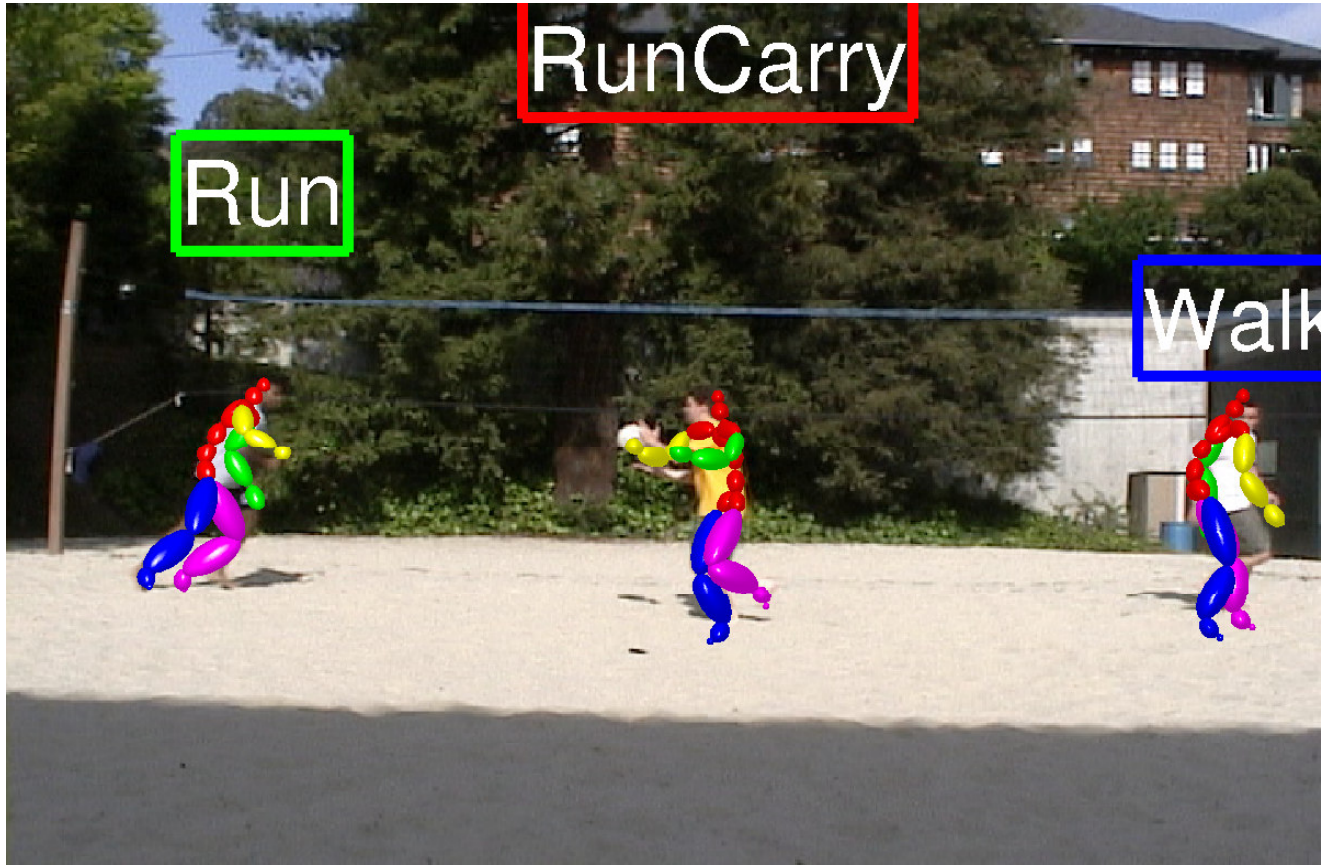
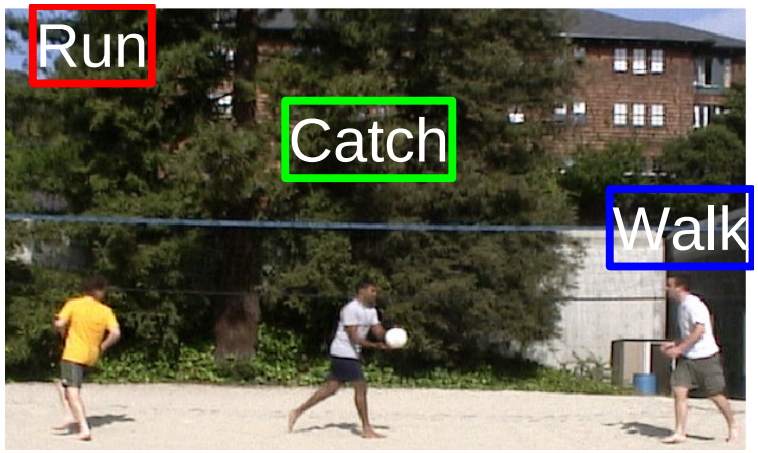
Looking at People

Deva Ramanan UC Berkeley
David Forsyth UIUC/UC Berkeley
Andrew Zisserman Oxford



Why?

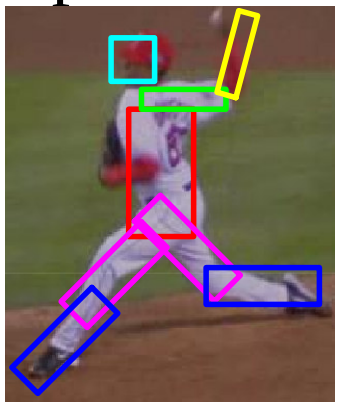
surveillance



interfaces

video data mining

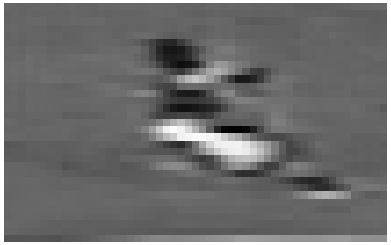
motion capture



Basic Approaches

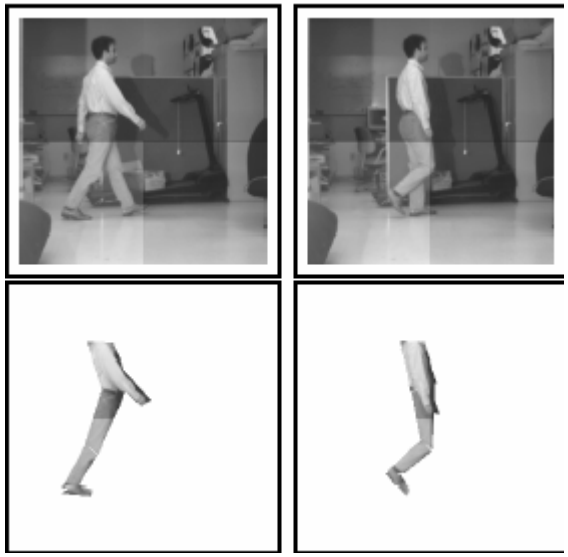
- Model-Free
  - annote video directly using local features

    (Efros et al. Cutler & Davis. Bobick. Nelnik Manor & Irani)



- Skeleton Model
  - track skeleton and then recognize pose

    (Yacoob & Black)



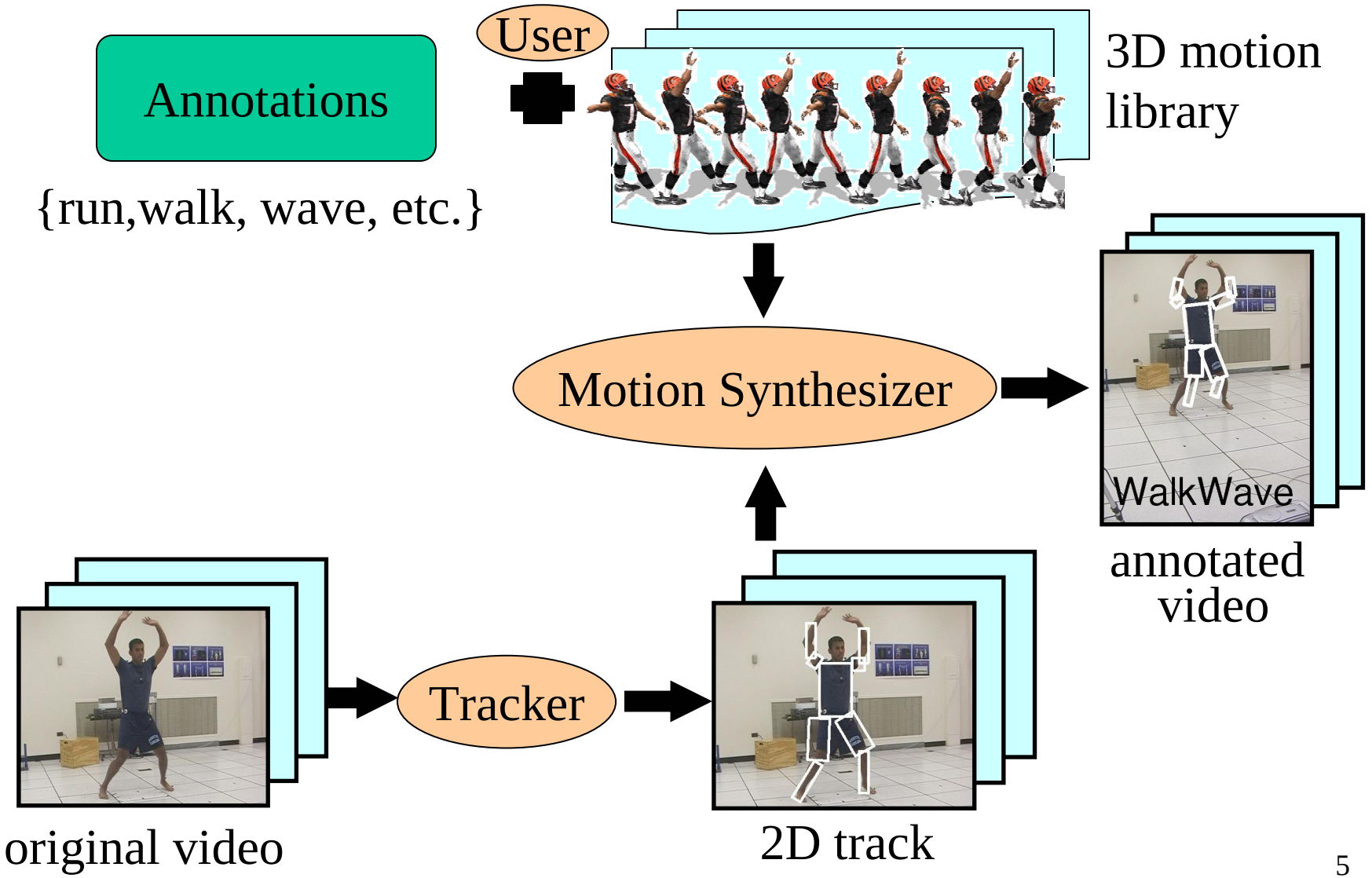
Common Issue: No vocab

Bobick & Davis. PAMI01

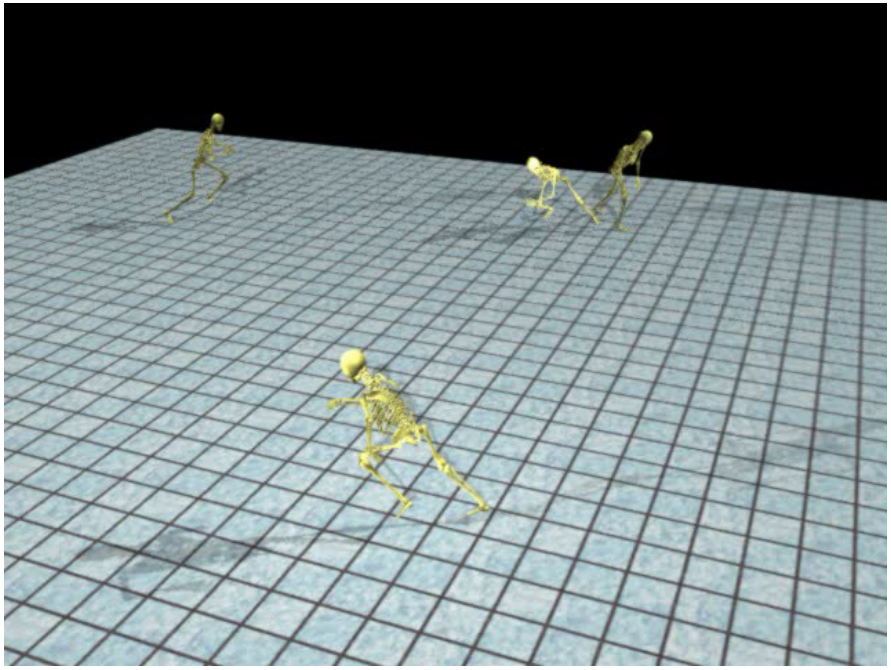
What's right annotation?

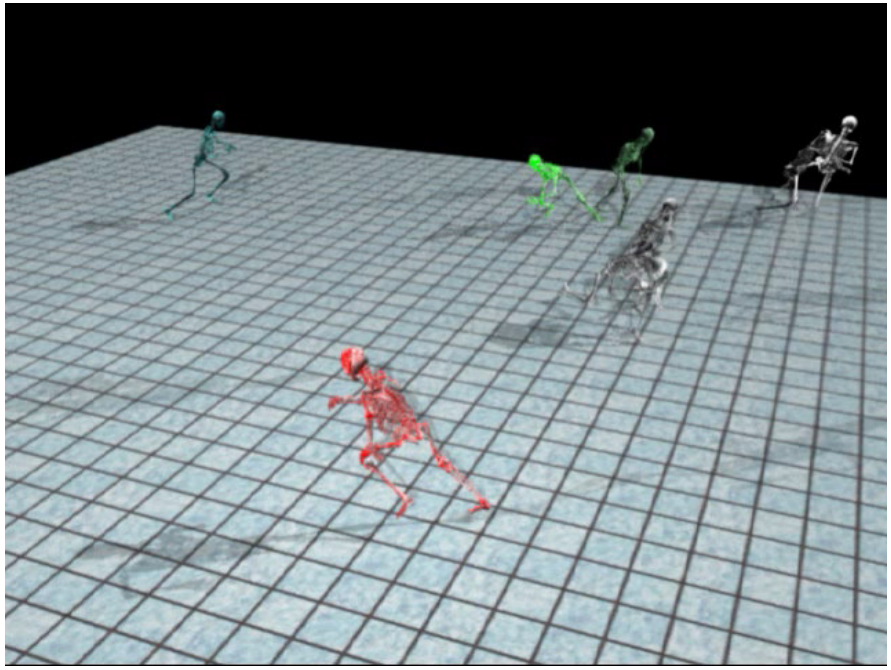
Our approach: Separate out vocab by thinking of **analysis as synthesis**

System Model



Motion Synthesis Under Constraints

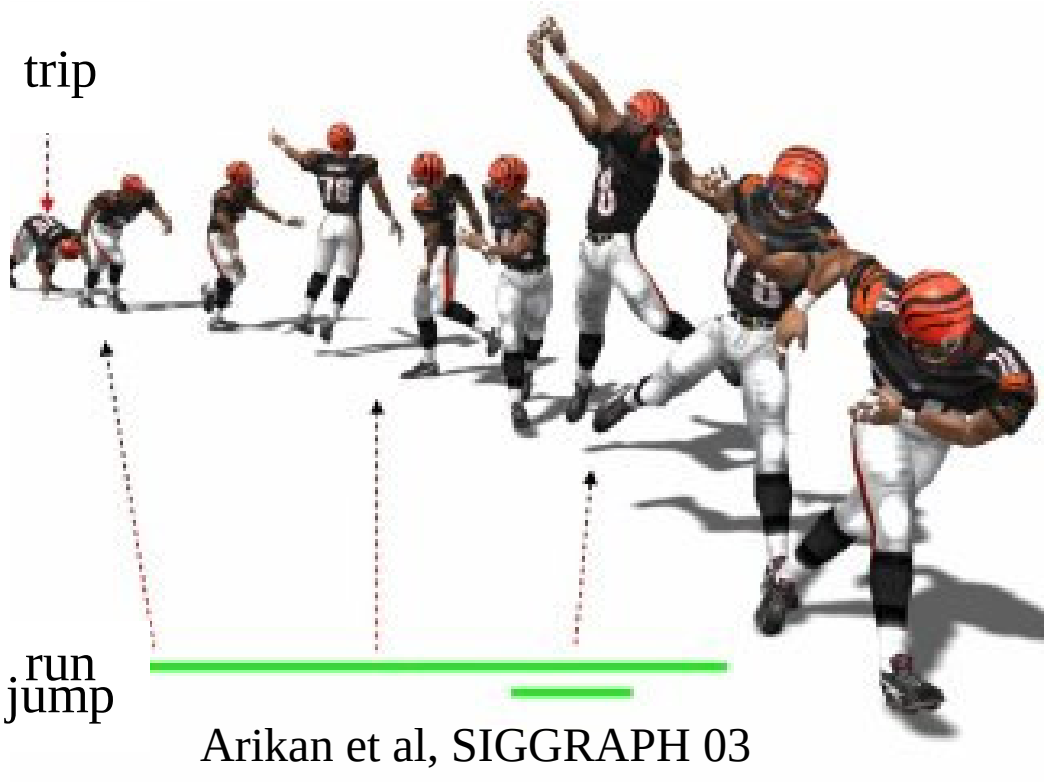




Satisfy pose constraints by
cutting & pasting existing clips of motion

Arikan et al. SIGGRAPH03
Arikan & Forsyth. SIGGRAPH02
Kovar et al. SIGGRAPH02
Lee at al. SIGGRAPH02

Motion Synthesis from Annotations



Arikan et al, SIGGRAPH 03

- synthesize by dynamic programming
- what's right annotation vocabulary?

Annotation Vocabulary

pick a set useful for synthesis

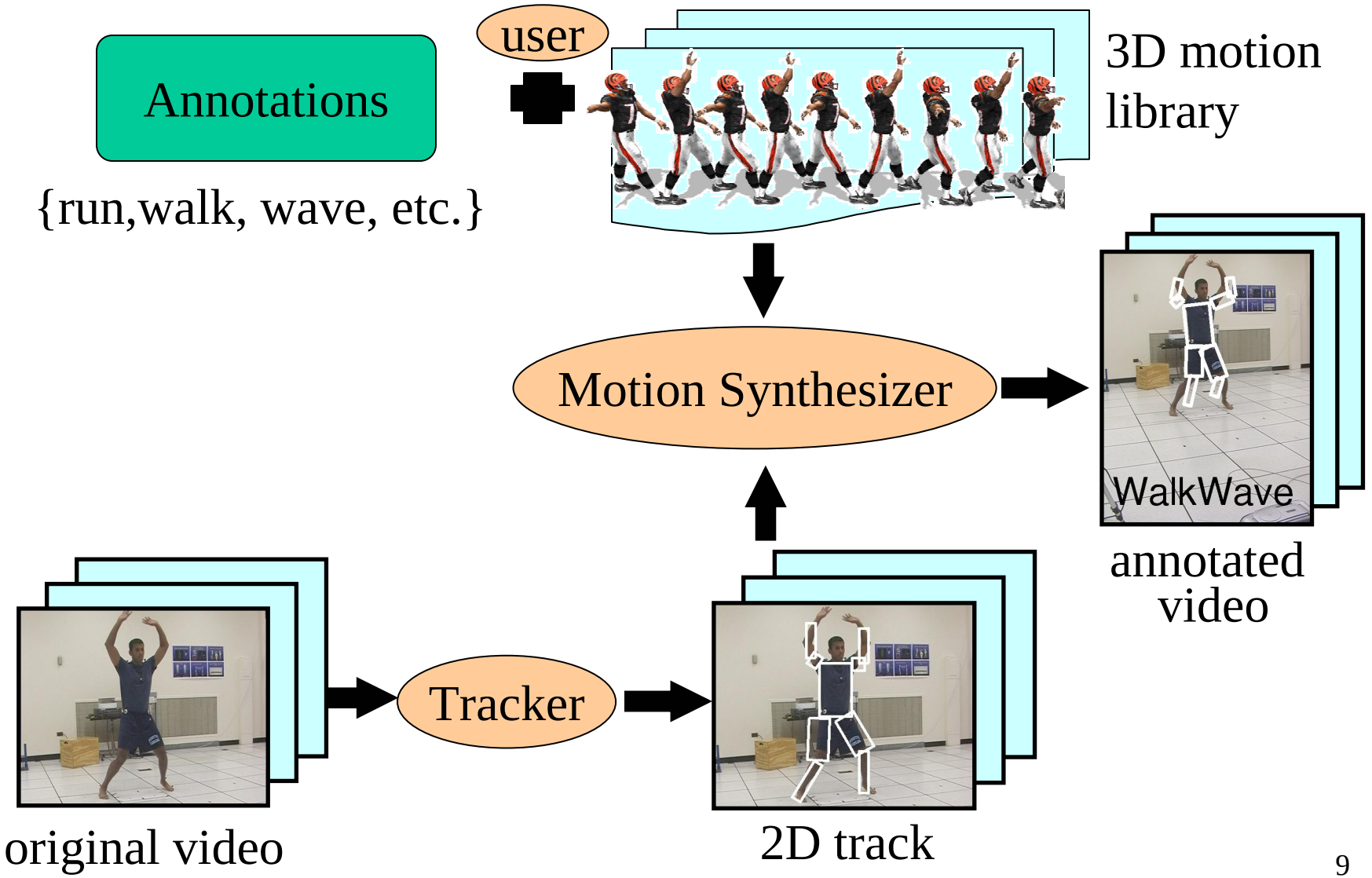
run	0
walk	1
wave	1
jump	0
turn left	1
turn right	0
catch	0
reach	0
carry	0
backwards	0
crouch	0
stand	0
pick up	0

-flags can be **composed**

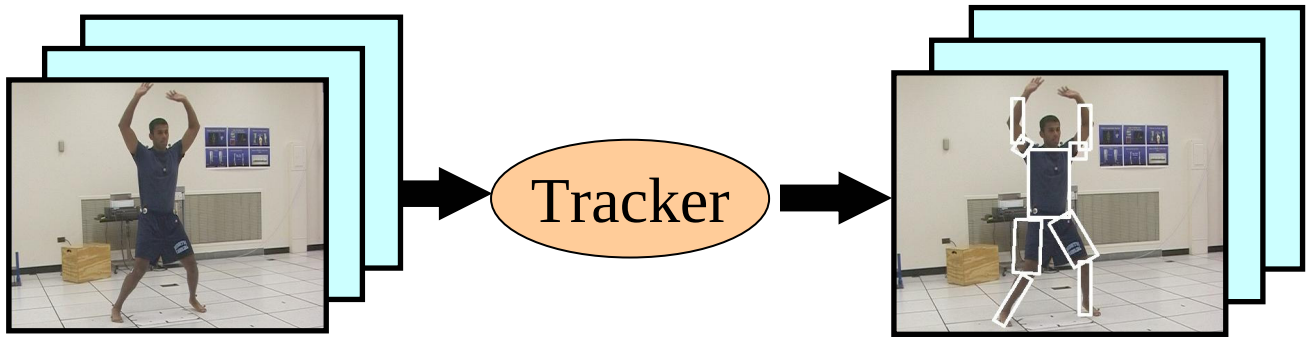
-2^{13} = 8192 annotations

-not all combos are valid

System Model



Tracking Algorithm



original video

2D track

Why is tracking hard?

fast

unpredictable

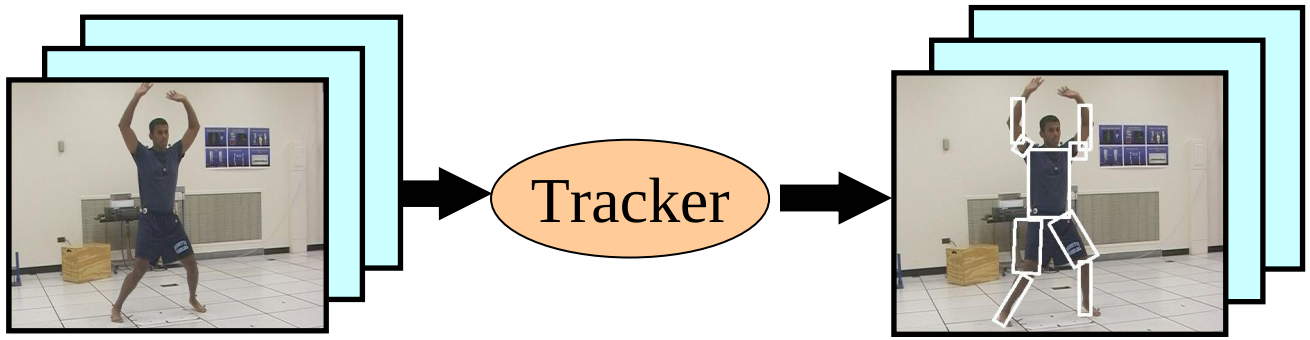
hard to track

variety of poses

interactions & occlusions

hard to detect

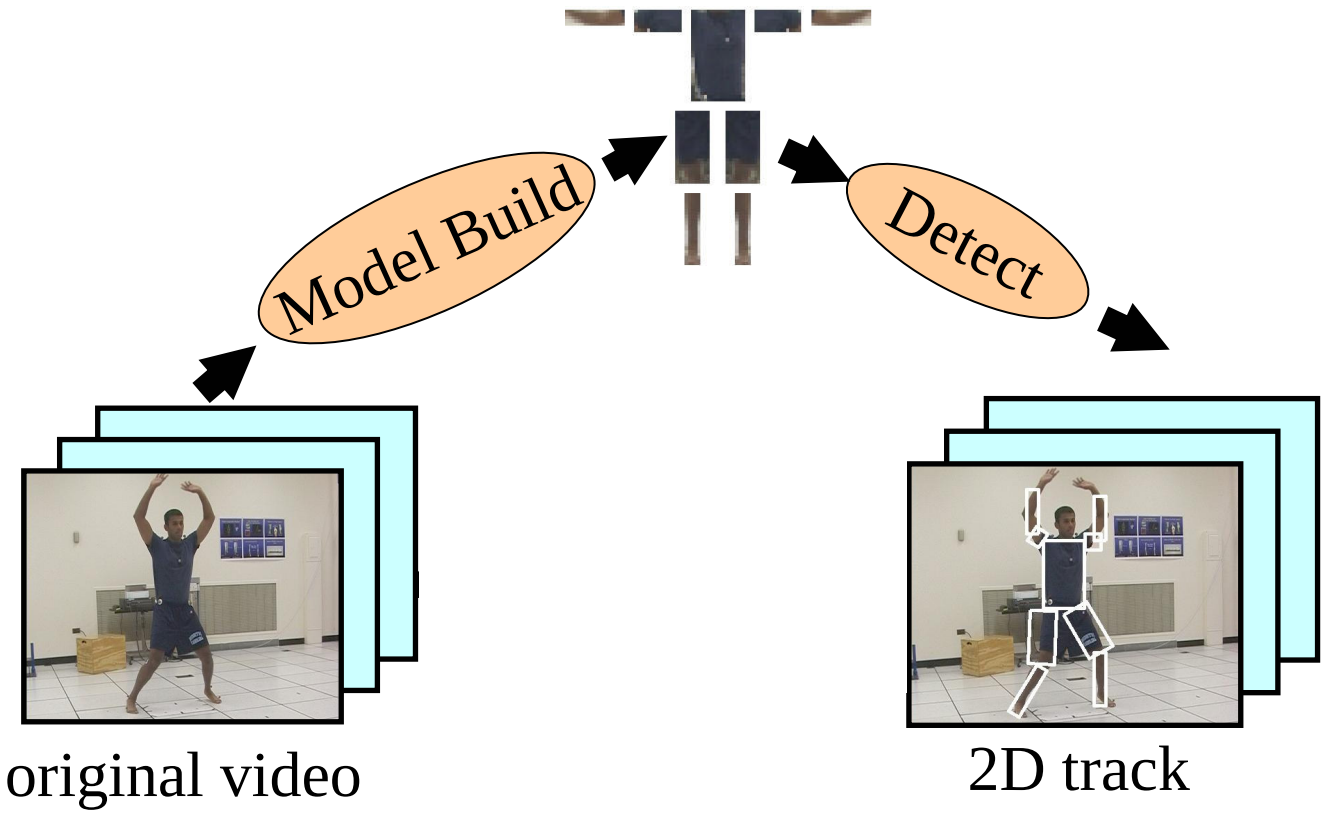
Tracking Algorithm



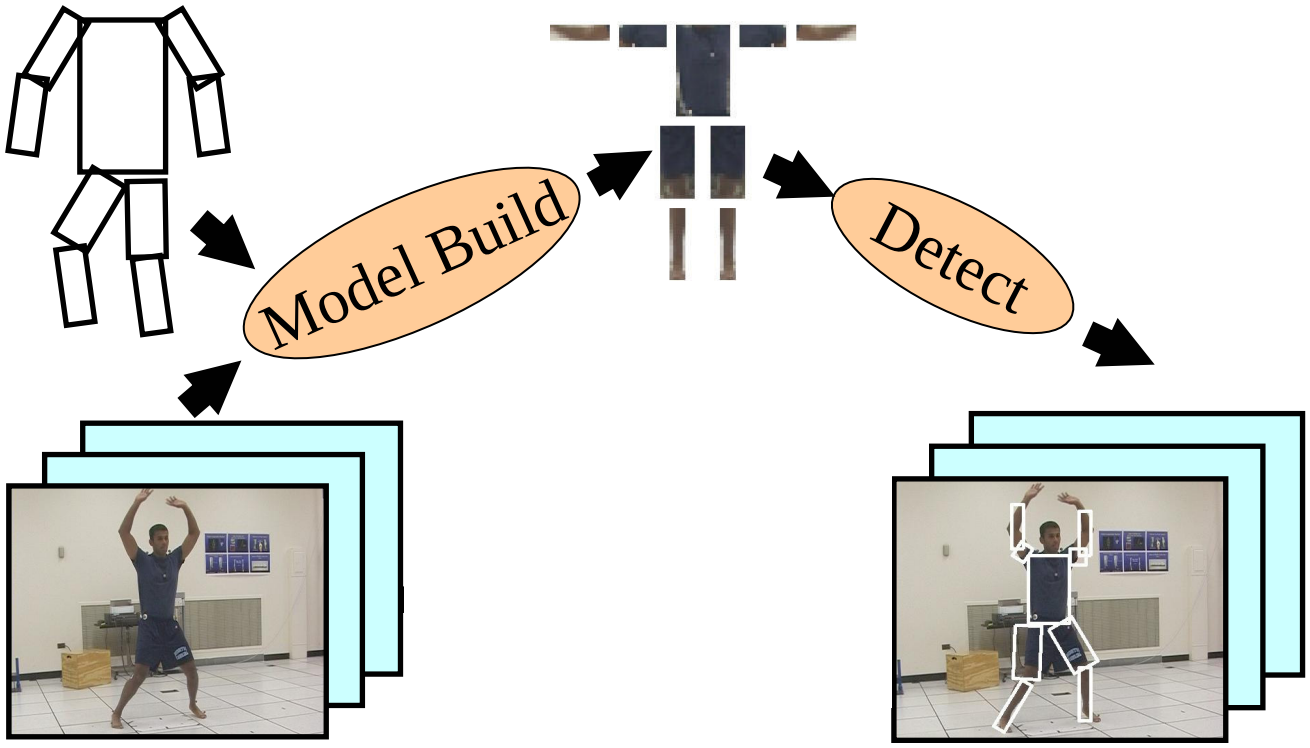
original video

2D track

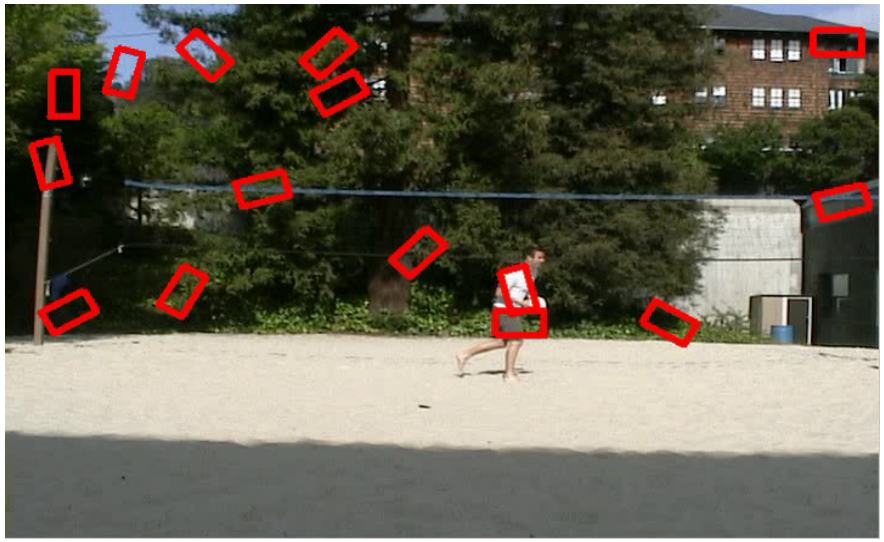
Tracking Algorithm



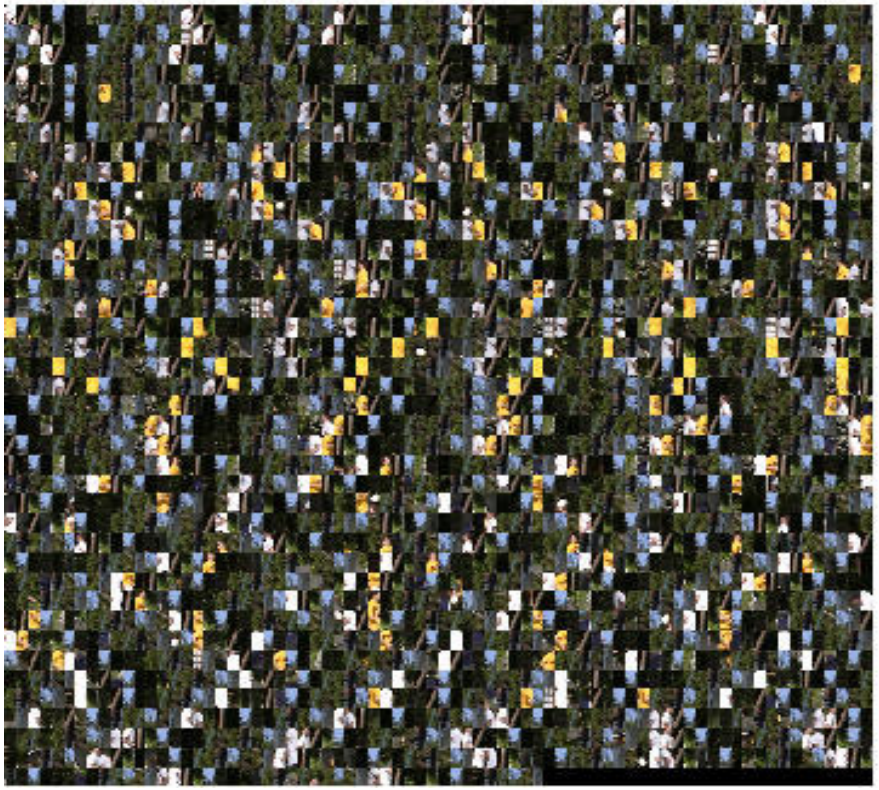
Tracking Algorithm



Look for candidate torsos

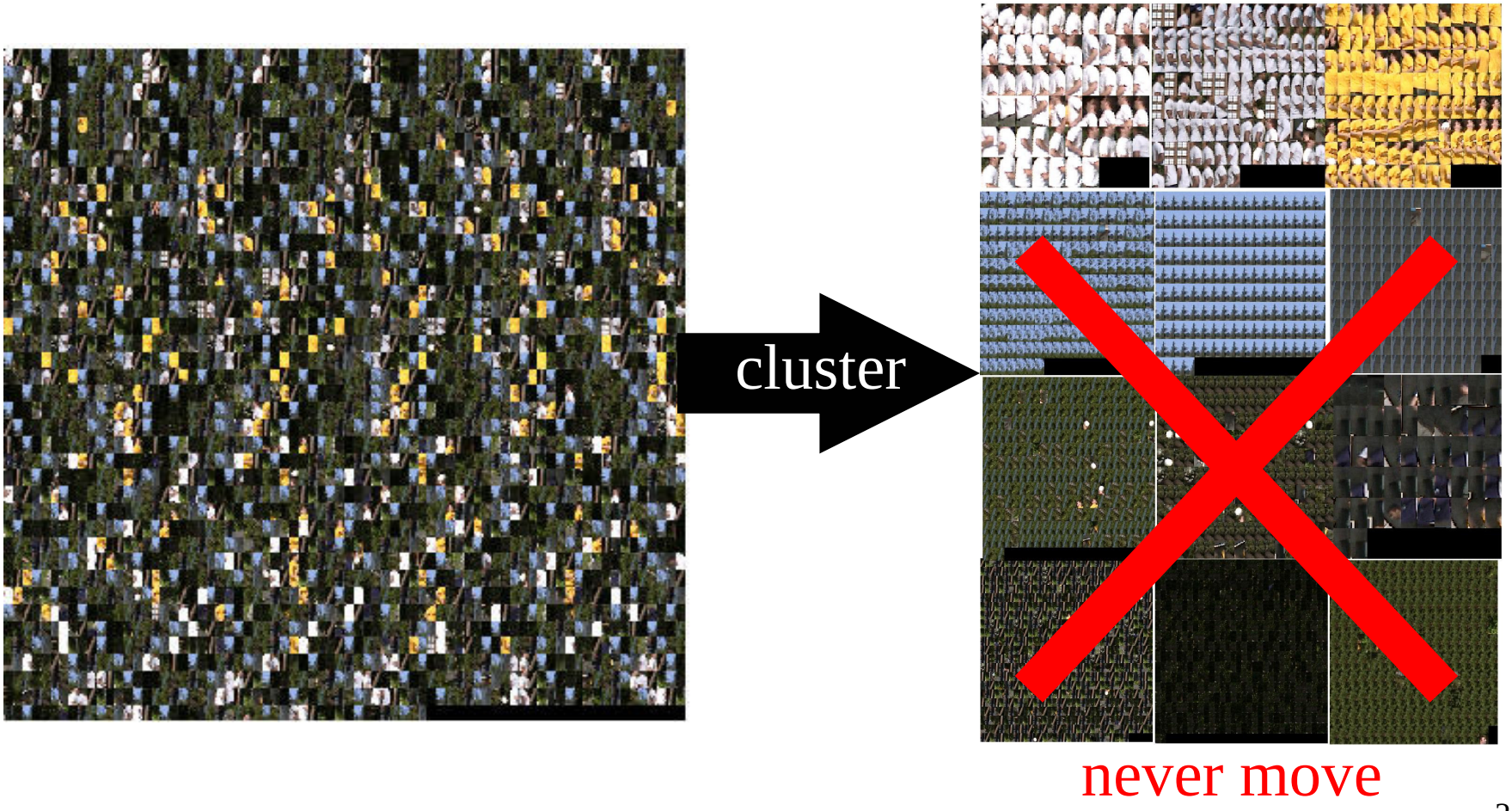


detected torsos

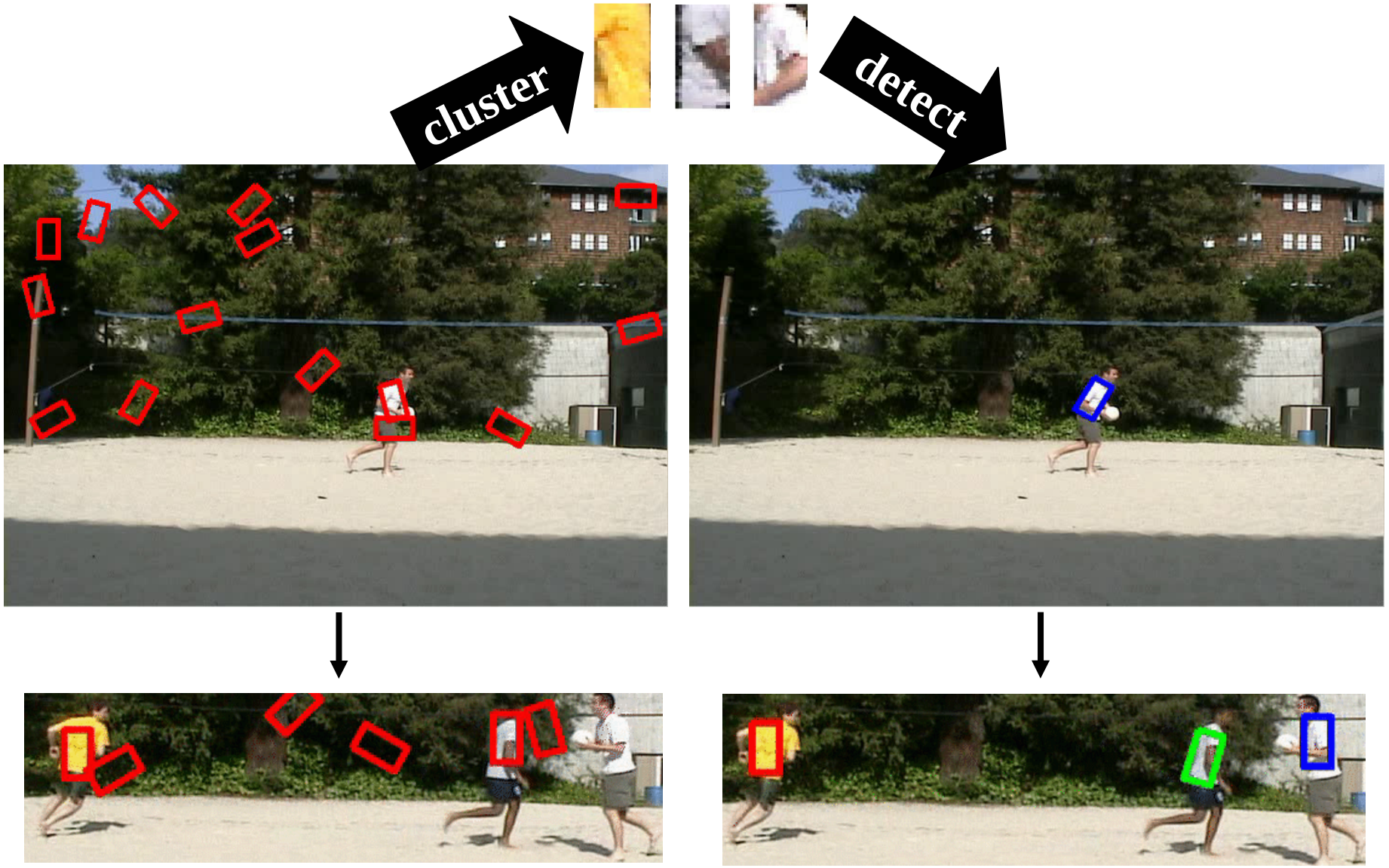


bag of detected torso patches

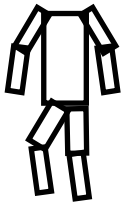
Cluster



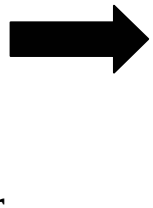
Find new torsos using appearance

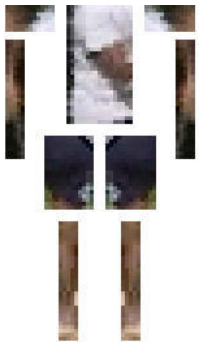


Find arms & legs near torsos



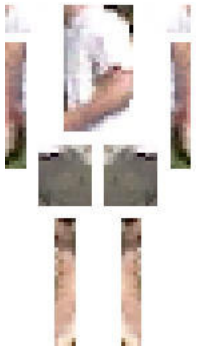
person detector





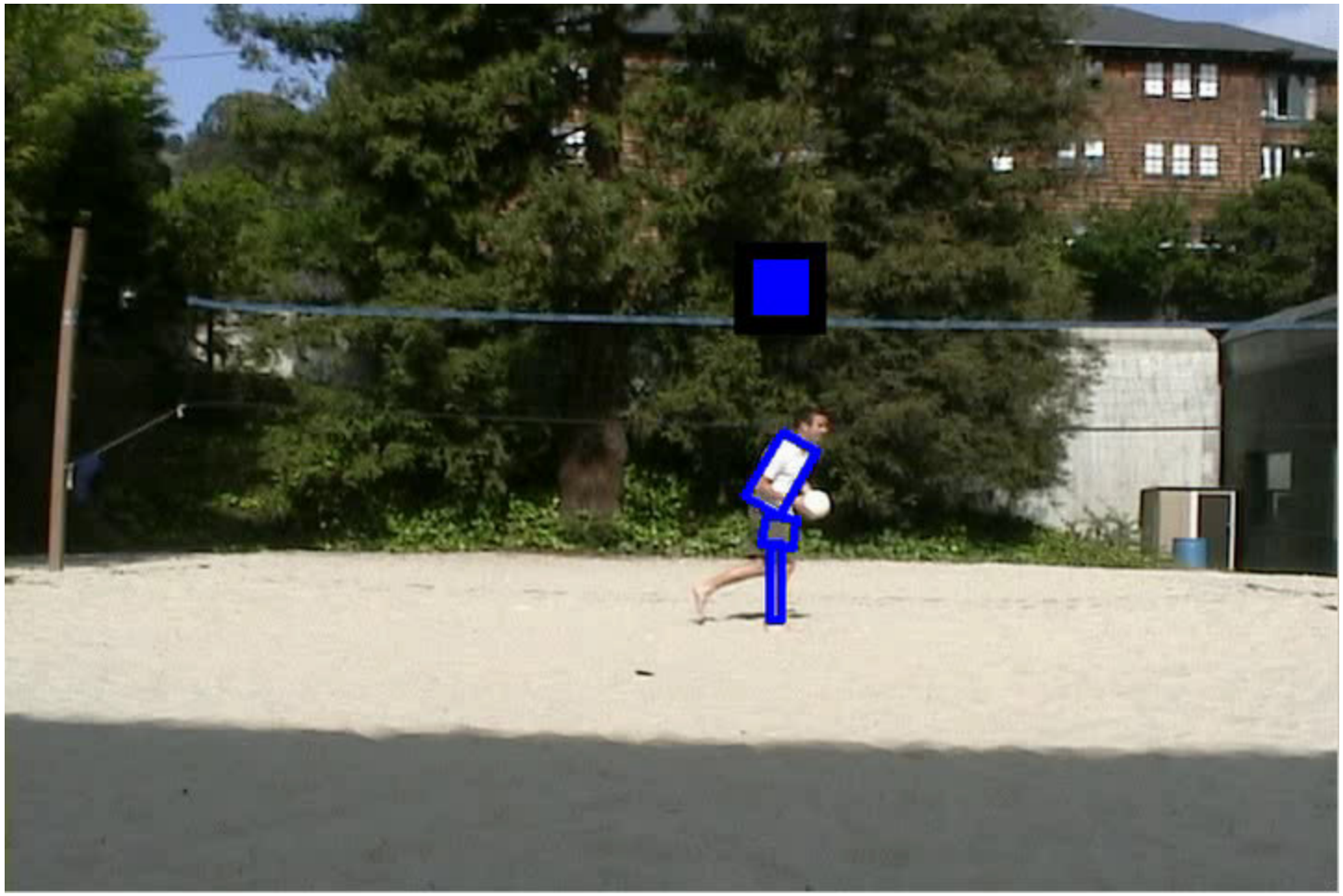
Deva detector

Bryan detector



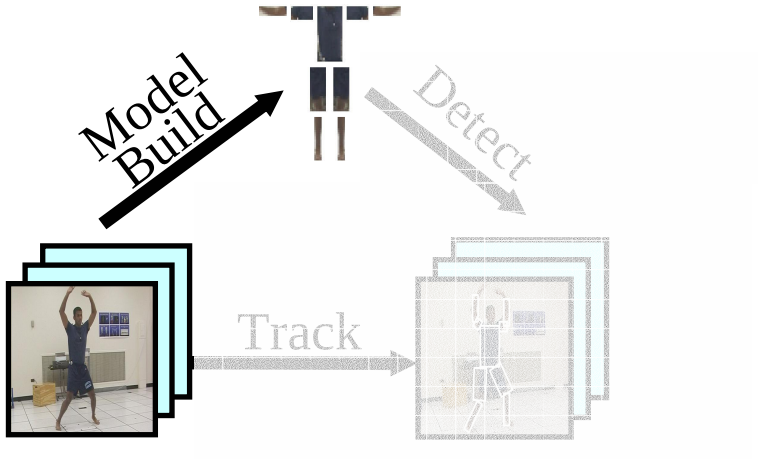
John detector

Final Tracks



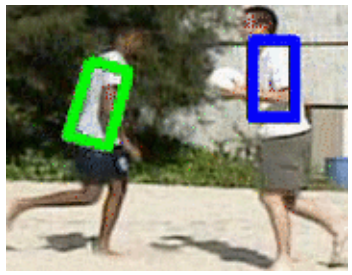
Multi-object tracking as detection

Detectors for Model-Building



Initial detectors do **NOT** detect;
they learn appearance

missed detection



find using learned appearance

opportunistic detection

What to detect?

People take on a variety of poses, aspects, scales

self-occlusion

rare pose

motion blur

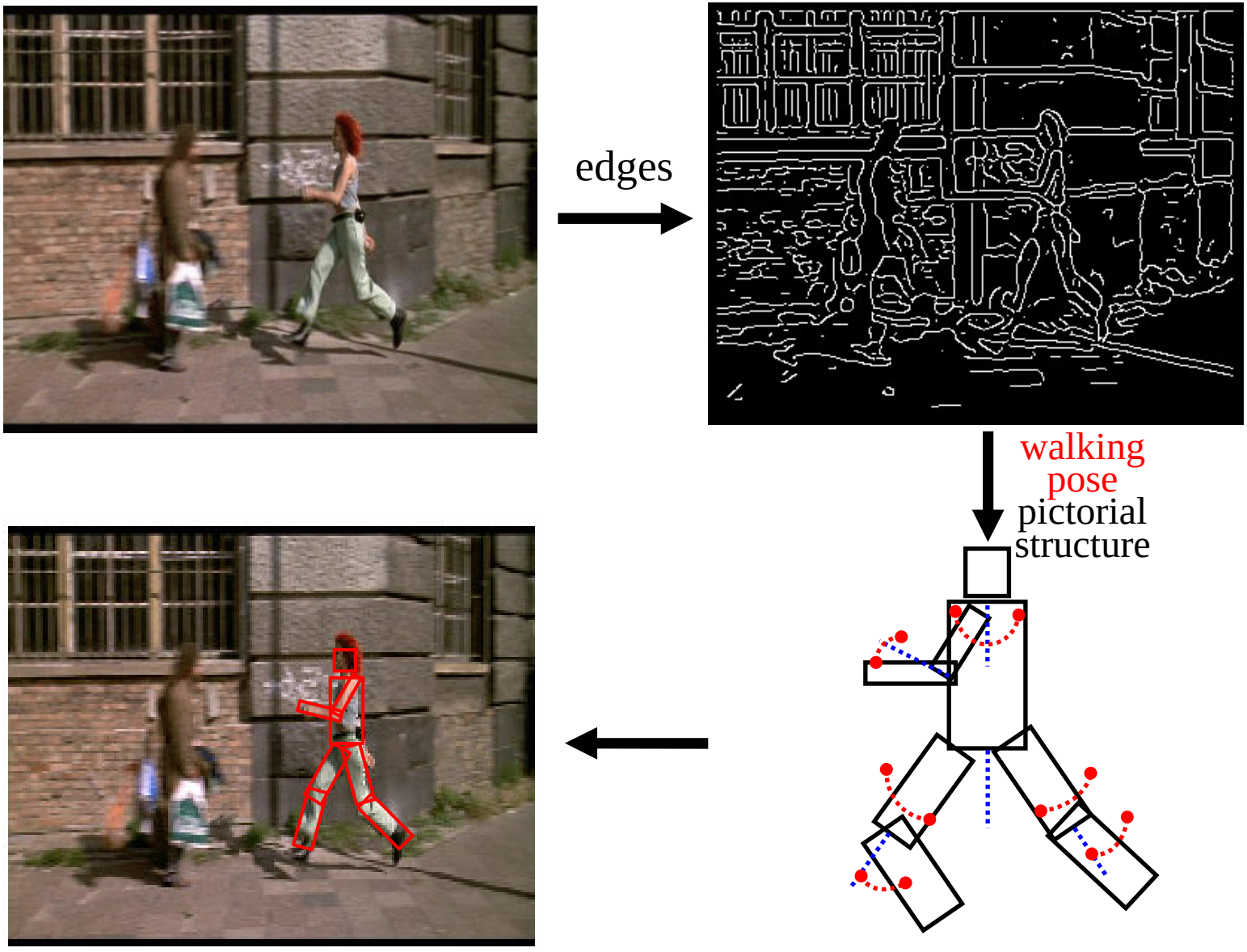
non-distinctive pose

too small

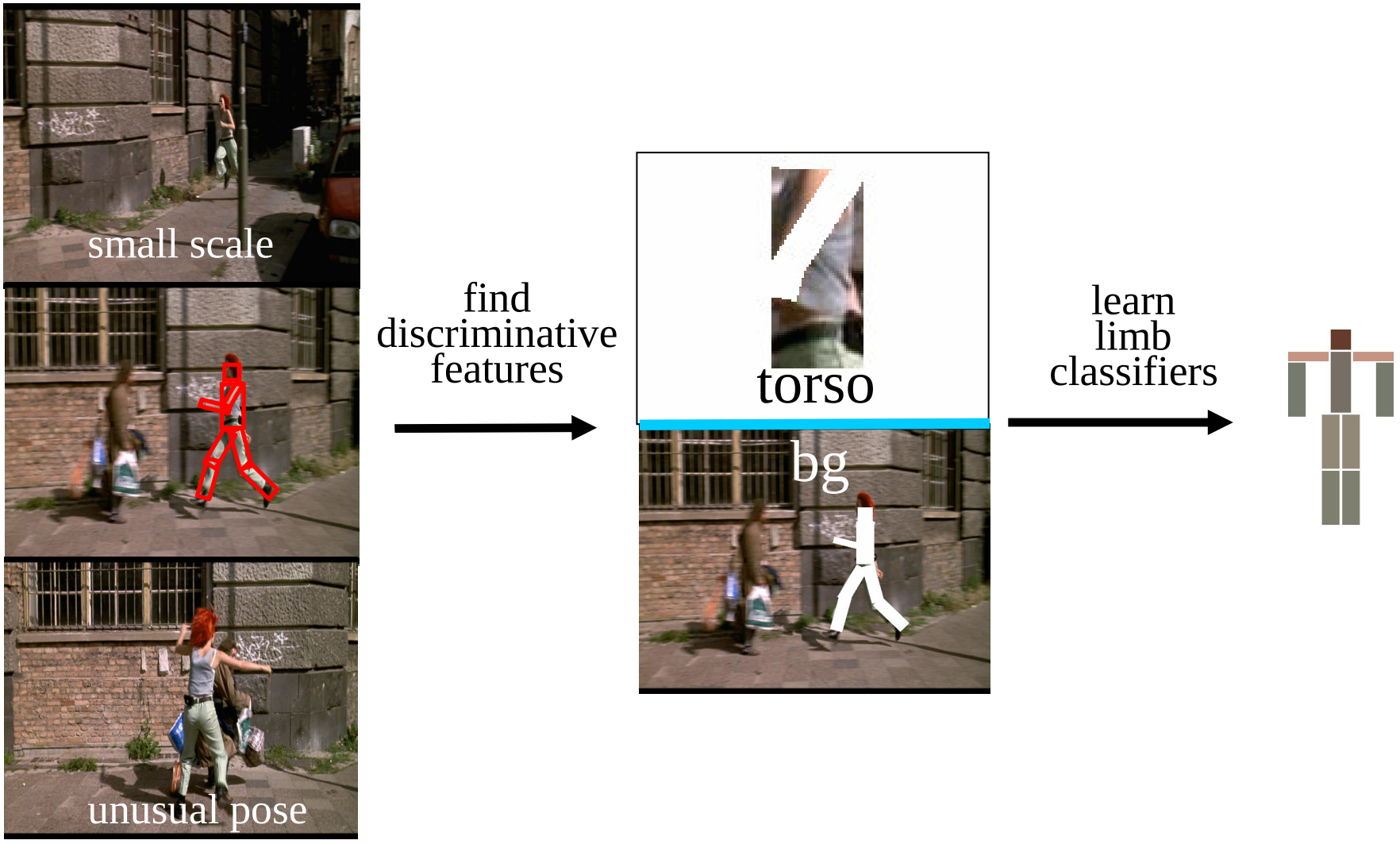
just right

detect this

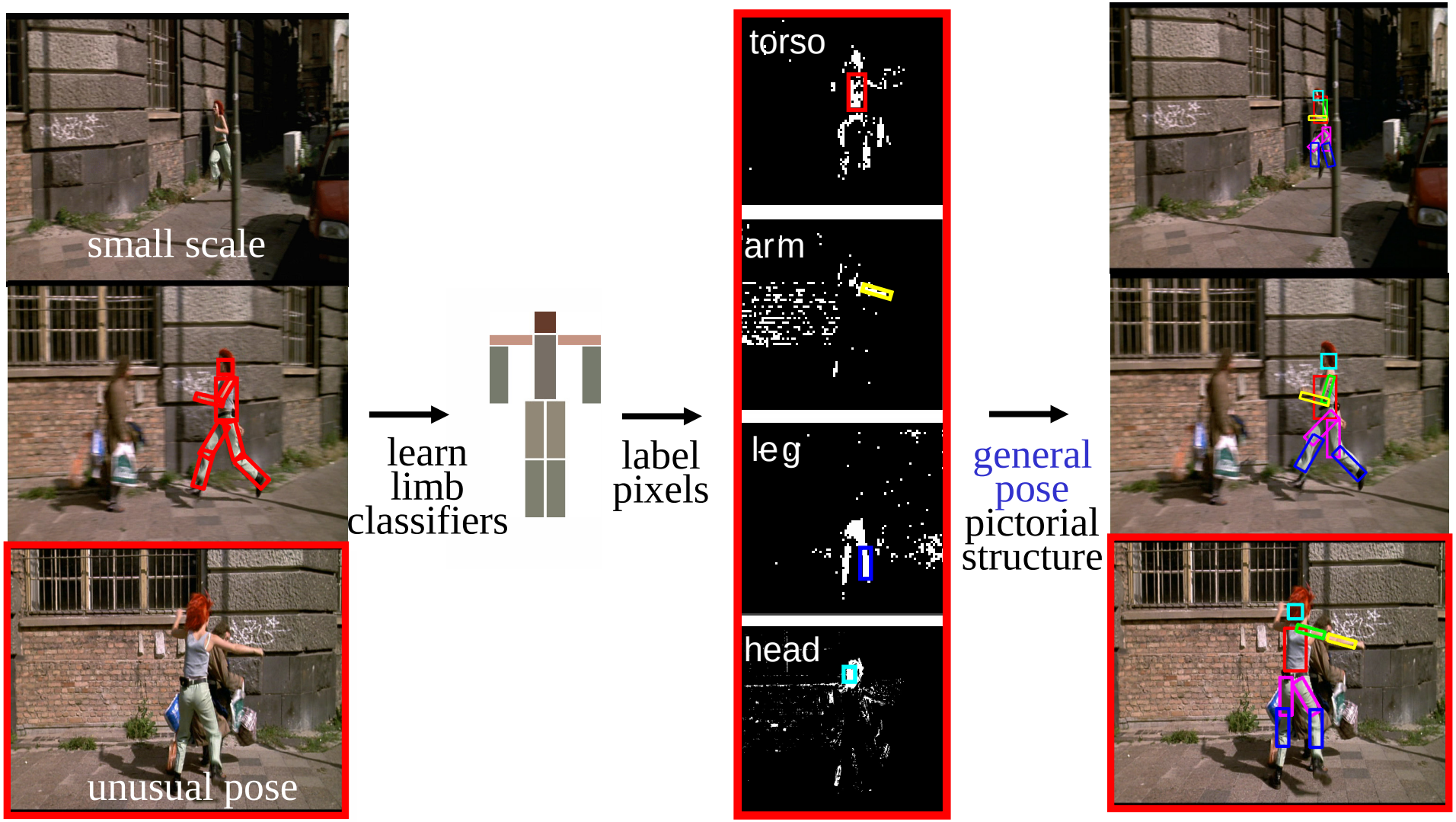
Stylized Pose Person Detector



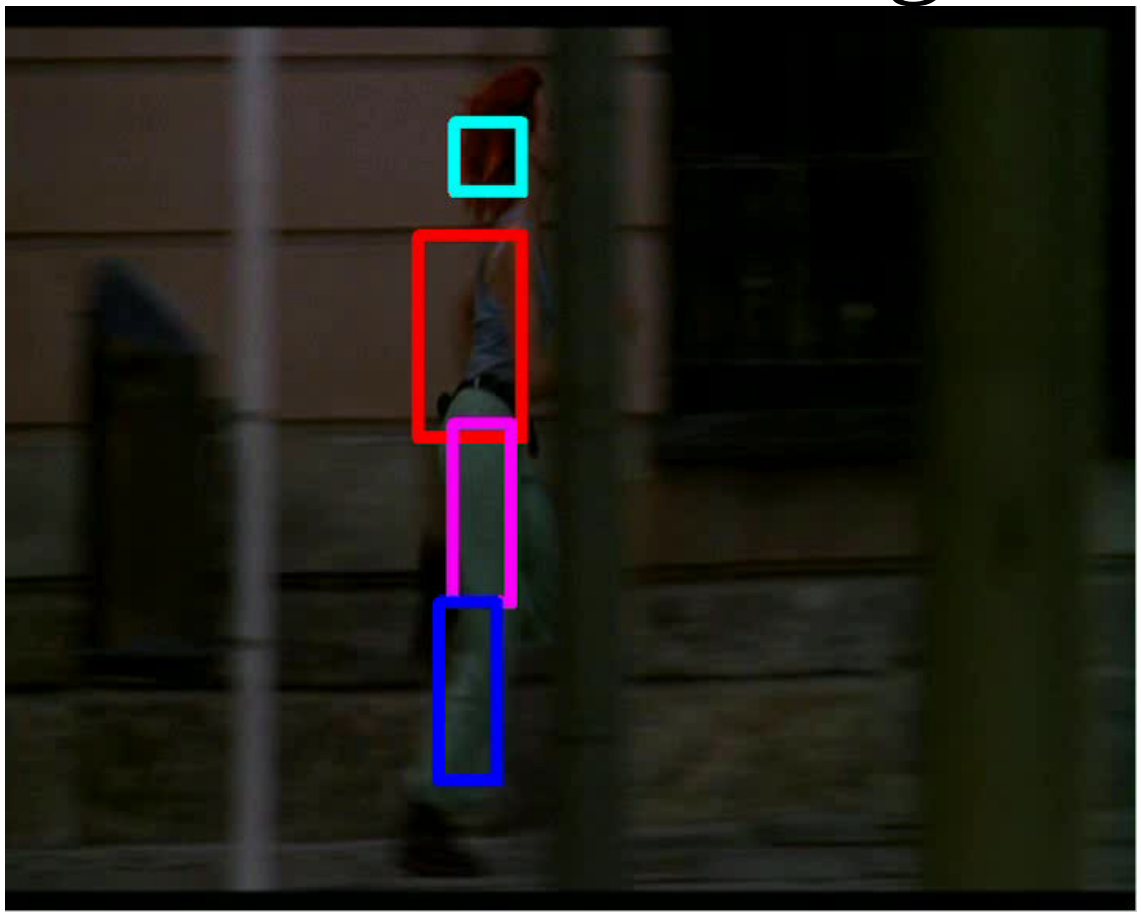
Build Model

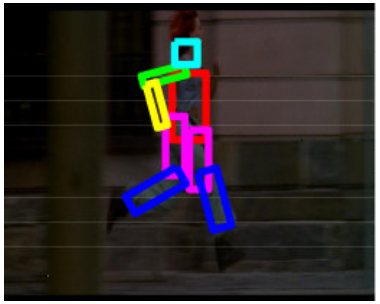

small scale
unusual pose
find
discriminative
features
torso
bg
learn
limb
classifiers

Build Model & Detect

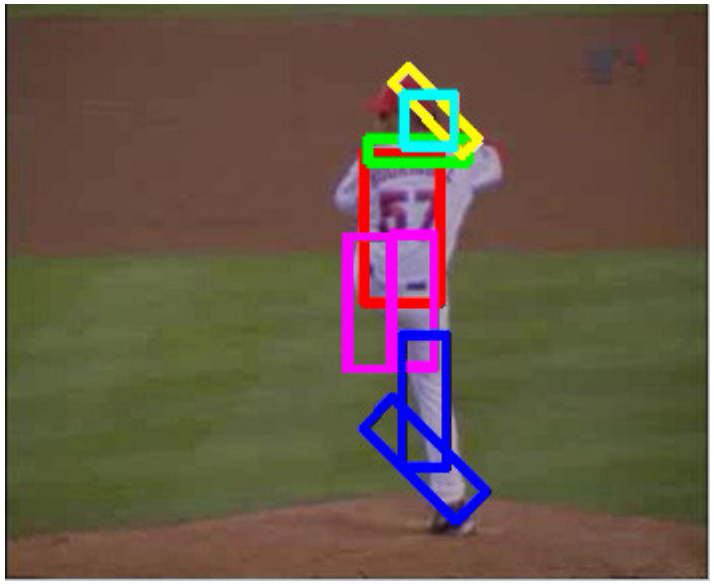


How well do classifiers generalize?

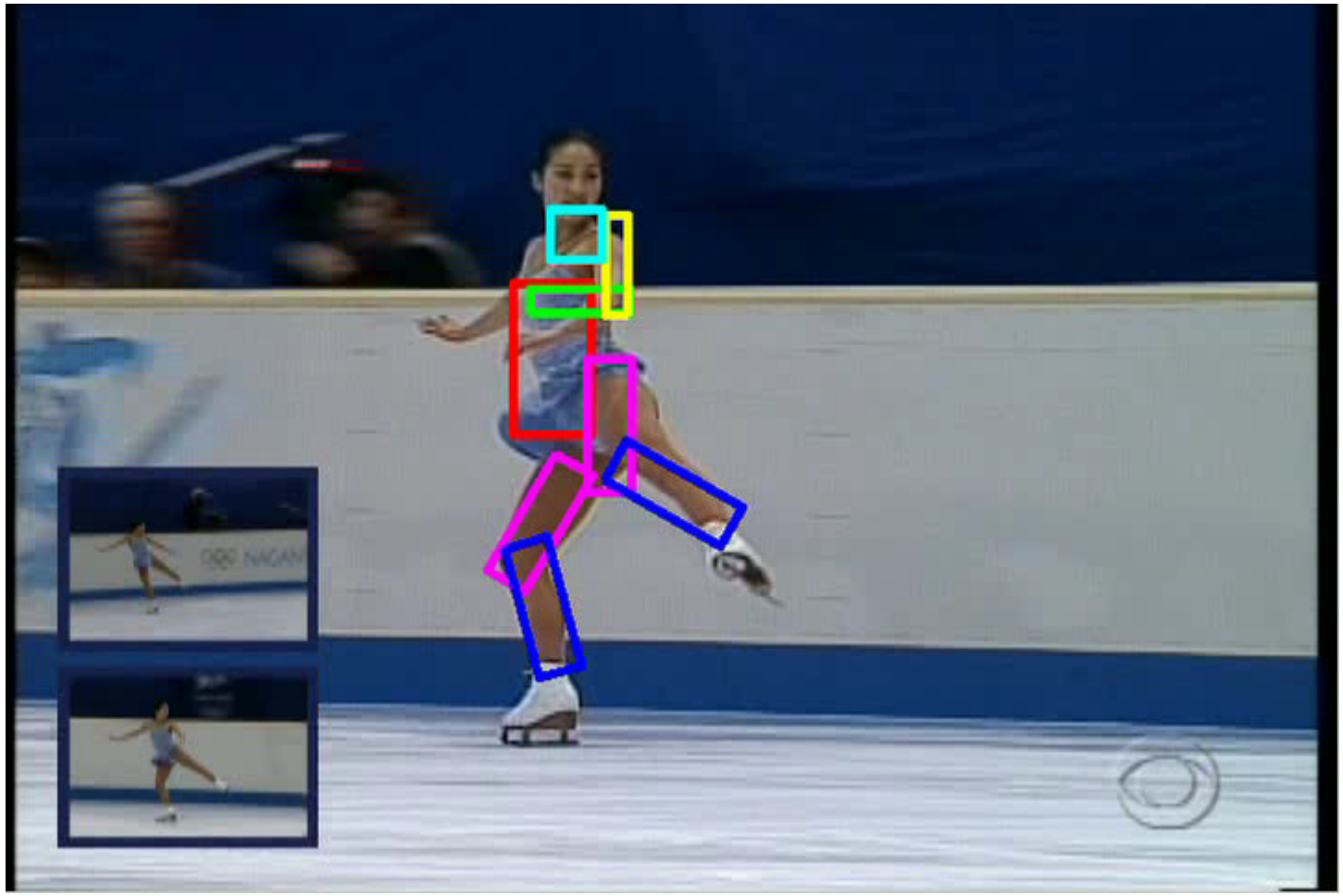


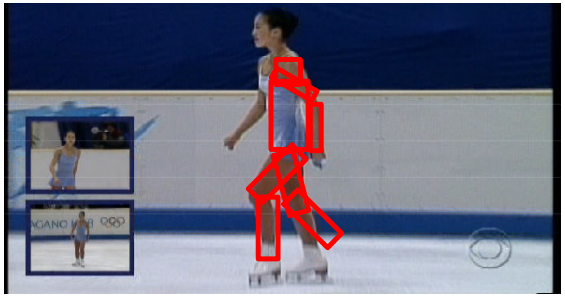


How likely is a ‘typical’ pose?



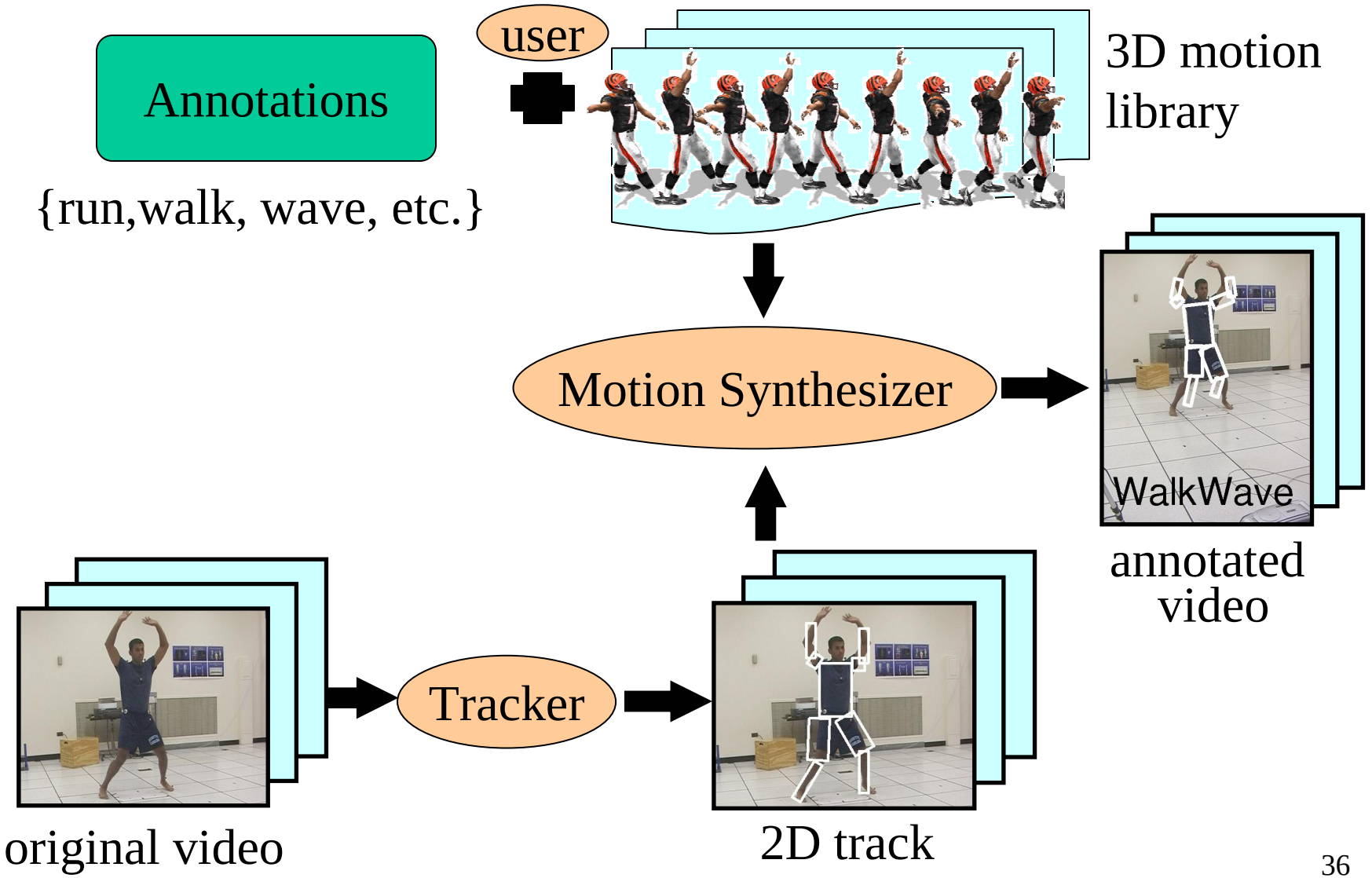
How likely is a ‘typical’ pose?





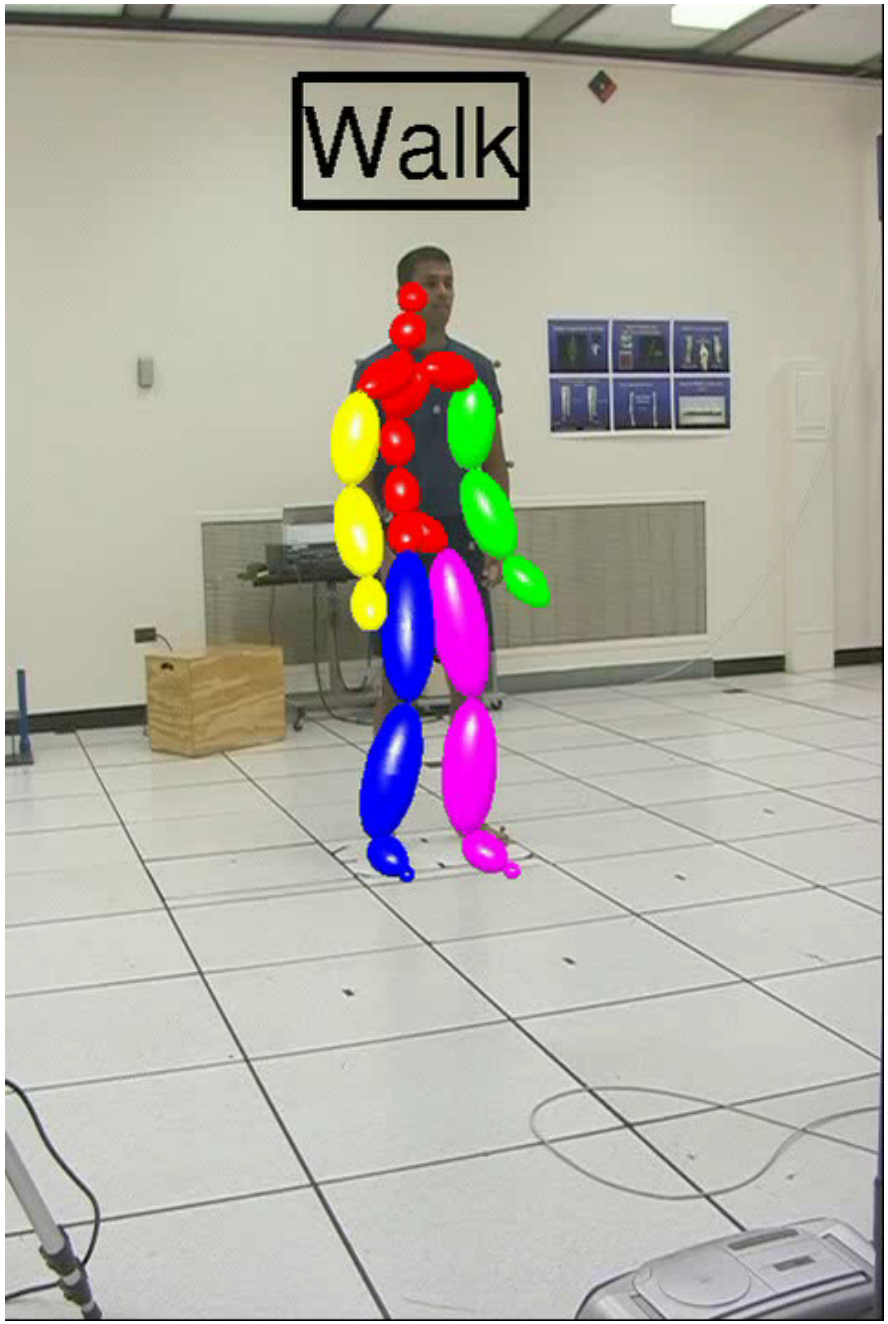
Running Example

Final System

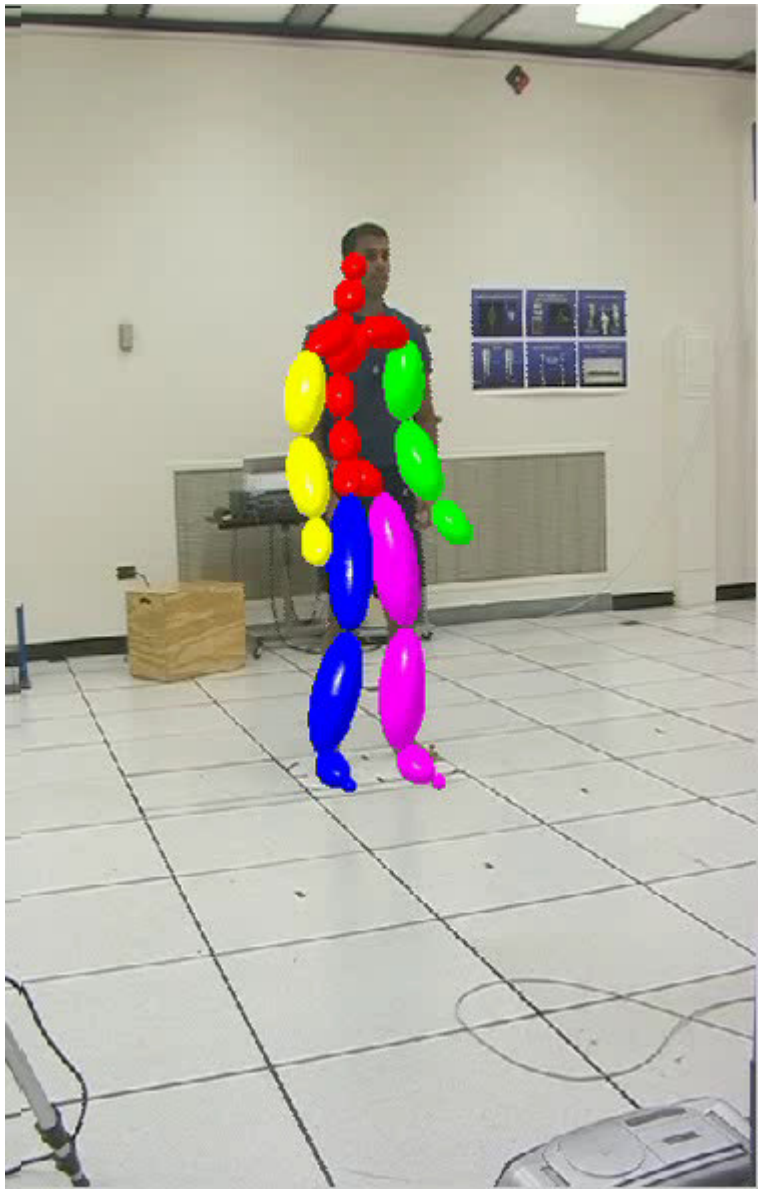


Novel Motion

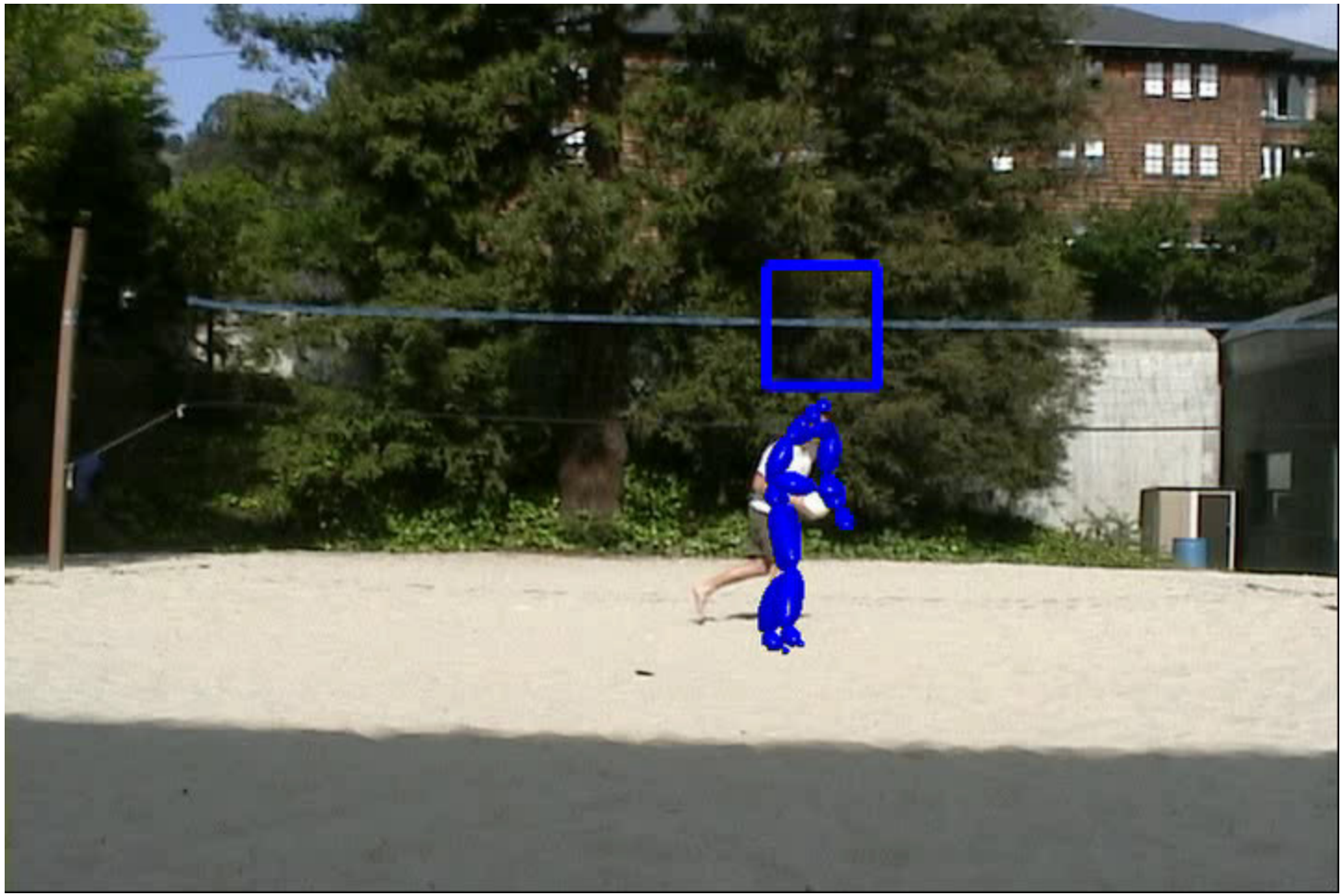
Novel Motion-Annotation



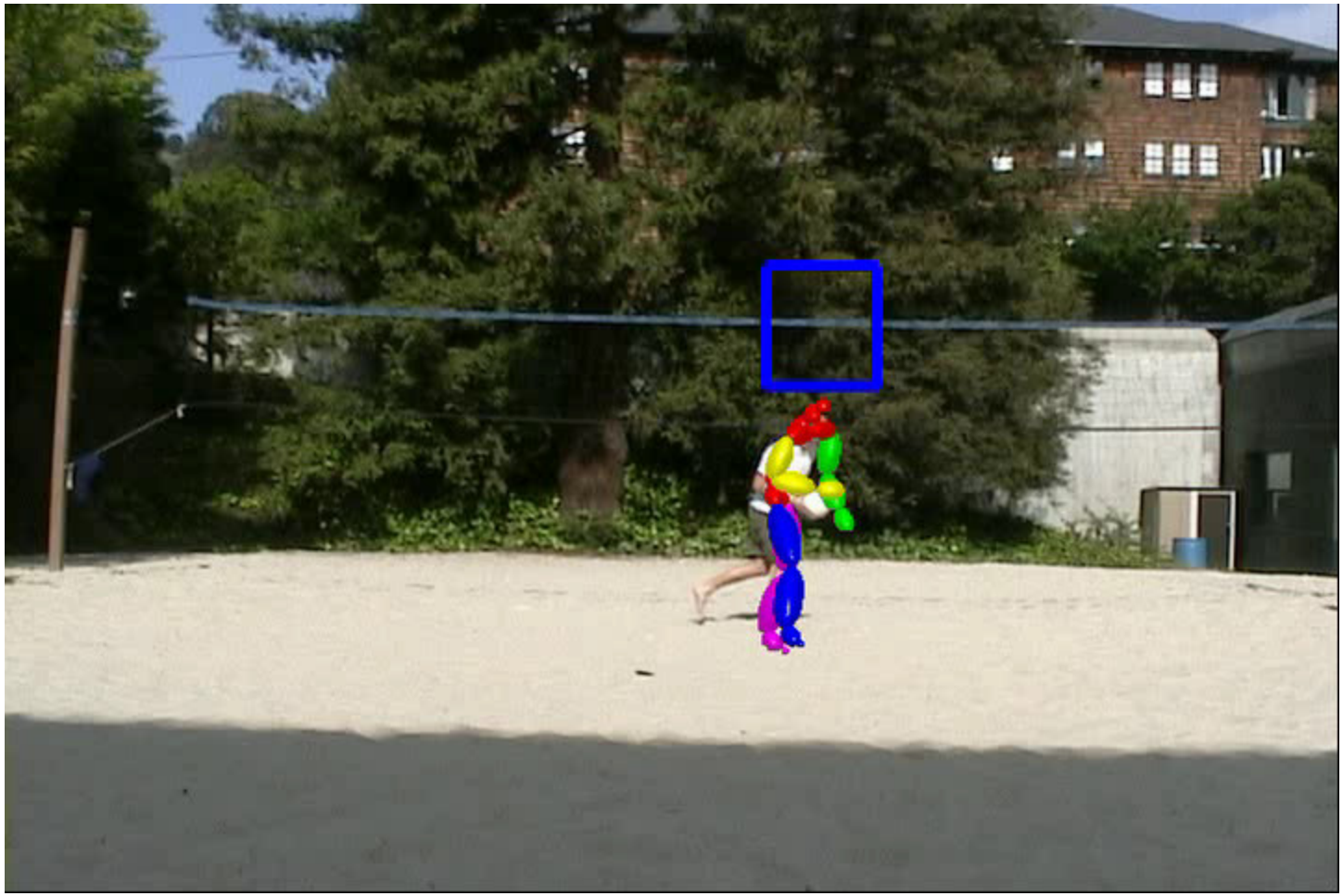
Novel Motion- Recover 3D



Automatic Annotation



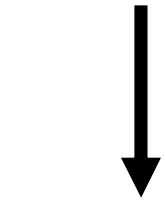
Automatic Annotation (1/3 Speed)



Conclusions

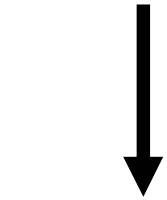
Motion Analysis =

Tracking + Motion Synthesis



Tracking =

Model Building + Detection



Build Models by

Opportunistic Detection