

CMPT 888 – Human Activity Recognition

Greg Mori

Outline

- Intro to class
- Administrative details

Overview

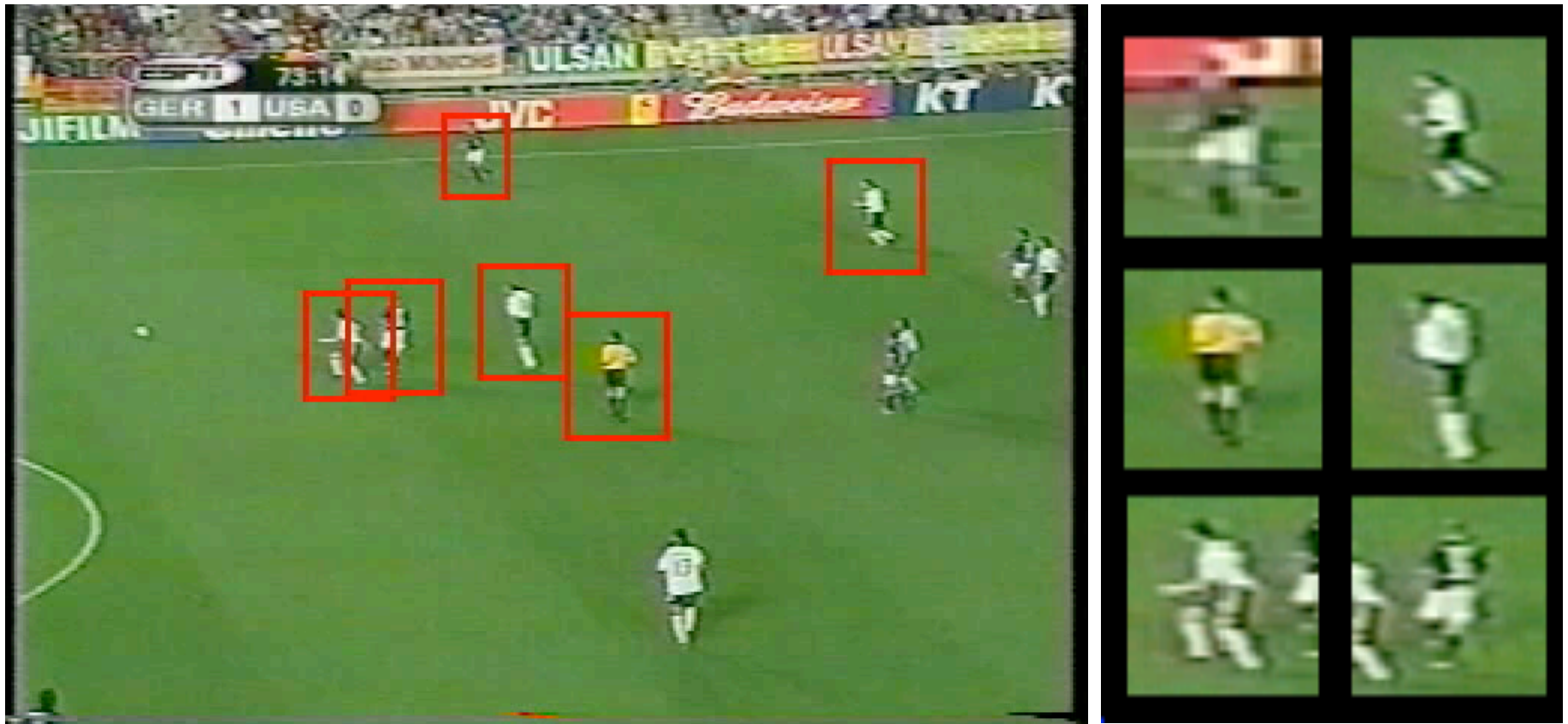
- This class is about vision-based action recognition
 - Input is images or videos
 - Output is description of what people are doing in the images/videos

Action Recognition Example



- Recognize human actions from raw video data

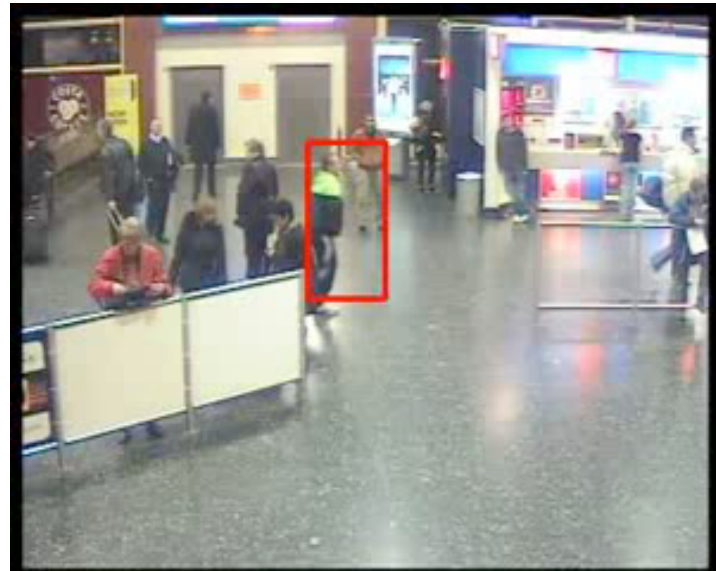
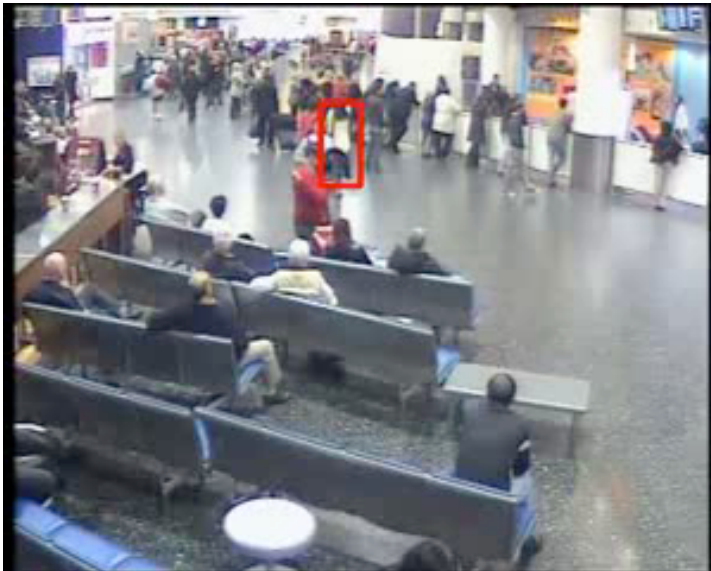
Gathering action data



- 3 components:
 - detect humans, track, recognize action

Applications I

- Automated video surveillance
 - Draw attention to actions of interest
 - Save human operator time



Applications II



- Collect data on pedestrian behaviour
 - Collaboration with Saunier and Sayed (UBC Civil Engineering)

Applications III



Automatically detect falls, near-falls
(with S. Robinovitch SFU)

Why use Computer Vision?

- Competing approaches
 - Wearable sensors
 - Manual labour
- Non-intrusive
 - Do not need cooperative subjects
- Inexpensive, no operator fatigue
 - Semi-automatic techniques

PROBLEM DEFINITION

What is Action Recognition?

- Terminology
 - What is an “action”?
- Output representation
 - What do we want to say about an image/video?

Unfortunately, neither question has satisfactory answer yet

Terminology

- The terms “action recognition”, “activity recognition”, “event recognition”, are used inconsistently
 - Finding a common language for describing videos is an open problem

Terminology Example

- **“Action”** is a low-level primitive with semantic meaning
 - E.g. walking, pointing, placing an object
- **“Activity”** is a higher-level combination with some temporal relations
 - E.g. taking money out from ATM, waiting for a bus
- **“Event”** is a combination of activities, often involving multiple individuals
 - E.g. a soccer game, a traffic accident
- This is contentious
 - No standard, rigorous definition exists

Output Representation

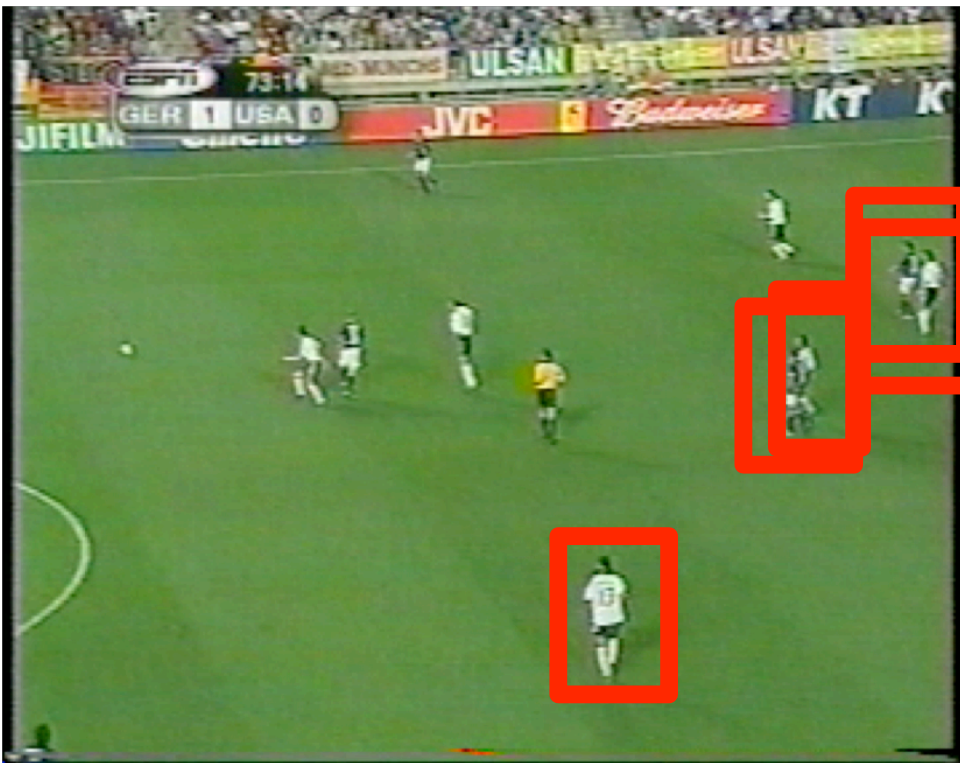
- Given this image what is the desired output?



- This image contains a man walking
 - Action classification / recognition
- The man walking is here
 - Action detection

Output Representation

- Given this image what is the desired output?



- This image contains 5 men walking, 4 jogging, 2 running
- The 5 men walking are here
- This is a soccer game

Output Representation

- Given this video what is the desired output?



- Frames 1-20 the man ran to the left, then frames 21-25 he ran away from the camera
- Is this an accurate description?
- Are labels and video frames in 1-1 correspondence?

Challenges in Recognition

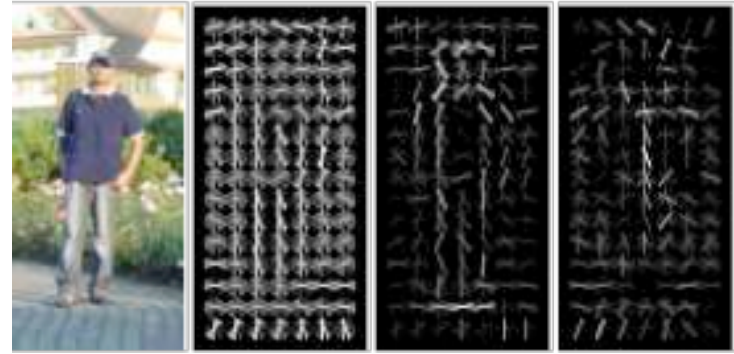
- Intra-class variation
- Object pose variation
- Background clutter
- Occlusion
- Lighting



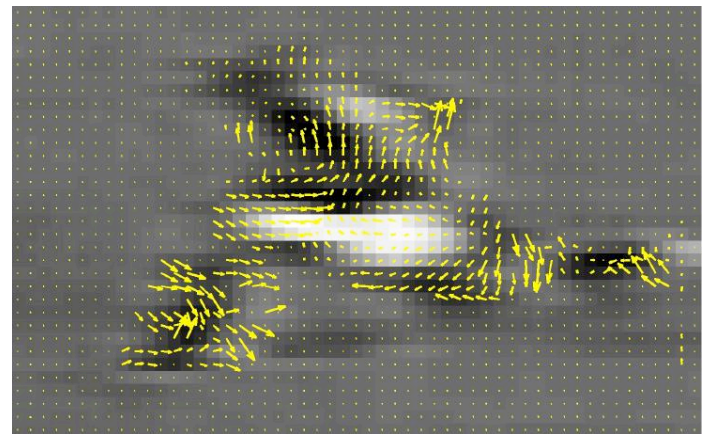
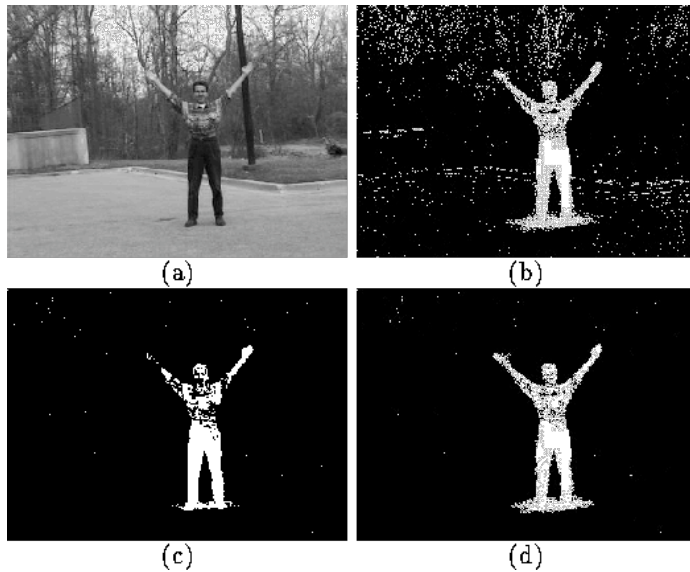
TRIMESTER PREVIEW

Week 2

- Preliminaries
 - Human detection
 - Background subtraction
 - Optical flow

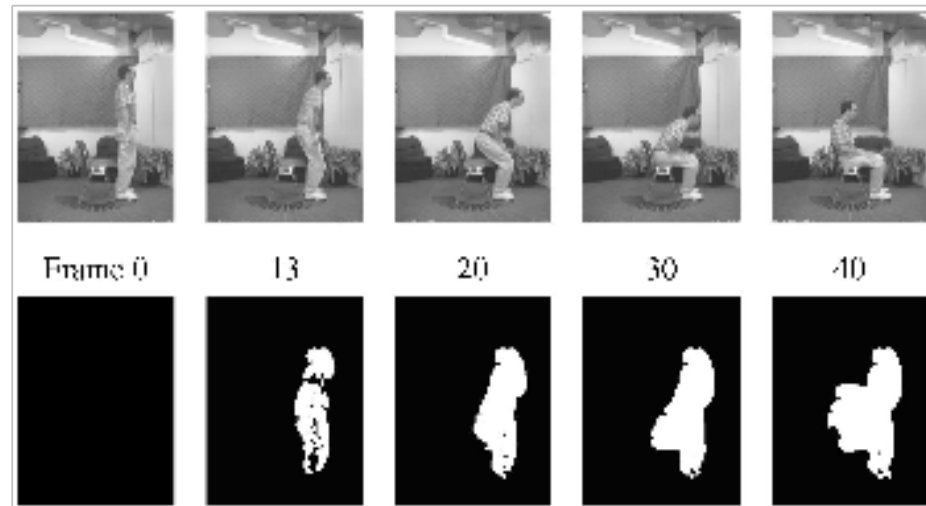


Dalal + Triggs CVPR05



Weeks 3-4

- Motion Templates

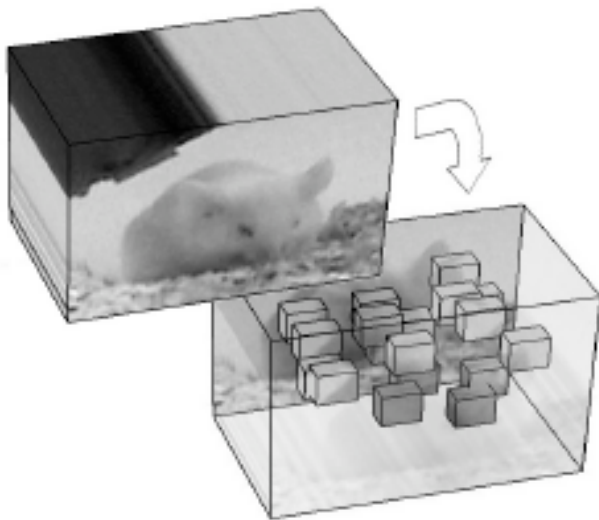


Bobick and Davis PAMI01

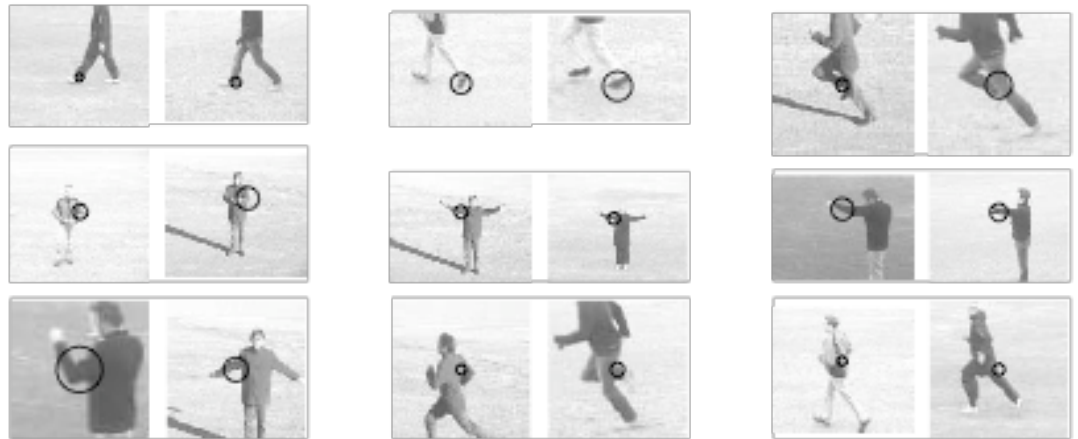
Efros et al. ICCV03

Weeks 5-6

- Local feature video representations



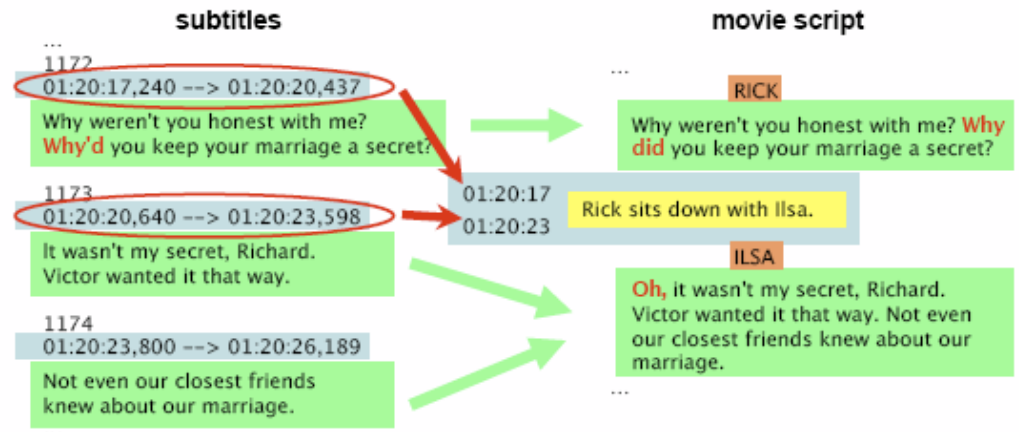
Dollar et al. VSPETS05



Schuldt et al. ICPR04

Week 7

- Unsupervised and weakly supervised methods



Learning realistic human actions
from movies

Demo

I.Laptev, M.Marszalek, C.Schmid and B.Rozenfeld
In Proc. CVPR 2008

For more information visit:
<http://www.irisa.fr/vista/actions>

Laptev et al. CVPR08

Week 8

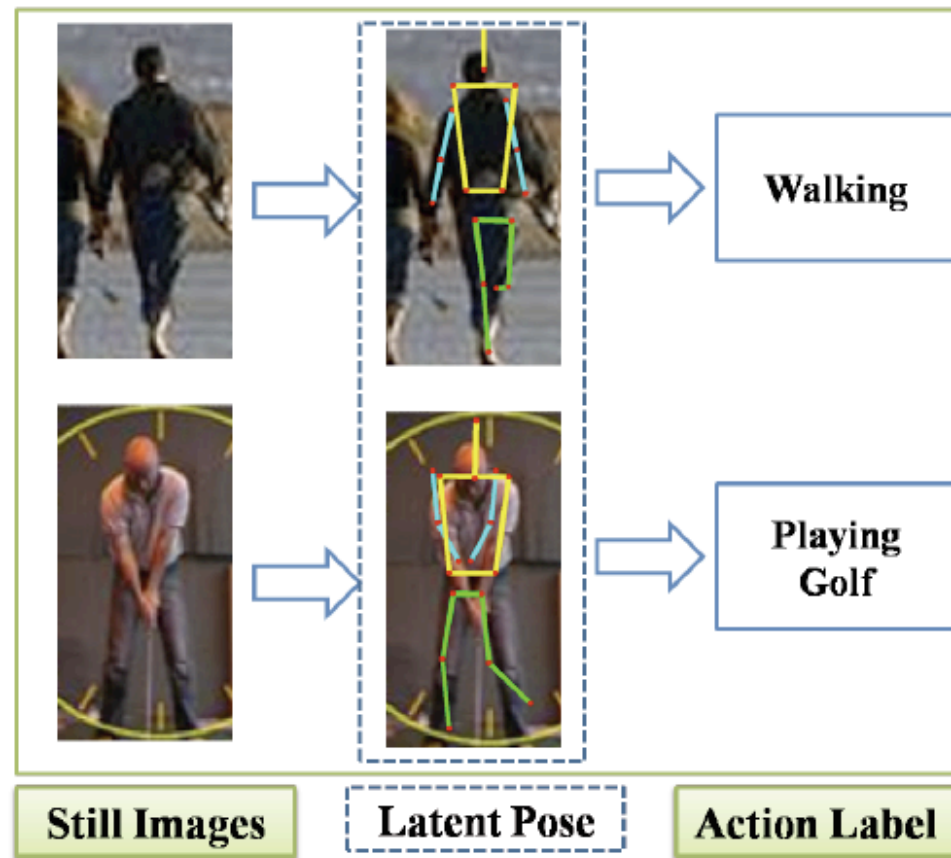
- Temporal models



? — ? — ? — ? — ? — ? — ? — ? — ? — ?

Week 9

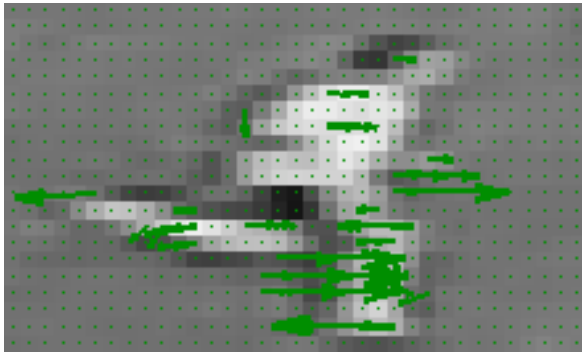
- Human pose estimation and pose retrieval



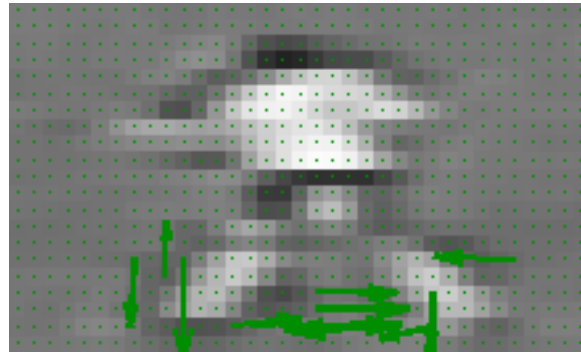
Yang et al. CVPR10

Week 10

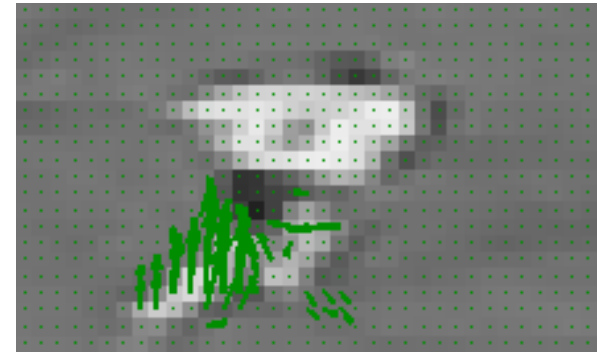
- Discriminative methods



Run right



Walk left



Run right 45

Fathi and Mori CVPR08

Week 11

- Human actions in still images



Wang et al. CVPR06

ADMINISTRIVIA

Course Plan

- Read research papers
 - For each topic I present important papers
 - Students each present a recent paper
 - We discuss
- Do a project
 - Gain in-depth experience on a problem and algorithm

Introductions

Prerequisite

- No formal prerequisites
 - But it would be best if you know some **computer vision / image processing** and some **machine learning**
- You will need to do the usual things
 - Math (continuous), programming, reading, writing, presenting
- Ask me if you are concerned

Grading Scheme

- 10% Class participation
 - Participate in discussions about papers, ask/answer questions
- 10% Reading assignments
 - 1 or 2 papers each week; subset of the ones I present
- 10% Paper presentation
 - Choose from list of papers online
- 10% Assignment
 - Small programming assignment on motion analysis
- 60% Project
 - Individual or in small groups
 - Presentation, written report

Reading Assignments

- Similar to mini paper review
 - One paragraph summarizing paper
 - Critical discussion (what you like / don't like)
 - Questions you have (for me to explain)
- Due before start of lecture via email
 - **First one due Monday**
- These details and list of papers are online

Paper Presentations

- Choose one paper that interests you
 - From list online / in syllabus
- 20 minute presentation
 - 10+ minutes questions/discussion
 - Feel free to use slides provided by authors

Assignment

- Short programming assignment
 - Background subtraction
 - Motion-based action recognition
- Out next week, due 2 weeks later

Project

- Major component of course
 - Recognize actions
- Implement existing technique
 - Or variant thereof
 - Can use something you're working on in your research
 - Must recognize actions
 - Must do something that didn't exist before this course
- Proposal, presentation, report

Course Plan

- Next week
 - Preliminaries
 - Background subtraction, human detection, motion
- After that
 - Papers, papers, papers