



Motion Interchange Patterns for Action Recognition in Unconstrained Videos

Orit Kliper-Gross, Yaron Gurovich, Tal Hassner, Lior Wolf



Weizmann Institute
of Science



The Open University
of Israel



Tel Aviv University

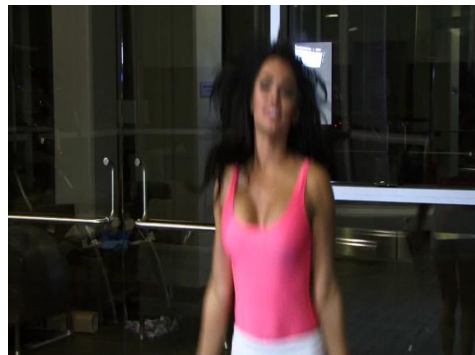
<http://www.openu.ac.il/home/hassner/projects/MIP/>

What is an Unconstrained Video?

Controlled Sets →



“In The Wild” Sets →



Challenges

Large
variability



Camera
Motion



Action
Ambiguity

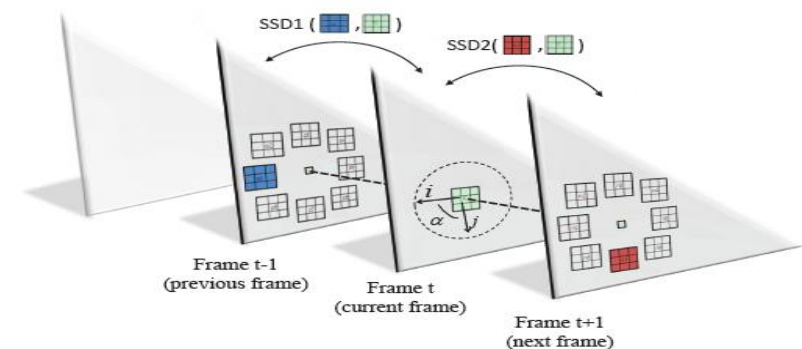


Others



Motion Interchange Patterns (MIP) - Highlights

- **A new video descriptor:**
 - Dense Characterization of motion changes
 - Captures shape of moving edges
 - Built-in stabilization mechanism
- State-of-the-art performance in the most recent and challenging benchmarks (ASLAN, HMDB51, UCF50, ...)



Action Recognition

Common Pipeline

Input
Videos

STIP +
Local Desc

STIP: On Space-Time Interest Points
Laptev IJCV'05

Local Descriptors: HOG, HOF and HNF
Learning realistic human actions from movies
Laptev et.al. CVPR'08



Action Recognition

Common Pipeline

Input
Videos

STIP +
Local Desc

Global
Descriptors

Bag of Words (BOW)

Following:

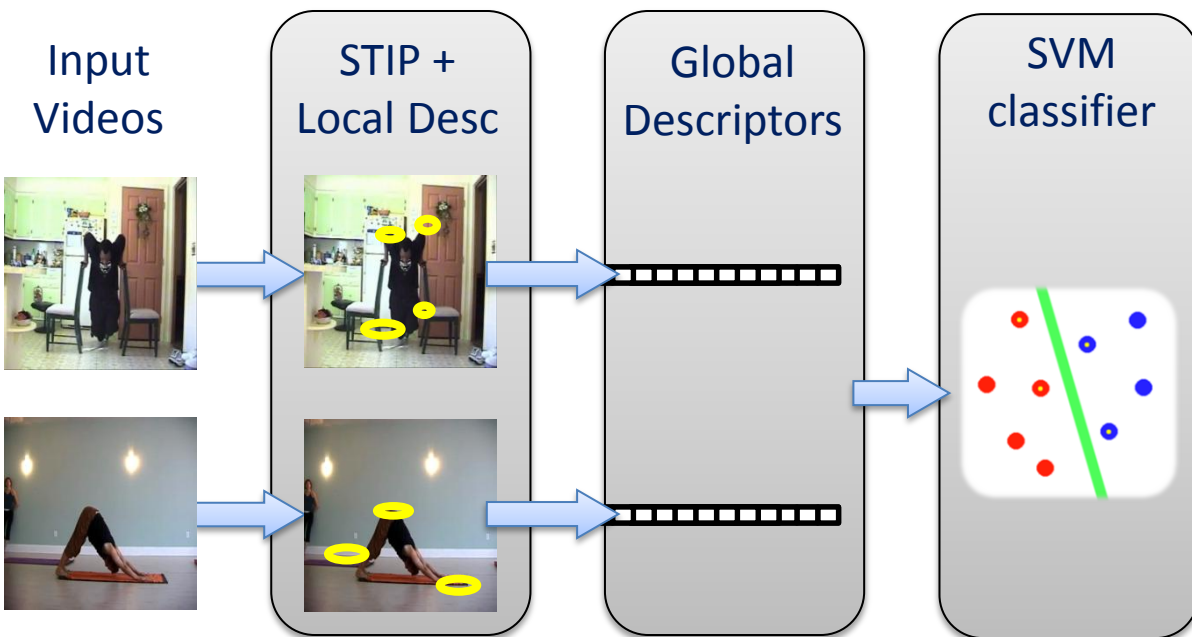
Learning realistic human actions
from movies

Laptev et.al. CVPR'08



Action Recognition

Common Pipeline



New Video Descriptor

Motion Interchange Patterns (MIP)

New Video Descriptor

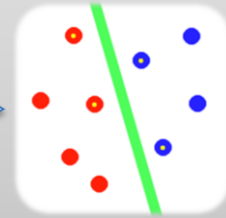
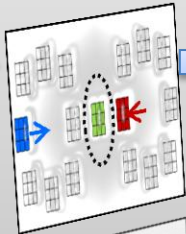
ECCV'12

Input
Video

LBP-
based
Methods

Global
Descriptors

SVM
classifier

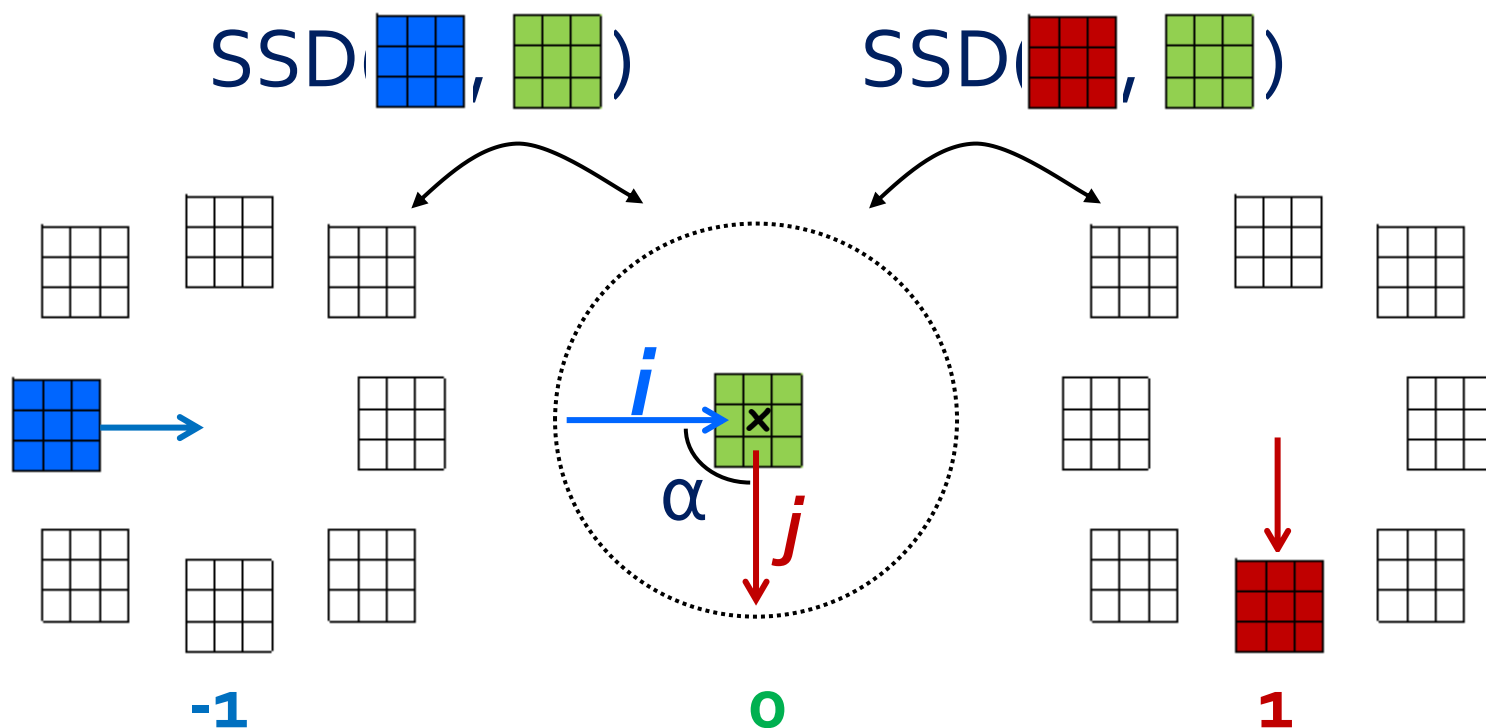


Local Binary Patterns (LBP) -based representations

- What:
 - Low-level, dense, local representation
- How:
 - Per-pixel encoding
 - Uses binary/trinary digits 0 1 / -1 0 1
 - The descriptor: frequencies of binary/trinary strings
- Very Successful:
 - Image textures [*Ojala et al. '96, Ojala et al. '02, Heikkila et al. '06*]
 - Face recognition [*Ahonen et al. '06, Zhang et al. '07, Wolf et al. '08*]
 - Facial expression [*Zhao and Pietikainen '07*]
 - Action recognition [*Yang et al. '07, Kellokumpu et al. '08, Yeffet & Wolf '09*]

Motion Interchange Patterns (MIP)

ECCV'12



Motion Interchange Patterns (MIP)

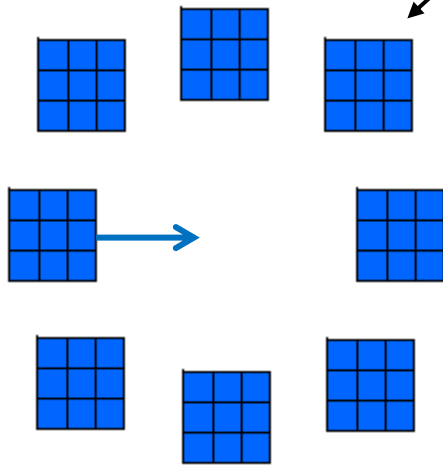
ECCV'12

$$\alpha = 0$$

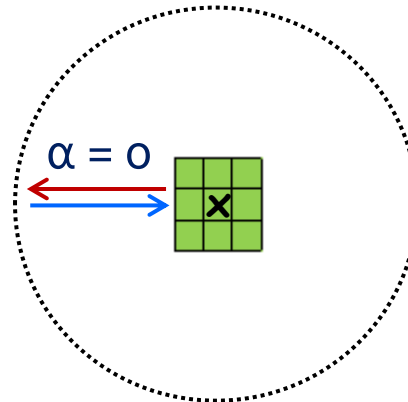


$SSD(\text{blue grid}, \text{green grid})$

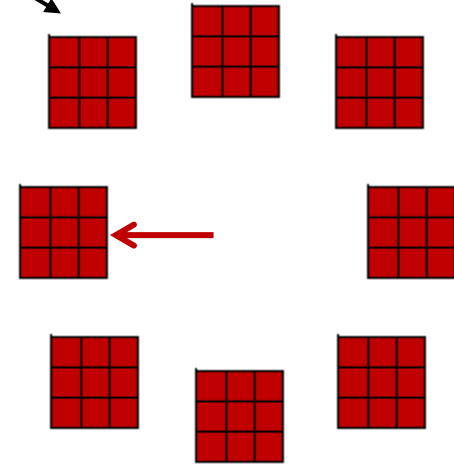
$SSD(\text{red grid}, \text{green grid})$



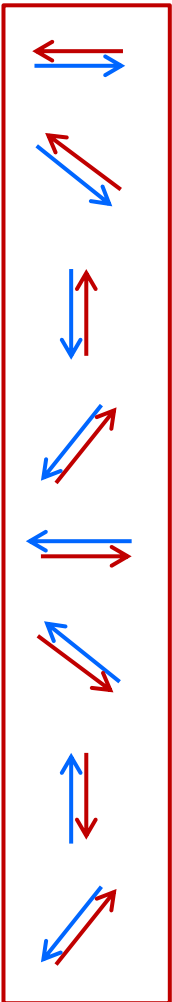
-1



0



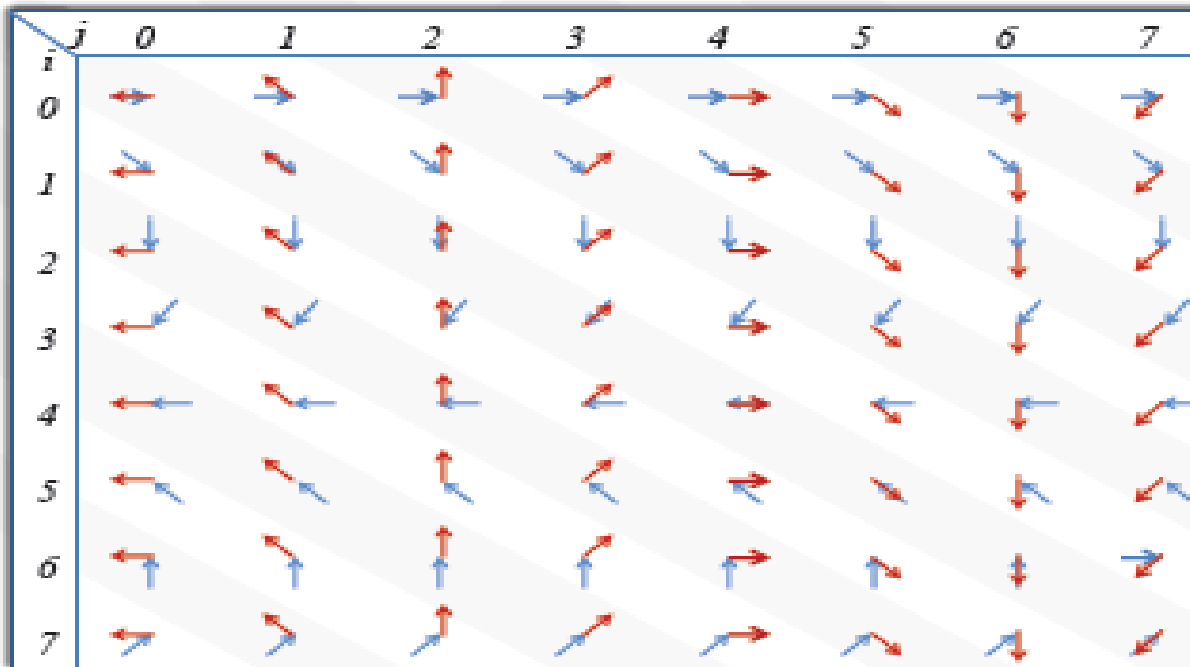
1



Motion Interchange Patterns (MIP)

64-digits trinary code

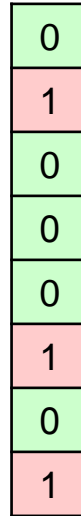
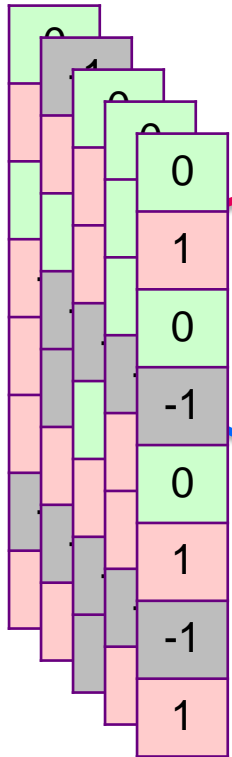
different α = different channels = diagonals



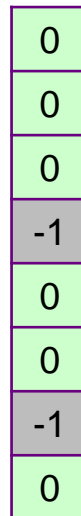
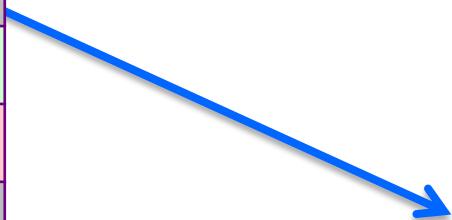
Motion Interchange Patterns (MIP)

Each α defines a channel \rightarrow 8 channels

Per-pixel 64-digits
trinary code



0-255 integer



0-255 integer

2 integers per-pixel
Per Channel

Motion Interchange Patterns (MIP)

An example - one channel basic coding

- Vote for next frame
- Vote for prev frame
- Static edges

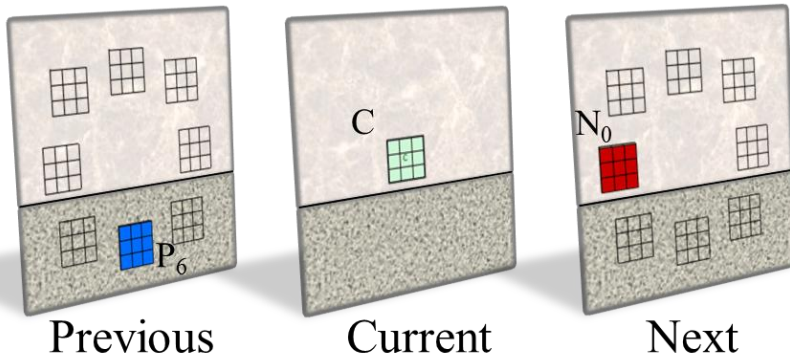


MIP captures:
Motion, Motion Changes, and Shape

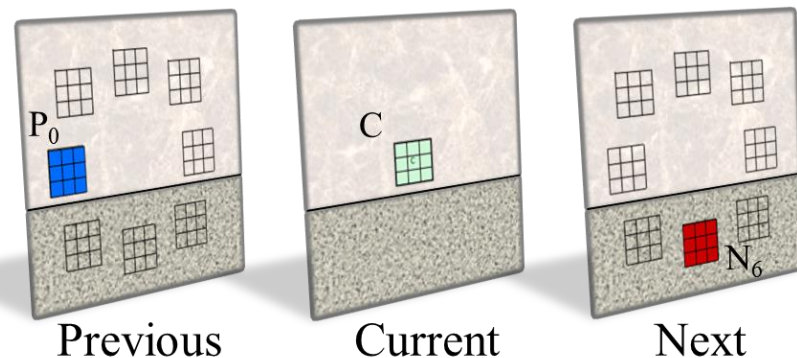
Suppression Mechanism

Suppress background structure and noise

Original Coding = 1



Switched Locations Coding = -1



Switched Patch Suppression

2 ways to look at this:

- No motion.
- Contradicted motion voting.
i.e.

Original coding voted down ←

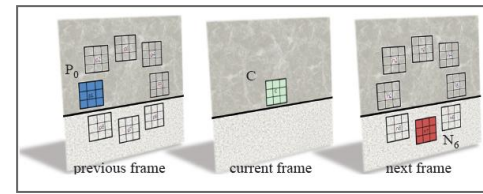
Switched patches voted up →



Suppress the code

MIP Suppression Mechanism

An Example



Without
Suppression



Original

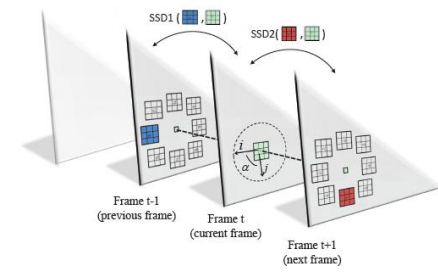


With
Suppression



Effect of Camera Motion

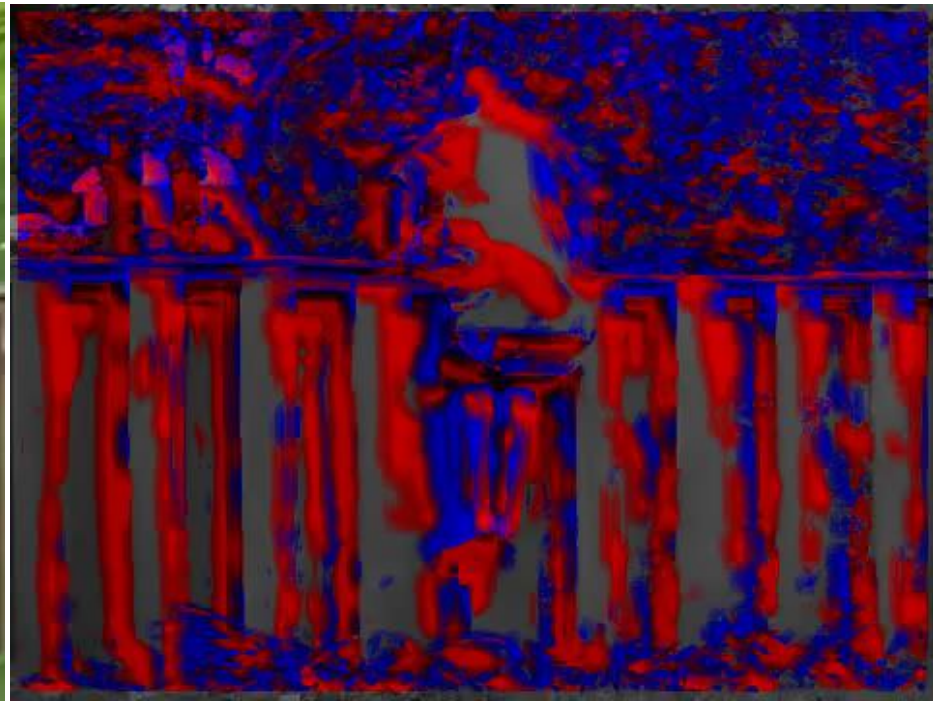
Motivating Example



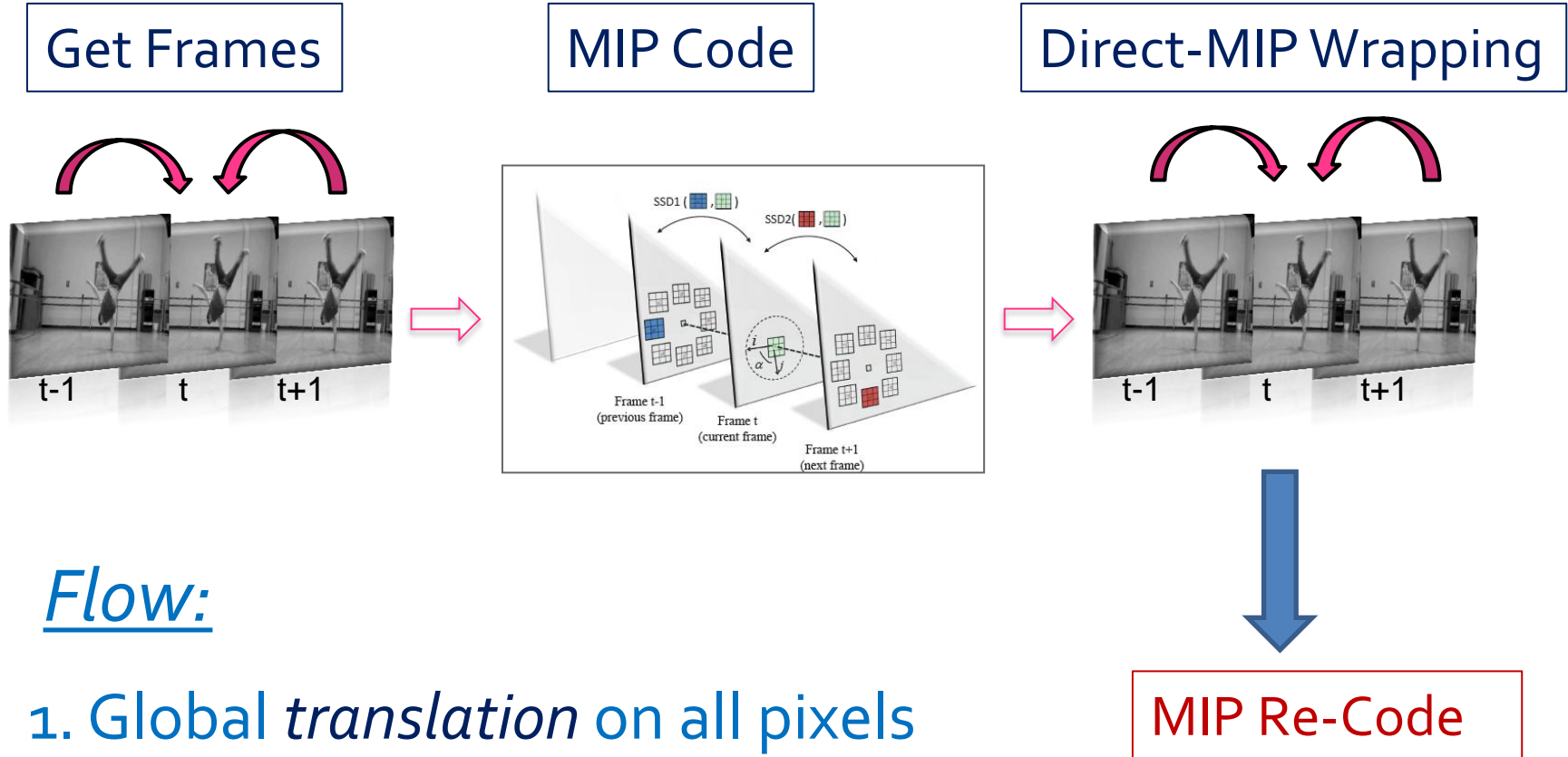
Original Movie



MIP Coding



MIP Stabilization Mechanism

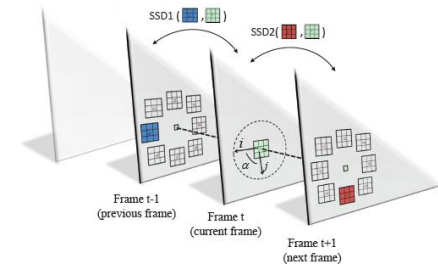


Flow:

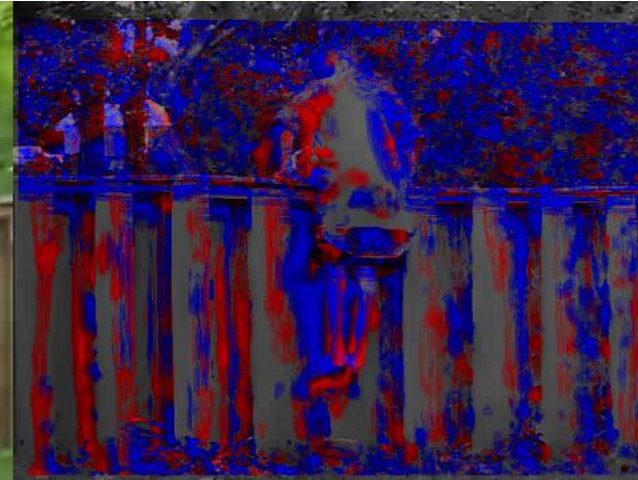
1. Global *translation* on all pixels
2. Code MIP.
3. Use **MIP** silent pixels for global *affine*

MIP Stabilization Mechanism

An Example



Without
Stabilization

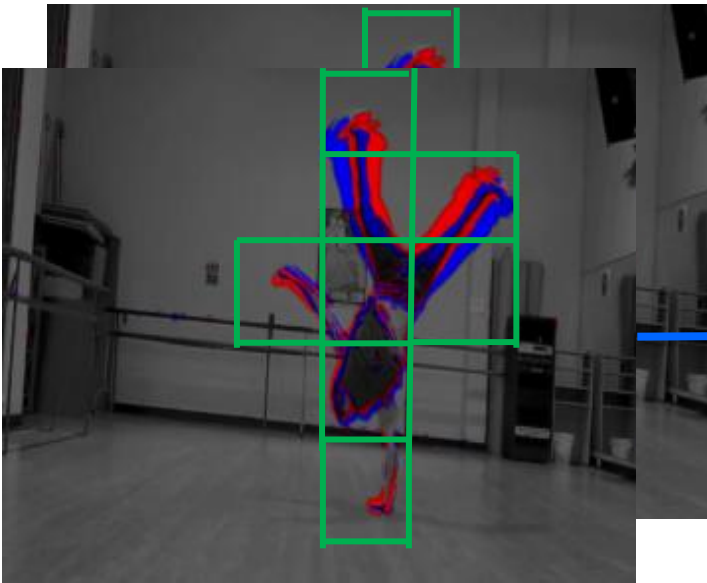


Original

With
Stabilization

Motion Interchange Patterns (MIP) Vectorization

16x16 patches



Two 256-bins
histograms
per *Patch*

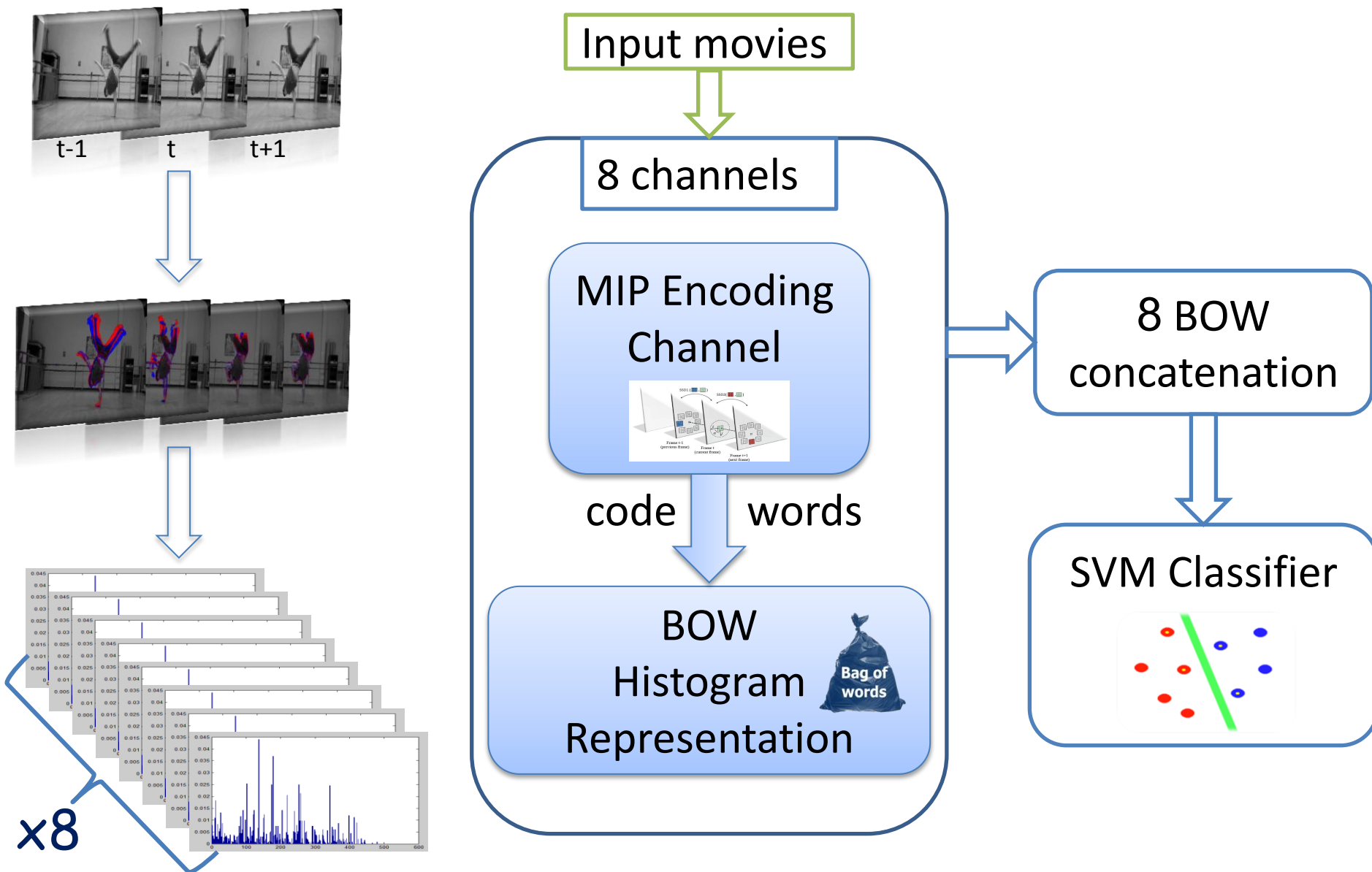
Concatenate:
512-bins histogram
per *Patch*



Vectorization: 512-dimensions *code words*

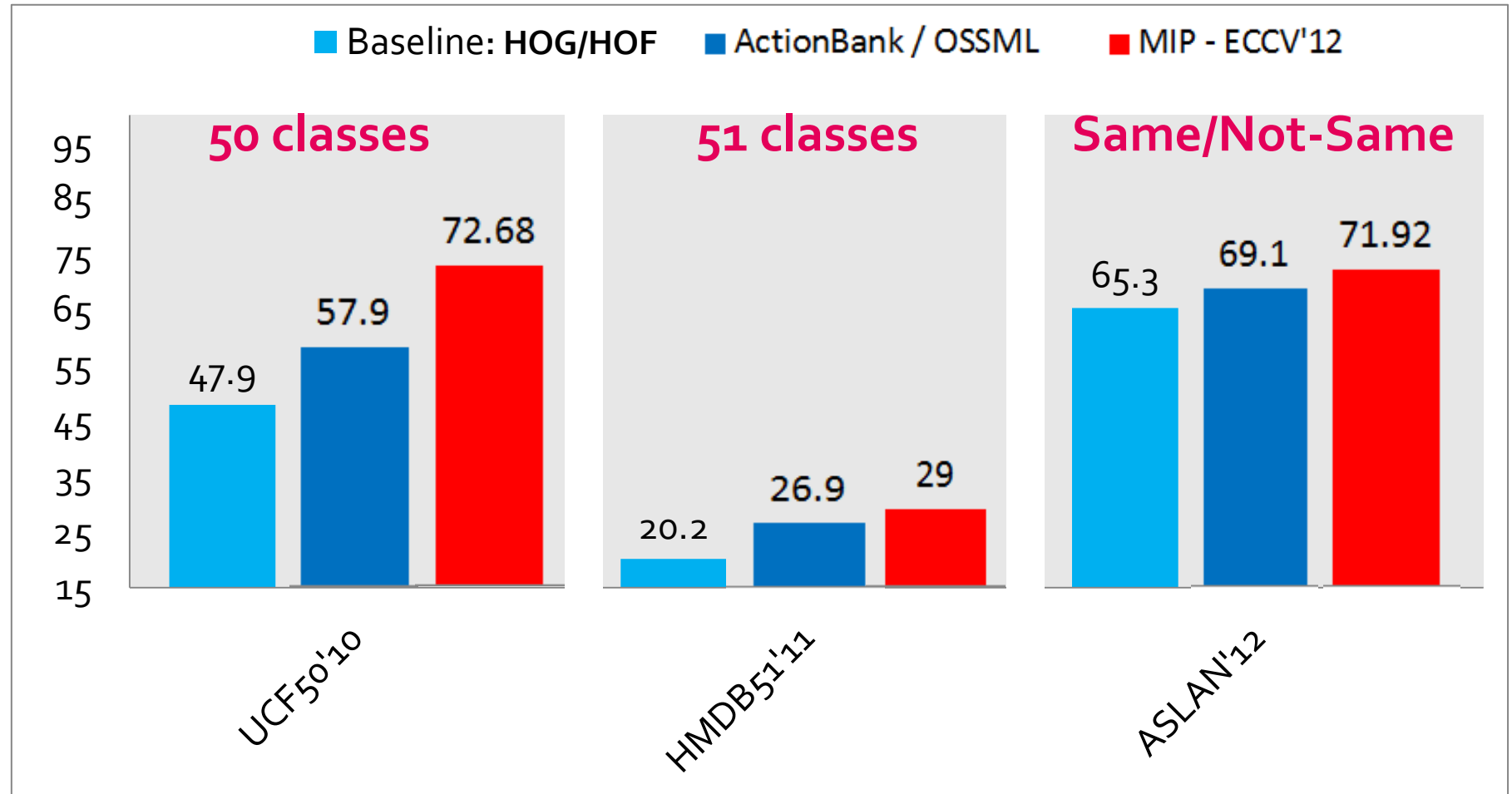


Motion Interchange Patterns – pipeline



MIP on Most Challenging AR Datasets

Performances



SotA on the most challenging Action Recognition DBs

Motion Interchange Patterns (MIP)

Examples

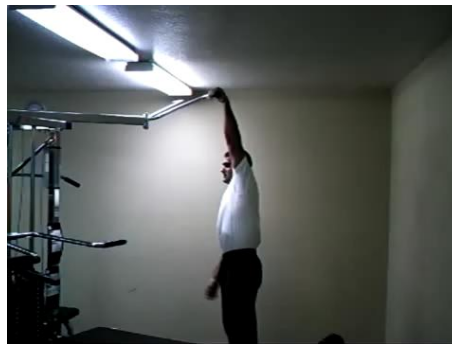
Results on ASLAN

Same classified as Same (TP) 😊

Jumping Jacks



Pull Ups – 1 hand



Moon Walk



Results on ASLAN

Same classified as Not-Same (FN) ☹️

Kissing



Squat



Talking on phone



Results on ASLAN

Not-Same classified as Not-Same (TN) 😊



Results on ASLAN

Not-Same classified as Same (FP) 😞



Results on HMDB51

Brush hair success :



False positive, miss of 'chew':



Results on HMDB51

Cartwheel success :



False positive, miss of 'flic flac':



Results on UCF50

Basketball success:



False positive, miss of 'Volleyball Spiking':



Results on UCF50

HighJump success :



False positive, miss of 'Pole-vault':



Results on UCF50

Nunchucks success :



False positive, miss of 'Pizza Tossing':



Summary

A New Video Descriptor:

- Efficient Low-level, dense, local representation
- Complete characterization of motion & motion changes
- Captures shape of moving edges
- Built-in suppression & stabilization mechanisms

Thank You !

<http://www.openu.ac.il/home/hassner/projects/MIP/>