
Frontiers of *Human Activity Analysis*

J. K. Aggarwal
Michael S. Ryoo
Kris M. Kitani

Introduction

Semantic video understanding

- Goal
 - Labeling of all objects, persons, and their events in a given video



Person 1 – *teases **P2**,
runs away*

Person 3 – *kicks **P1***

Person 4 – *stops fighting*

- Develop automated algorithms for the video recognition

Semantic video understanding

- Goal
 - Labeling of all objects, persons, and their events in a given video



Lioness

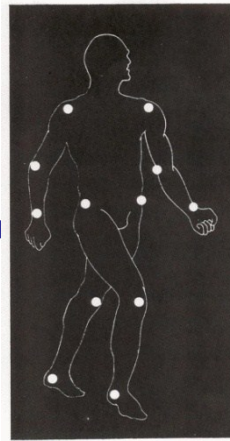
Baby zebra

Hunting – chasing

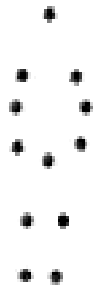
Succeeded

- Develop automated algorithms for the video recognition

Beginnings of Activity Recognition



- Johansson's experiments (1973) - lights attached to major joints of a person, dressed in black and human recognition of activity.
- Representing each rigid body part by two points, and determining the structure of jointed objects under orthographic projection.



BeginningsContd.

- Hoffman data/MIT.
- Six points on a walking man, 0.26 sec.
- Artificial Intelligence, vol.19, 1982, 107-130, Webb and Aggarwal.

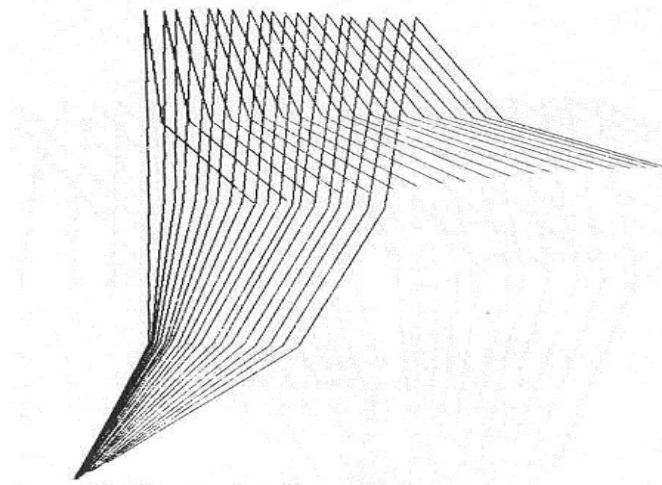


TABLE 1. Rigid part lengths (in meters); Hoffman data

From	To	Estimated length	Actual length	Relative error
shoulder	elbow	0.344	0.335	2.53%
elbow	wrist	0.283	0.274	3.35%
shoulder	hip	0.584	0.579	0.808%
hip	knee	0.438	0.437	0.175%
knee	ankle	0.435	0.437	1.90%

- This may be considered the beginning of estimation of structure and action recognition of jointed objects

Levels of video understanding

- Object-level understanding
 - Locations of persons and objects
 - E.g., 'lion' appeared in the video
- Tracking-level understanding
 - Object trajectories – correspondence
- Pose-level understanding
 - Human body parts
- Activity-level understanding
 - Recognition of human activities and events

Object detection

- Pedestrian (i.e. human) detection
 - Detect all humans in the given video



Object Tracking :

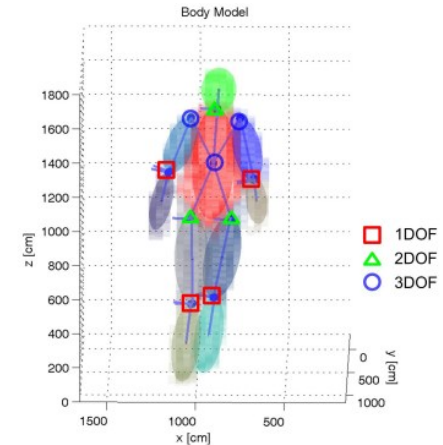


I-LIDS[®] (AVSS 2007)

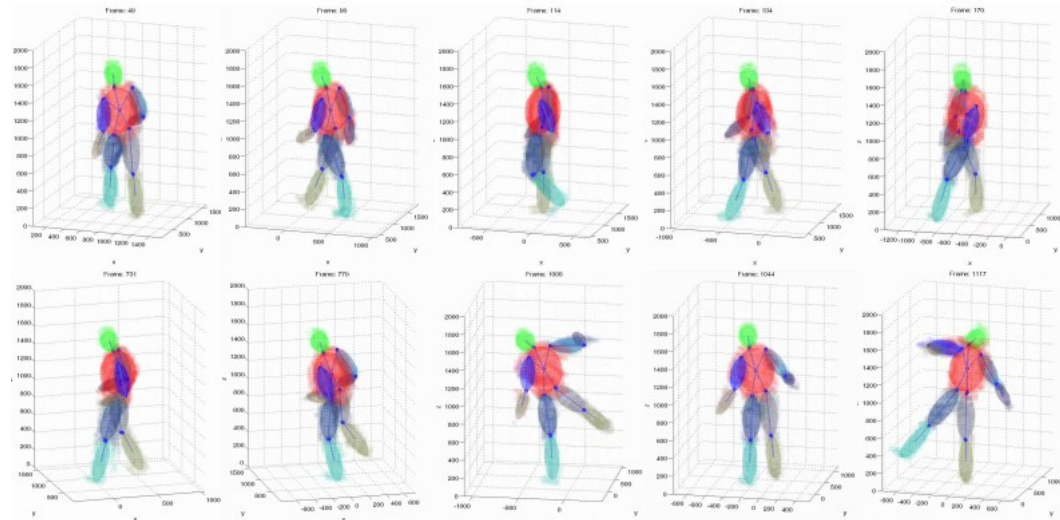
Ryoo and Aggarwal,
CVPR 2008

Posture recognition

- Human pose
 - Joint locations of a person measured per frame
 - 3-D body parts



- Video as a sequence of poses



Activity recognition

- Group activity
 - Stealing in an Apple store



What is activity recognition?

- Human activity recognition
 - Automated detection of ongoing events from video data.
 - Computer analysis of inputs from cameras.
 - Human actions, human-human interactions, human-object interactions, group activities.

Input video:



← Punching ($p2$, $p1$) →

Human activity

- Human activity
 - A collection of human/object movements with a particular semantic meaning
 - i.e., particular structure
- Activity recognition
 - Finding of video segments containing such movements
 - Must **search** for video segment that display properties of the movements

Introduction

Applications

Surveillance

- Ubiquitous cameras in public places (e.g. CCTVs).
 - In London, an average person is monitored 300 times / day.
- Goal
 - Monitor suspicious **activities** for real-time reactions.
 - 'Fighting', 'stealing'.
 - Currently, surveillance systems are mainly for recording.
- **Activity recognition** is essential for surveillance and other monitoring systems in public places

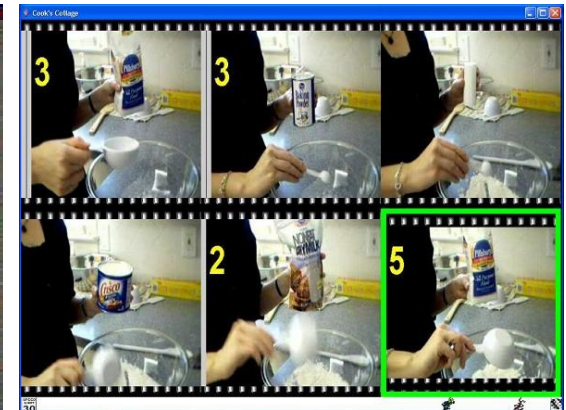
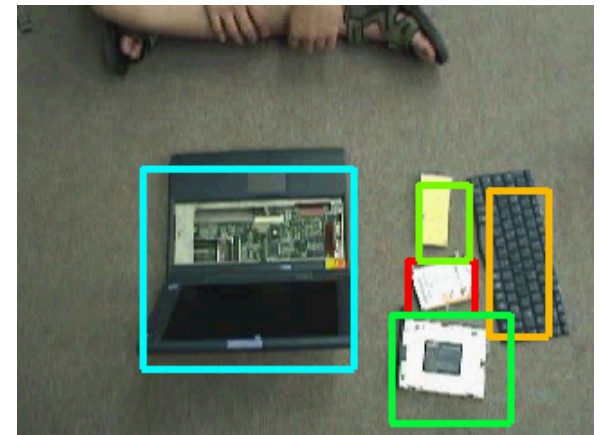


Intelligent environments (HCI)

- Intelligent home, office, and workspace
 - Monitoring of elderly people and children.
 - Support one's quality of life.
 - Recognition of ongoing **activities** and understanding of current context is essential.

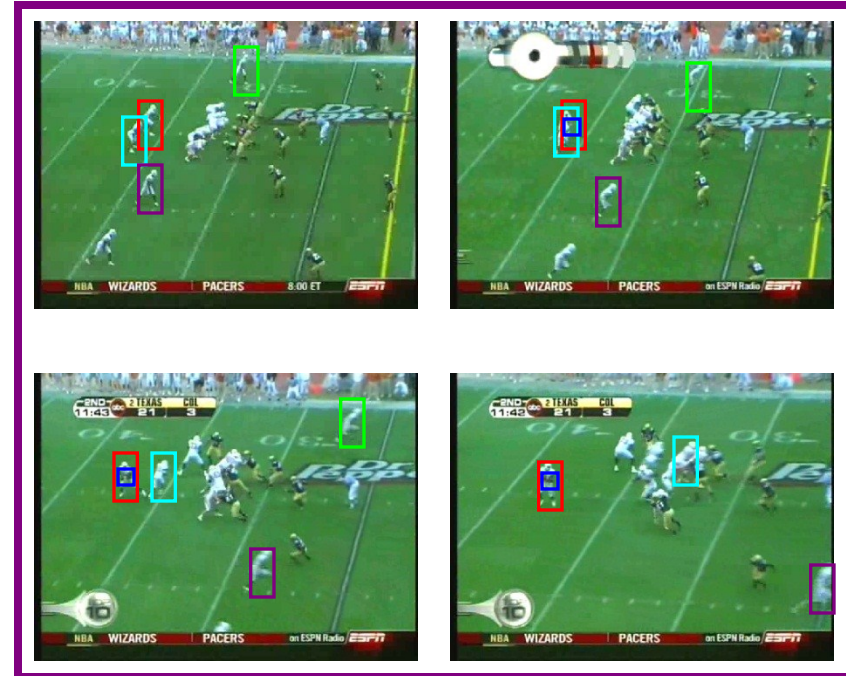
Task-aware intelligent workspace (assembly).

Ryoo et al., IJCAI 07, CVIU 10



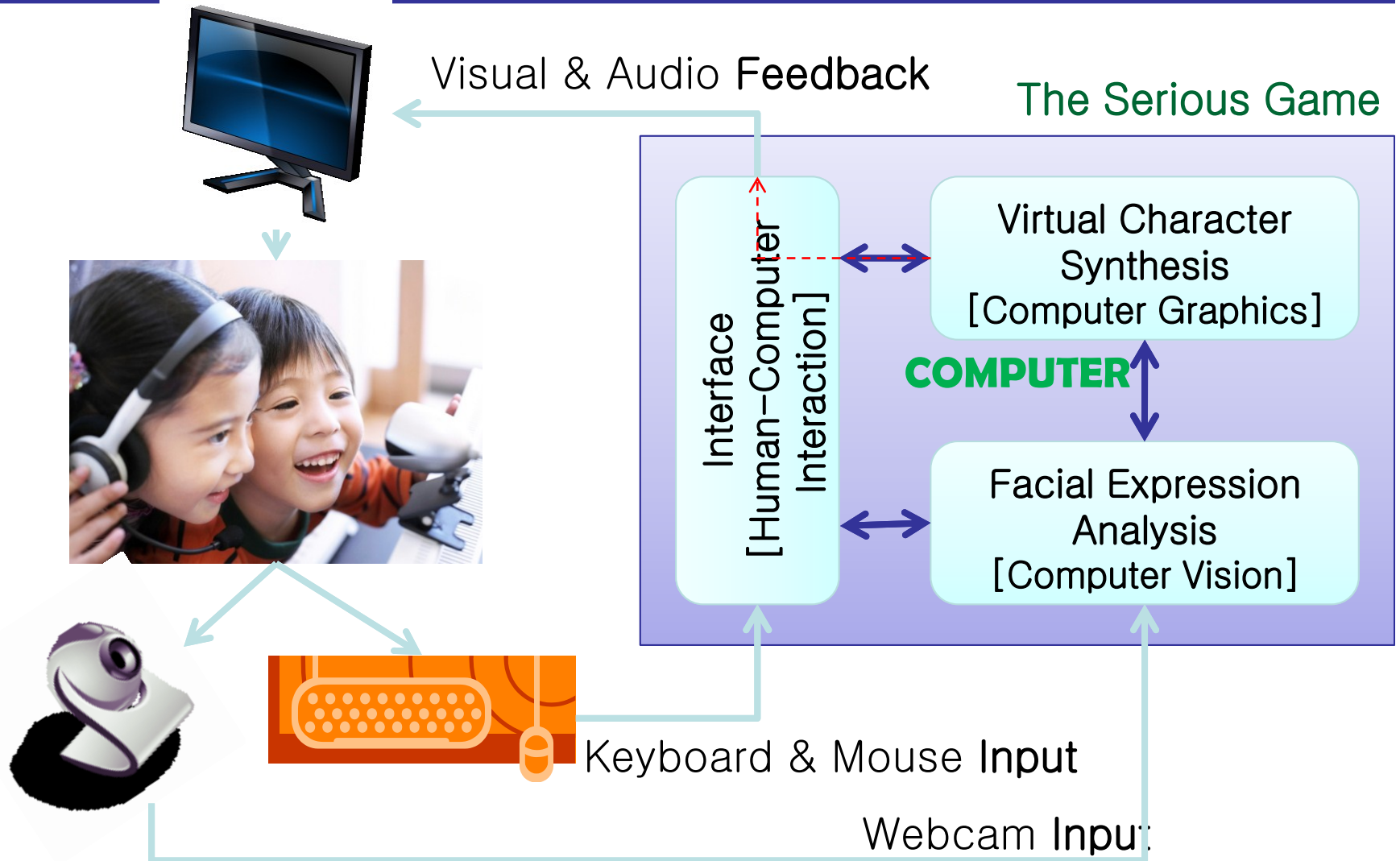
Sports play analysis

- Example: American football



Analyze what play this is!

A system to enable autistic children



Web-based video retrieval

- YouTube
 - 20 hours of videos uploaded every minute
 - Content-based search
 - Search based on contents of the video, instead of user-attached keywords
 - Example: search 'kiss' from long movies



Introduction

Types of activities

Levels of human activities

- Gestures
 - Atomic movements
- Actions
 - A single actor
- Interactions
 - Human-human interactions
 - Human-object interactions
- Group activities
 - Physical/conceptual groups

Human activities

- Categorized based on their complexity
 - Hierarchy
 - # of participants

Gestures:

Single body-part
movements

Atomic components



stretching, withdrawing, ...

Human activities

- Categorized based on their complexity
 - Hierarchy
 - # of participants

Actions:

Single actor
movements



bending, waving, ...

Human activities

- Categorized based on their complexity
 - Hierarchy
 - # of participants

Interactions:

Human-human/
human-object
interactions



punching, pushing, ...

Human activities

- Categorized based on their complexity
 - Hierarchy
 - # of participants

Group
activities:

Activities of
groups



group stealing, ...

Introduction

Challenges

Challenges – robustness

- Environment variations
 - Background
 - Moving backgrounds – trees



- Pedestrians
 - Occlusions
 - View-points – moving camera

Challenges – robustness

- Actor movement variations
 - Each person has his/her own style of executing an activity
 - Who stretches his hand first?
 - How long does one stay his hand stretched?



Challenges – various activities

- There are various types of activities
 - The ultimate goal is to make computers recognize all of them reliably.

gestures



actions



interactions



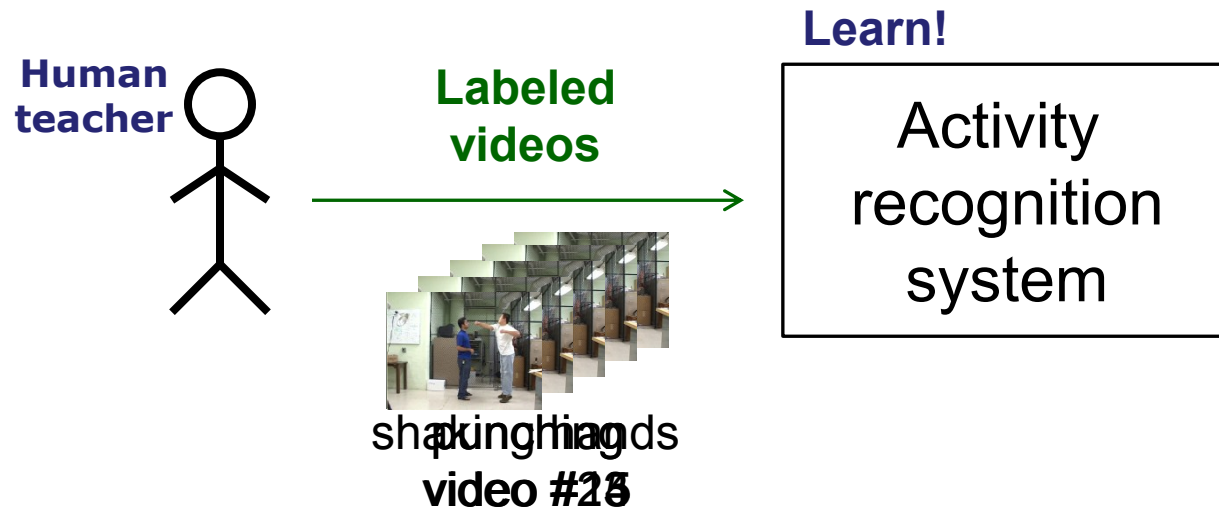
group activities



Levels of human activities

Challenges – learning

- Insufficient amount of training videos
 - Traditional setting: Supervised learning



- Human efforts are expensive!
 - Unsupervised learning
 - Interactive learning

This tutorial

- Targeted for broad CVPR audience
 - Assuming basic background in computer vision and machine learning
 - Not assuming significant activity recognition background
- Goal
 - State-of-the-arts of activity recognition
 - Past research progress and current research directions
 - Future challenges

Outline

- Introduction
 - Overview
- Single layered approaches
 - Sequences (HMMs)
 - Spatio-temporal features
- Hierarchical approaches
 - Syntactic/Statistical approaches
 - Description-based approaches
 - Human interactions, group activities
- Applications and challenges