output: A
Procedure:

1: **while** $R > r$ **do** % r is threshold value
2: **for** $k = 1{:}w$ **do**
3: $G = \begin{cases} \{[G|k], if\ E'_k \le E_{min} \\ G \quad otherwise \end{cases}$ % G represents non-action locations
4: **end for**
5: $E_{min} = E_{min} + 0.01$ % 0.01 is the step size
6: $R = card(G)/w$ % *card(.)* finds the length
7: **end while**
8: $A = comp(G, [1{:}w])$ % comp(.) finds the complement

Clustering temporal action locations into proposals P

$$P_a = \begin{cases} [P_a|A(i)]\ if\ A(i+1) - A(i) = 1 \\ a = a + 1 \quad otherwise \end{cases}$$ where i=1:length of A

Results

Datasets:
MuHAVi-uncut: Untrimmed videos from 8 cameras. **CCD:** Untrimmed videos from 5 different scenarios.

Comparison with Action Proposals from dense Trajectories (APT) [12]

	Recall	Precision	Recall	Precision
	PMHI (our)		APT [12]	
Video Name	MuHAVi- uncut dataset			
C1:Camera1	94.1	80.0	50.0	53.0
C2:Camera2	53.0	50.0	100	54.4
C3:Camera3	94.1	94.1	90.0	56.0
C4:Camera4	88.2	62.5	29.4	52.0
C5:Camera5	94.1	67.0	43.1	30.3
C6:Camera6	71.0	67.0	70.0	44.1
C7:Camera7	94.1	76.2	50.0	39.0
C8:Camera8	100	100	50.0	53.0
Average	**86.1**	**74.6**	60.3	48.0

	Recall	Precision	Recall	Precision
	PMHI (our)		APT [12]	
Video Name	CCD dataset			
V1:corridor	80	66.7	20.0	35.7
V2:diningRoom	100	50.0	100	90.0
V3:lakeSide	100	100	33.3	85.0
V4:library	100	100	100	100
V5:park	50	100	33.3	100
Average	**86.0**	**83.3**	57.3	82.1

PMHI outperformed APT for both datasets.

Conclusion

- PMHI segments the uncut video by producing non-overlapping action proposals.
- It is unsupervised and hence it saves time for long and complex videos.
- Results show that detection of Energy minima from the Energy of MHIs can discriminate between actions and non-action regions accurately.