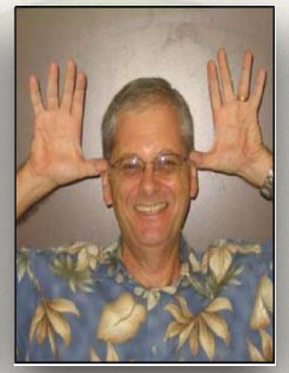


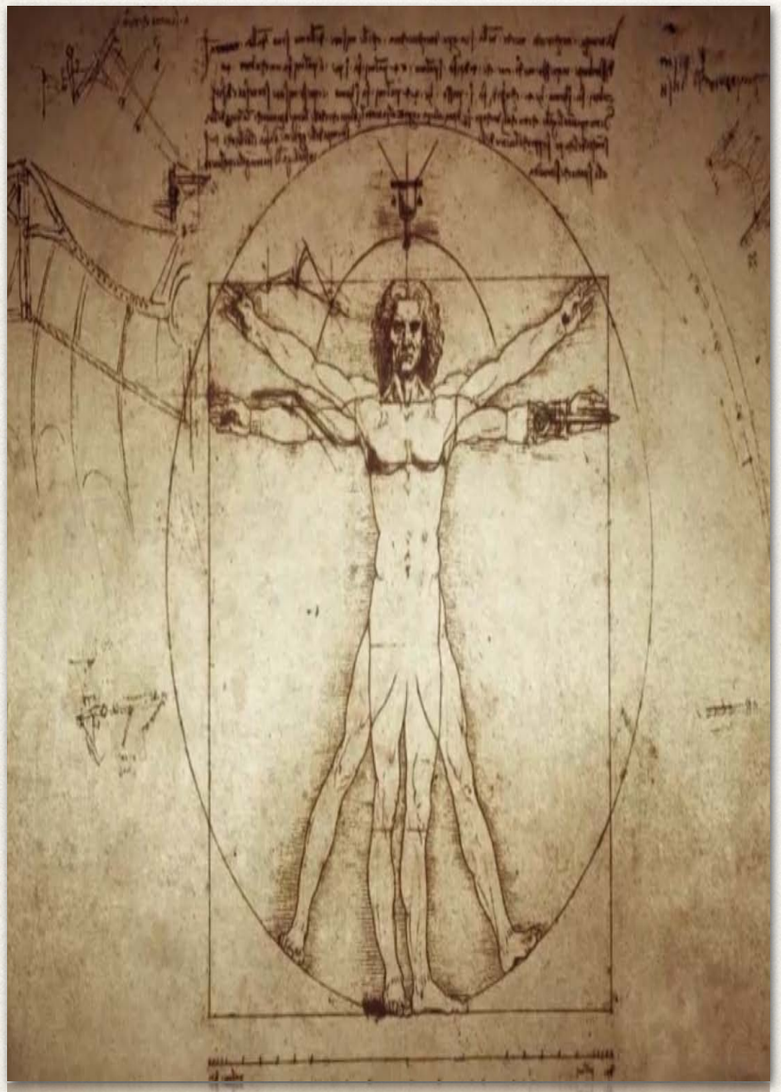
Computer Vision:

A Computational Intelligence Perspective – Part II

Derek T Anderson, James M Keller, Chee Seng Chan

24 July 2016

Tutorial Overview

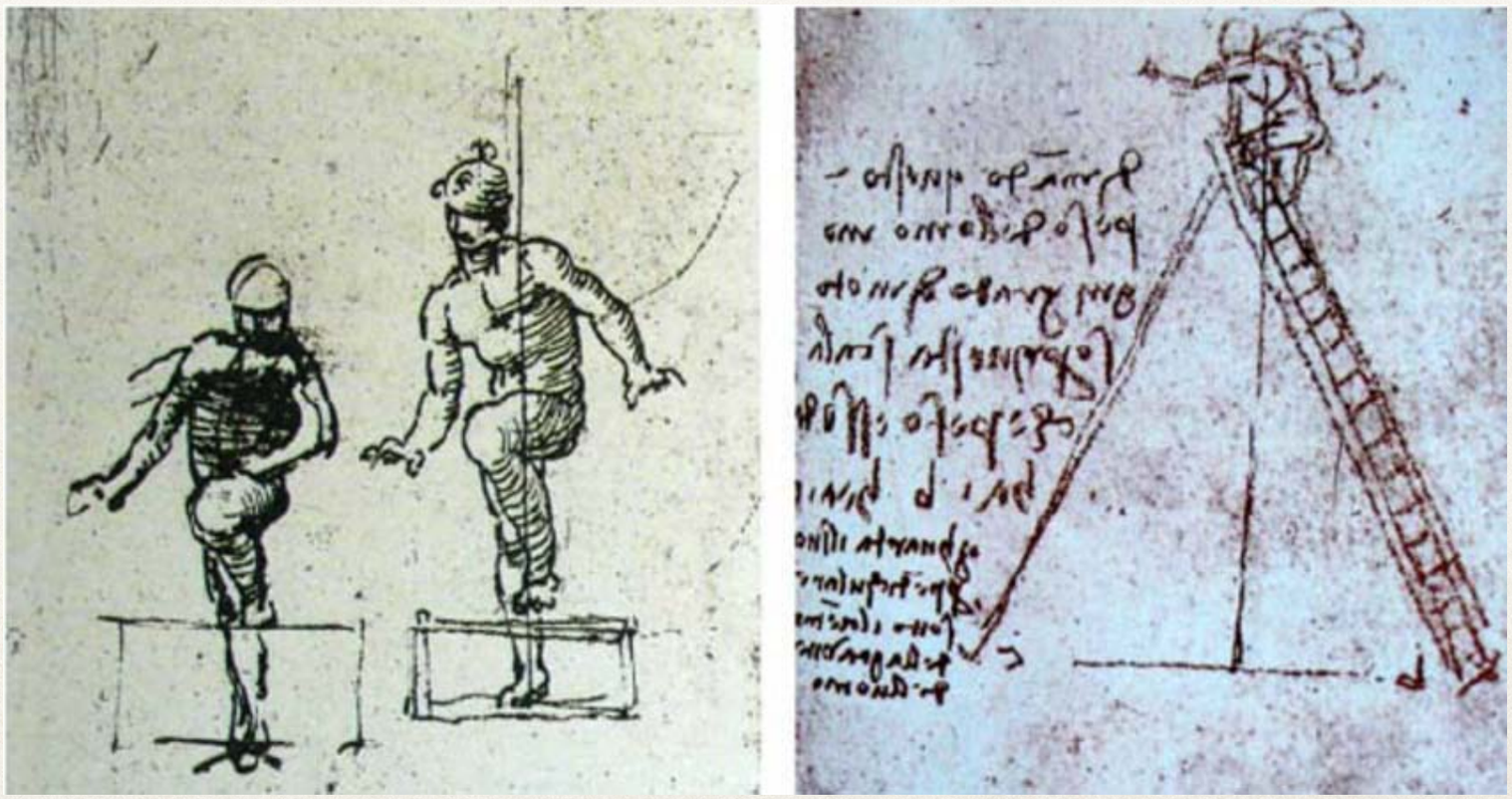


The Vitruvian Man,
Leornardo da Vinci, 1490

- Motivation
  - Historical review
  - Applications and challenges
- Appearance-based method
- Motion-based methods
- Deep-based methods
- Datasets

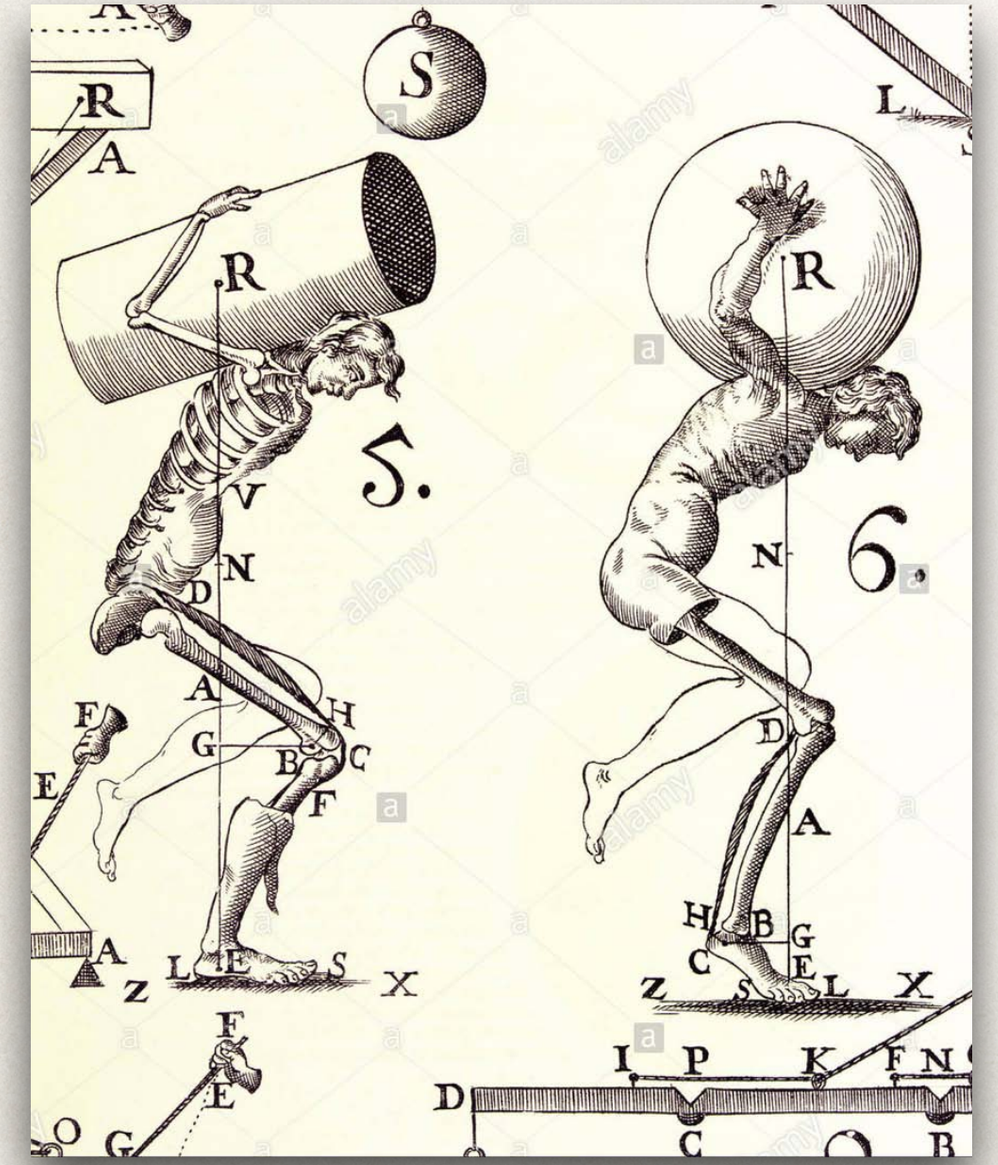
Motivation I: Artistic Representation

- Early studies were motivated by human representation in **Arts**:
- Da Vinci: *"it is indispensable for a painter, to become totally familiar with the anatomy of nerves, bones, muscles and sinews, such that he understand foe their various nations and stresses, which sinews or which muscle causes a particular motion".*



Leonardo da Vinci (1452-1519): A man going upstairs, or up a ladder

Motivation II: Biomechanics

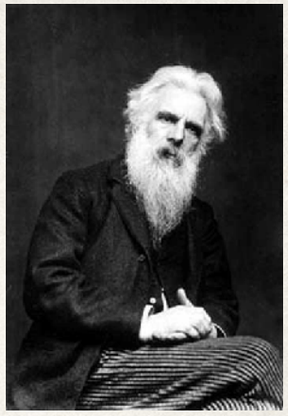


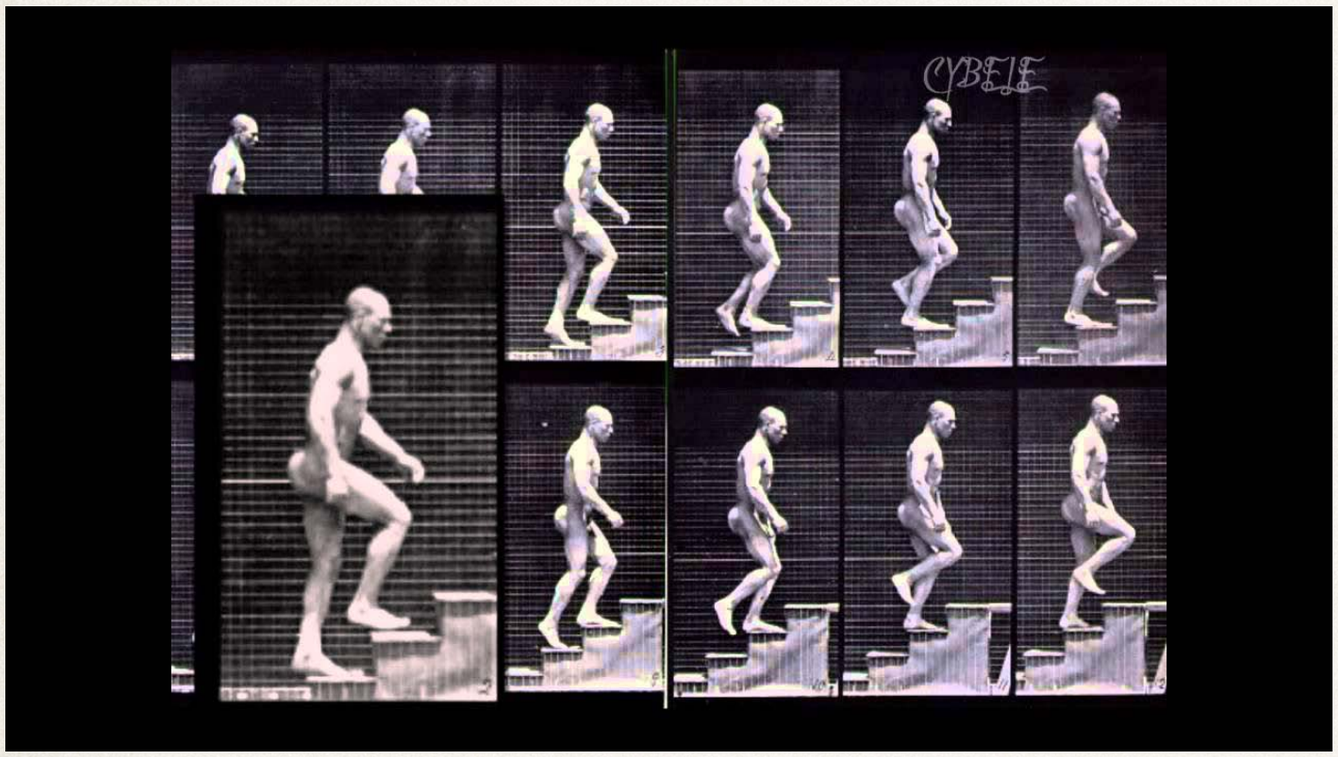
- The emergence of ***biomechanics.***
- Borelli applied to biology the analytical and geometrical methods, developed by Galileo Galilei.
- He was the first to understand that bones serve as levers and muscles function according to mathematical principles.

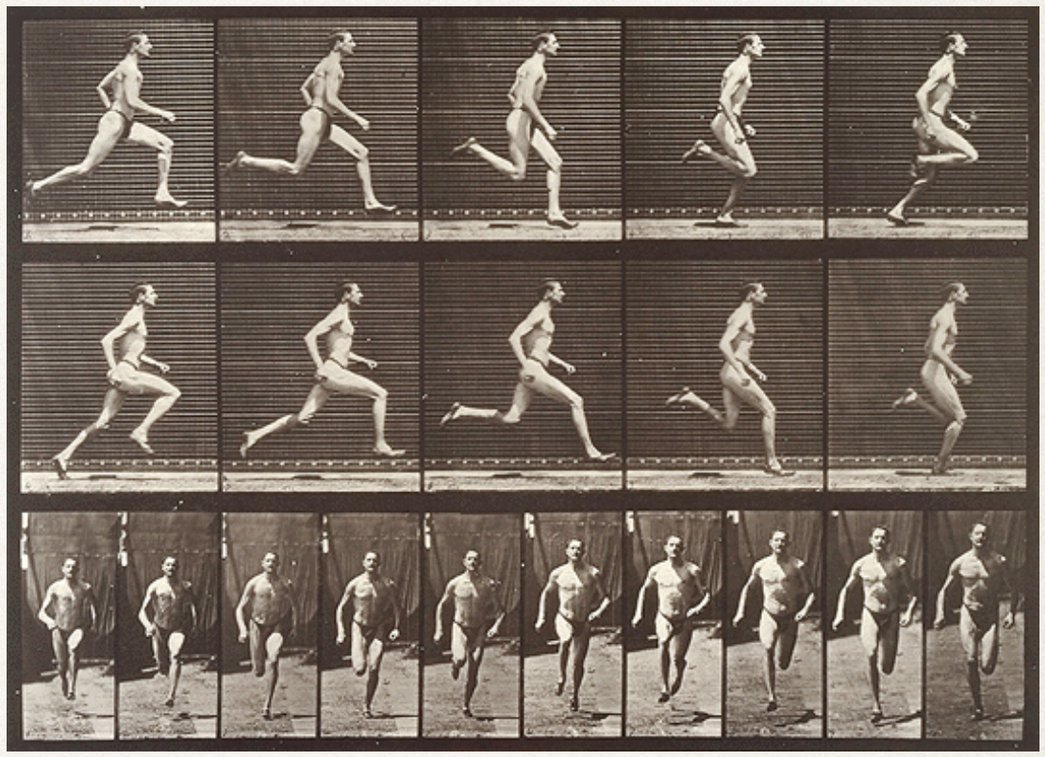
Giovanni Alfonso Borelli (1608-1679) : De Motu Animalium (On Animal Motion)

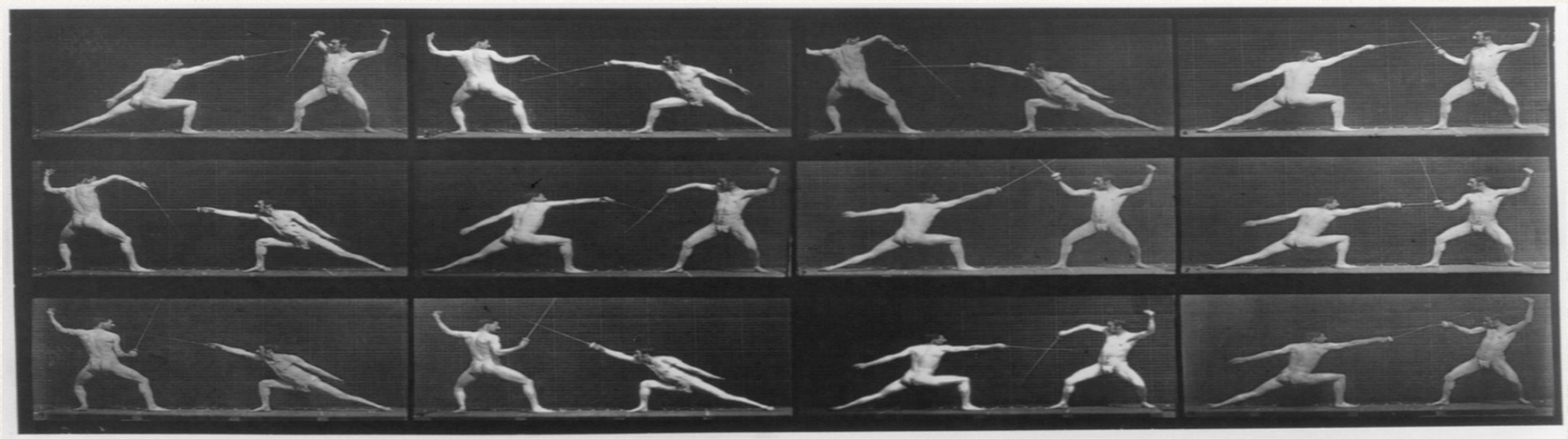
Motivation III: Motion Perception

Eadweard Muybridge (1830 - 1904)









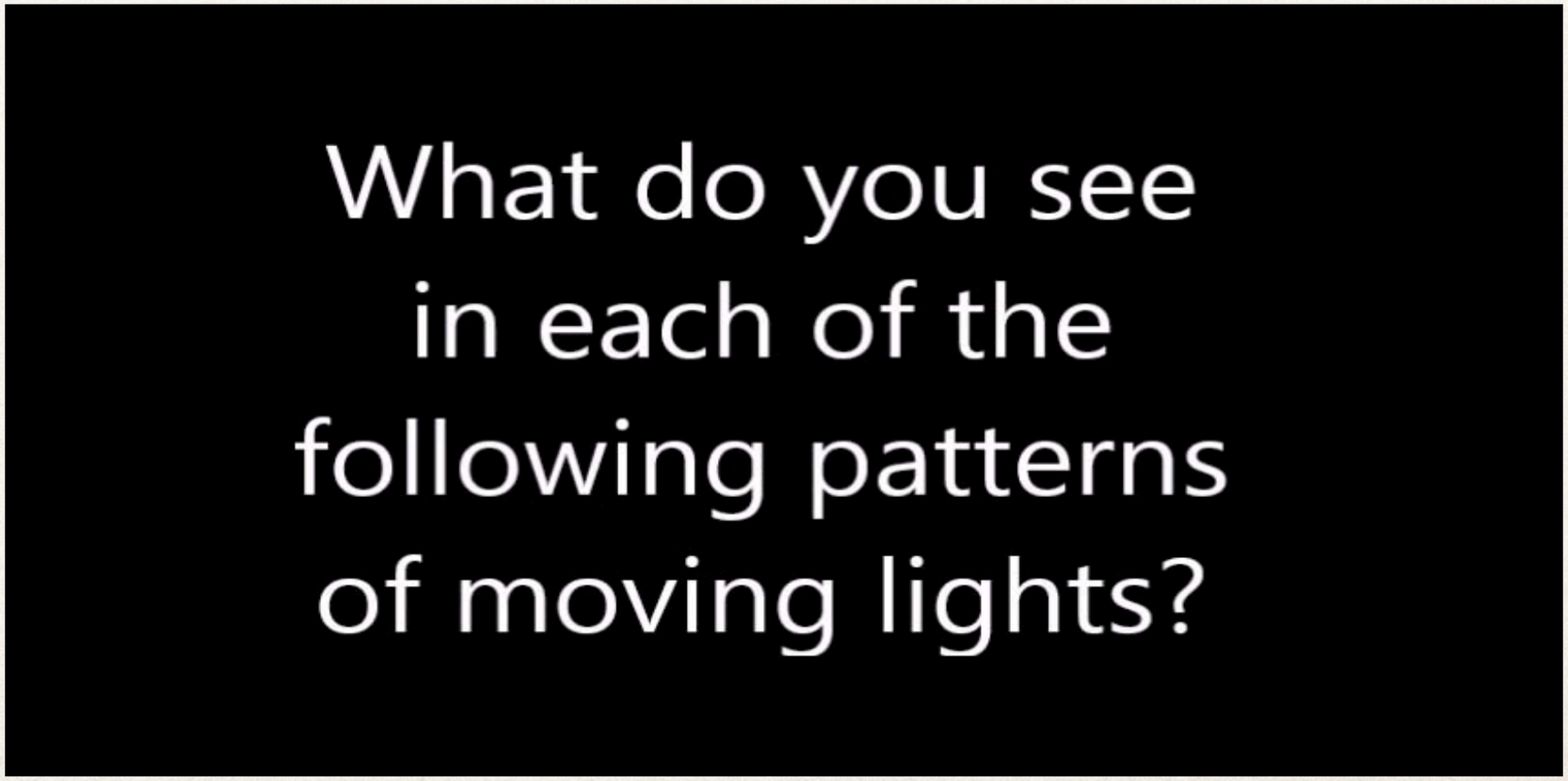
Motivation III: Motion Perception

Étienne-Jules Marey (1830-1904)

Motivation III: Motion Perception

Gunnar Johannson

- Gunnar Johannson (1971) pioneered studies on the use of image sequence for a programmed human motion analysis.
- **Moving Light Display (LED)** enable identification of familiar people and the gender and inspired many works in computer vision.

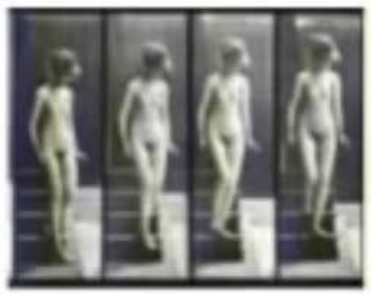


Gunnar Johansson, ***Perception and Psychophysics****, 1973*

Summary

15th century studies of anatomy

17th century emergence of *biomechanics*



19th century emergence of *cinematography*

1971 studies of human motion perception

Modern computer vision

Applications

Security

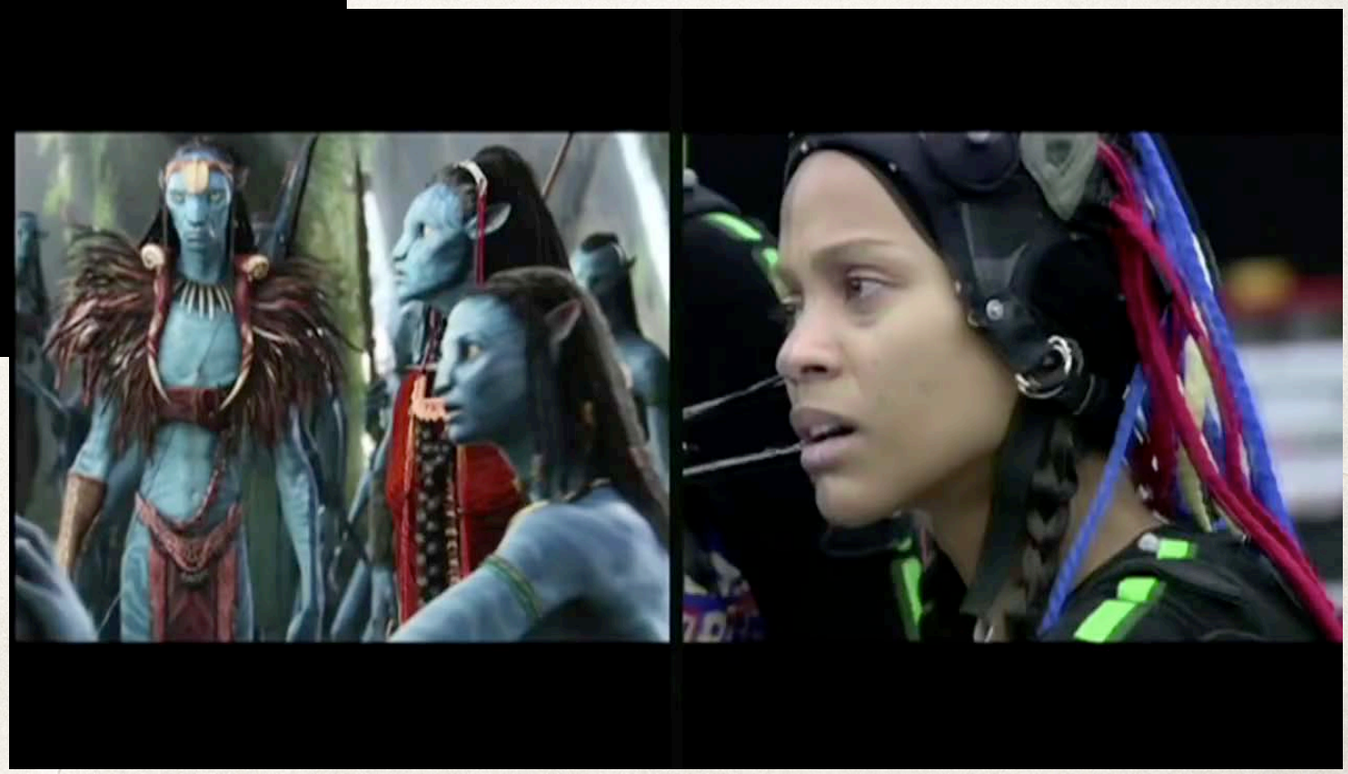
Applications

Motion Capture

- ***Example: Film***

The Hobbit - The Unexpected Journey, 2012

Avatar. 2009



Applications

Motion Capture

- ***Example: Film***

Applications

Sports Analysis

- ***Example: Football/Soccer***

Other Applications

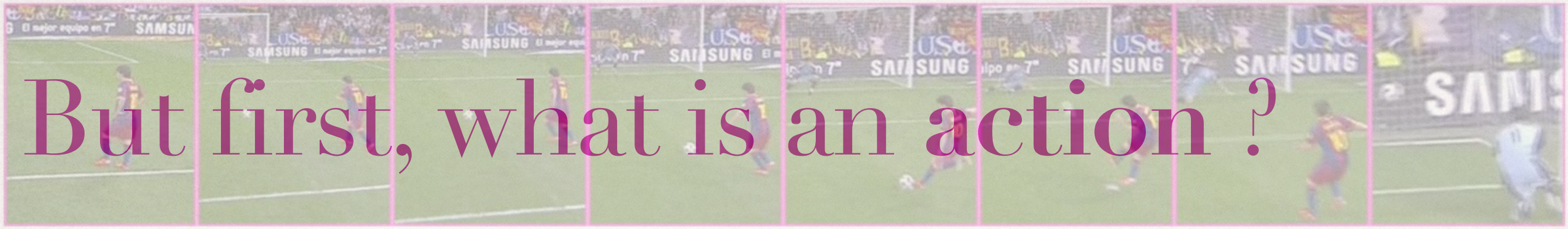
TV & Web:
e.g.
"Fight

Home
entertainment

Sociology research:

Manually
analyzed smoking
actions in
900 movies

Surveillance:
260K views
in 7 days on
YouTube



But first, what is an **action** ?

Human motions extend from the simplest movement of a limb to complex joint movement of a group of limbs and body.

- Moeslund and Granum (CVIU, 2006); Poppe (IMAVIS, 2010) define *action primitives* as **"an atomic movement that can be described at the limb level"**. Accordingly, the term *action* defines a diverse range of movements, from "simple and primitive ones" to "cyclic body movements". For instance, left leg forward is an action primitive of running.
- Turaga et al. (T-CSVT, 2008) define *action* as **"simple motion patterns usually executed by a single person and typically lasting for a very short duration (Order of tens of seconds)".** For example, actions are walking or swimming, activities are two persons shaking hands or a football team scoring a goal.
- Wang et al. (CVPR, 2016); Lim et al (PR, 2016) suggested that the true meaning of an *action* lies in **"the change or transformation an action brings to the environment", *e.g.*, kicking a ball.**
- In the Oxford Dictionary, *action* is defined as **"the fact or process of doing something, typically to achieve an aim". and *activity* is "a thing that a person or group does or has done".**

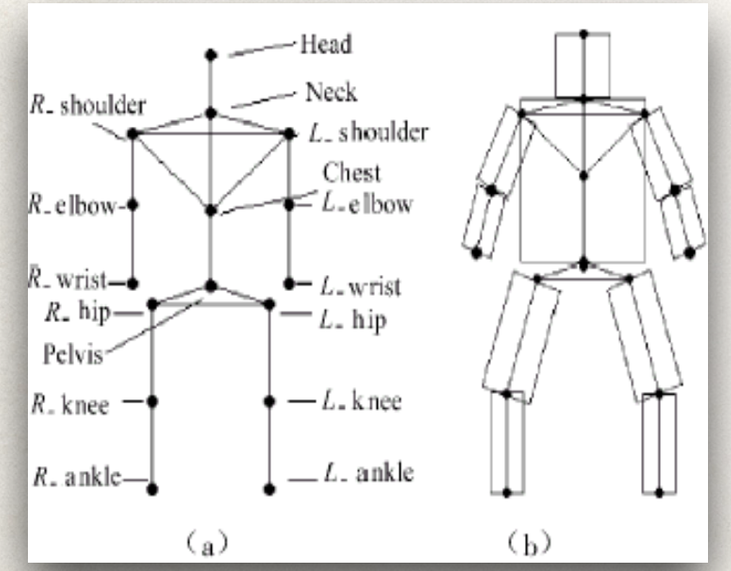
"Action is the most elementary human-surrounding interaction with a meaning."

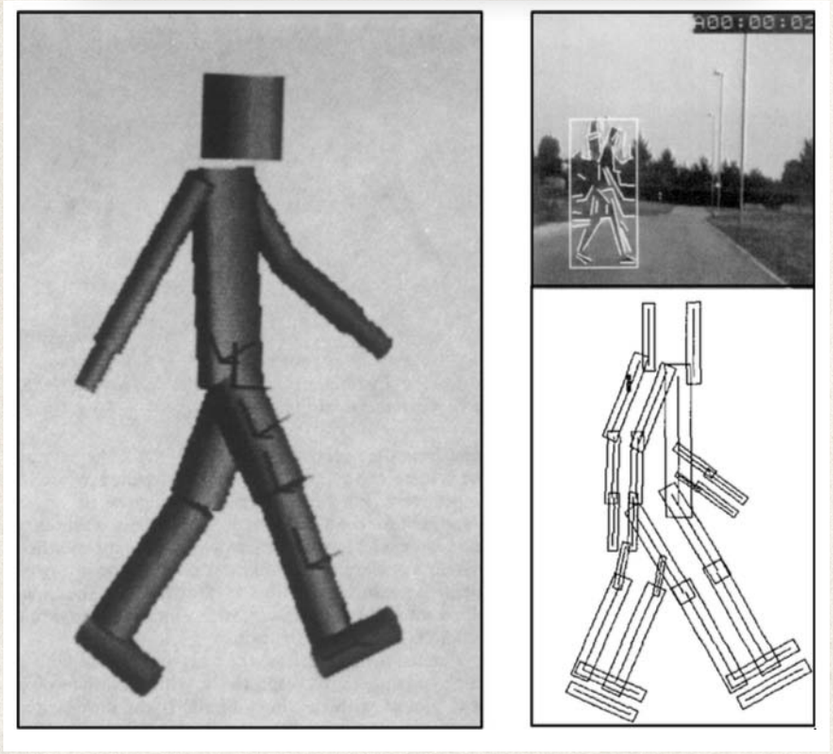
Tutorial Overview

The Vitruvian Man,
Leornardo da Vinci, 1490
Florence, Tuscany, Italy

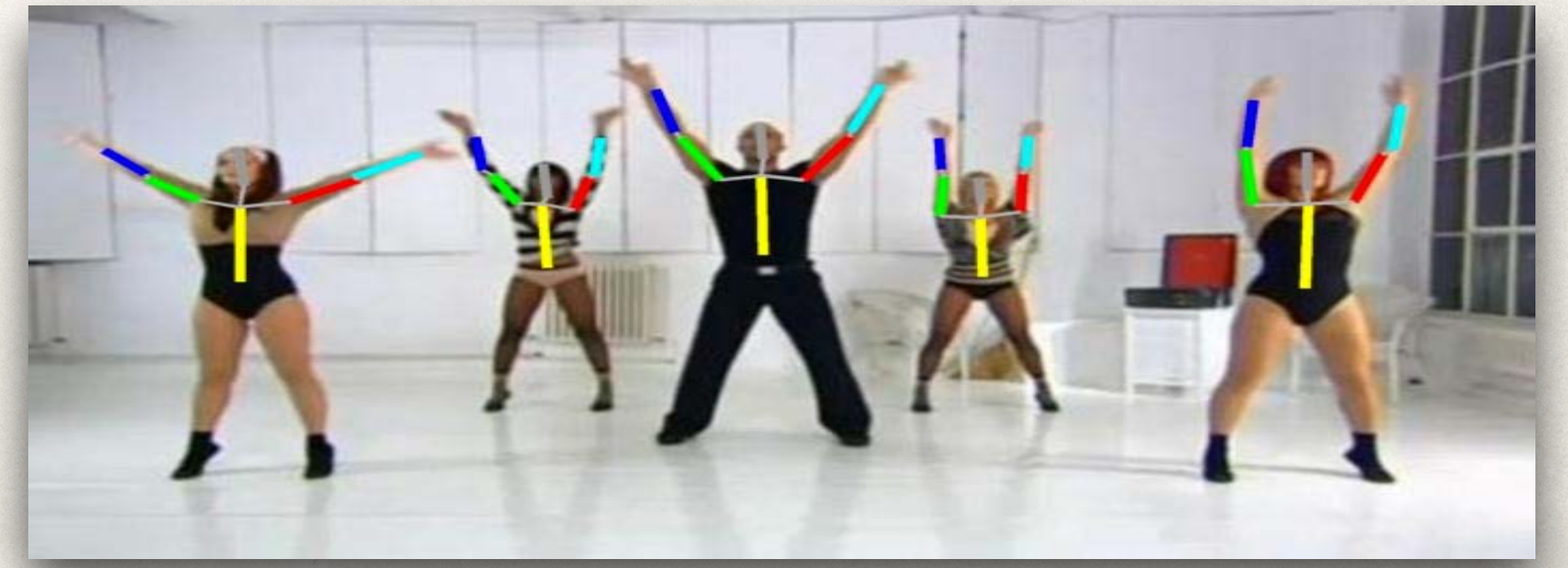
- Motivation
  - Historical review
  - Applications and challenges
- **Appearance-based method**
- Motion-based methods
- Deep-based methods
- Datasets

Appearance-based models





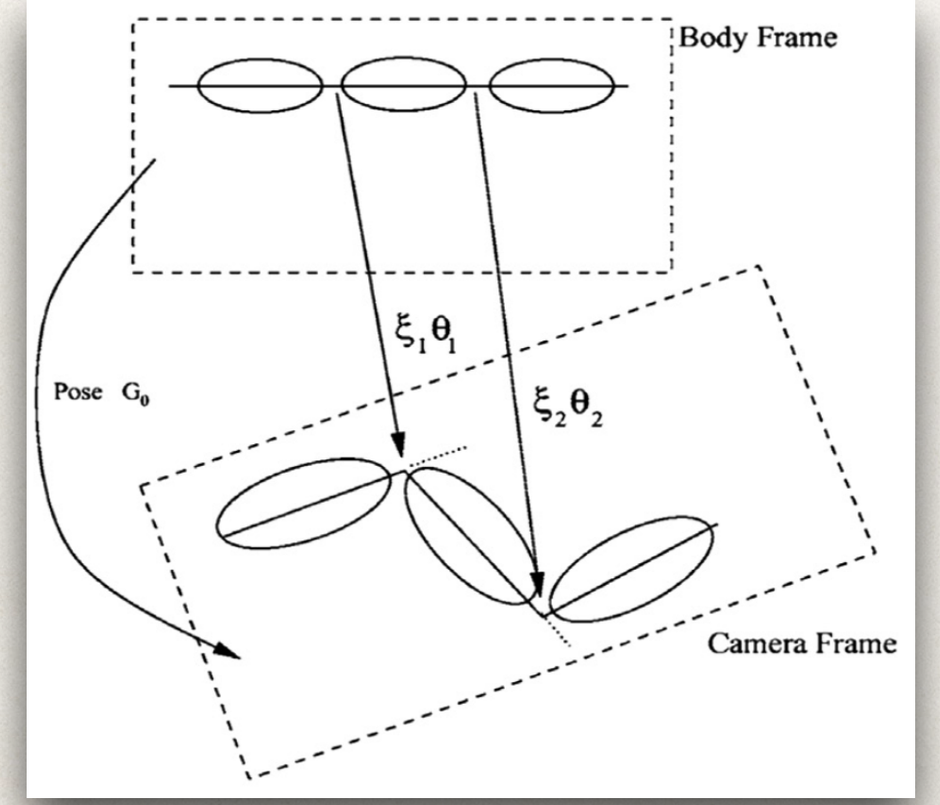
- One the earliest work in action recognition make use of 2D/3D models to describe actions.
- The notable example is the **WALKER** hierarchical model introduced in Hogg (1983). Other examples included connected cylinders in Rohr (1994), skeletonization, kinematic chain.

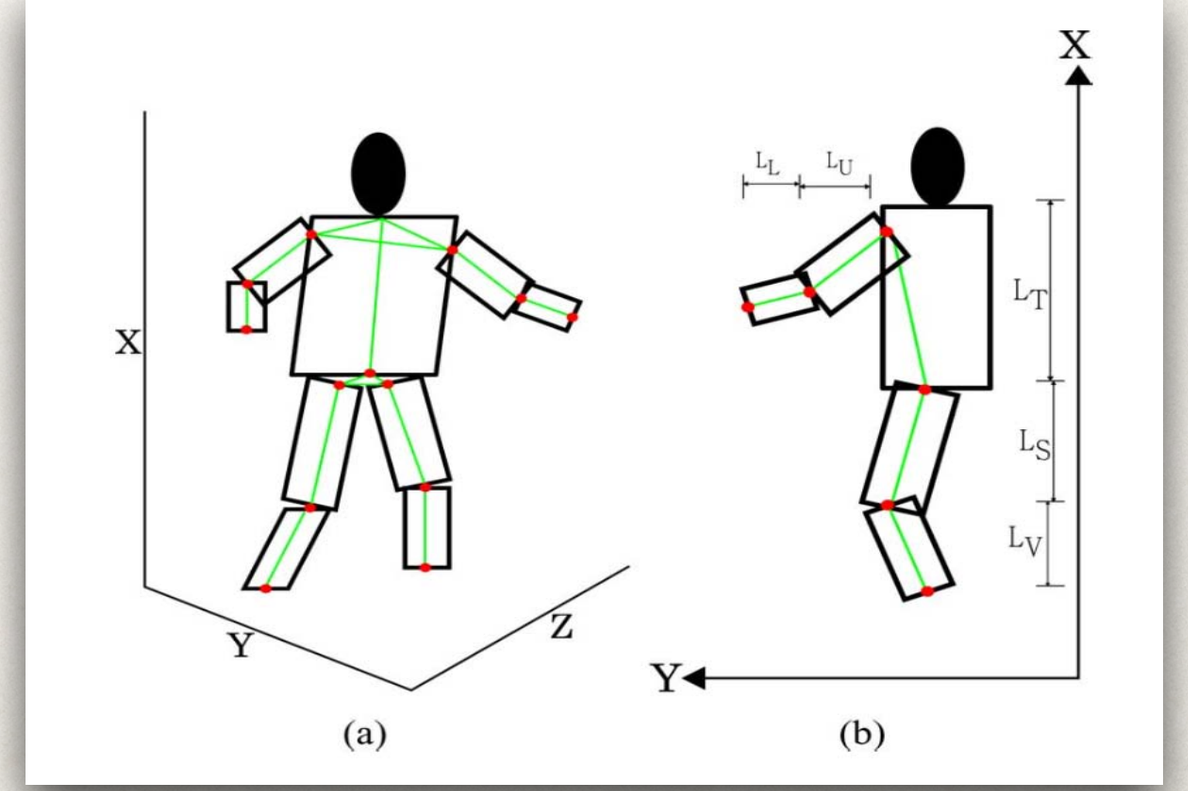


Appearance-based models

Kinematic Model (Fuzzy)

- The idea is human body parts do not move independently.
- So we can use a kinematic chain to built the human model.

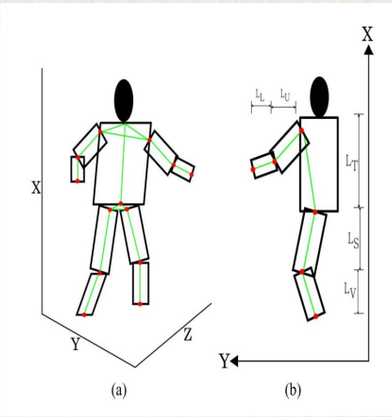




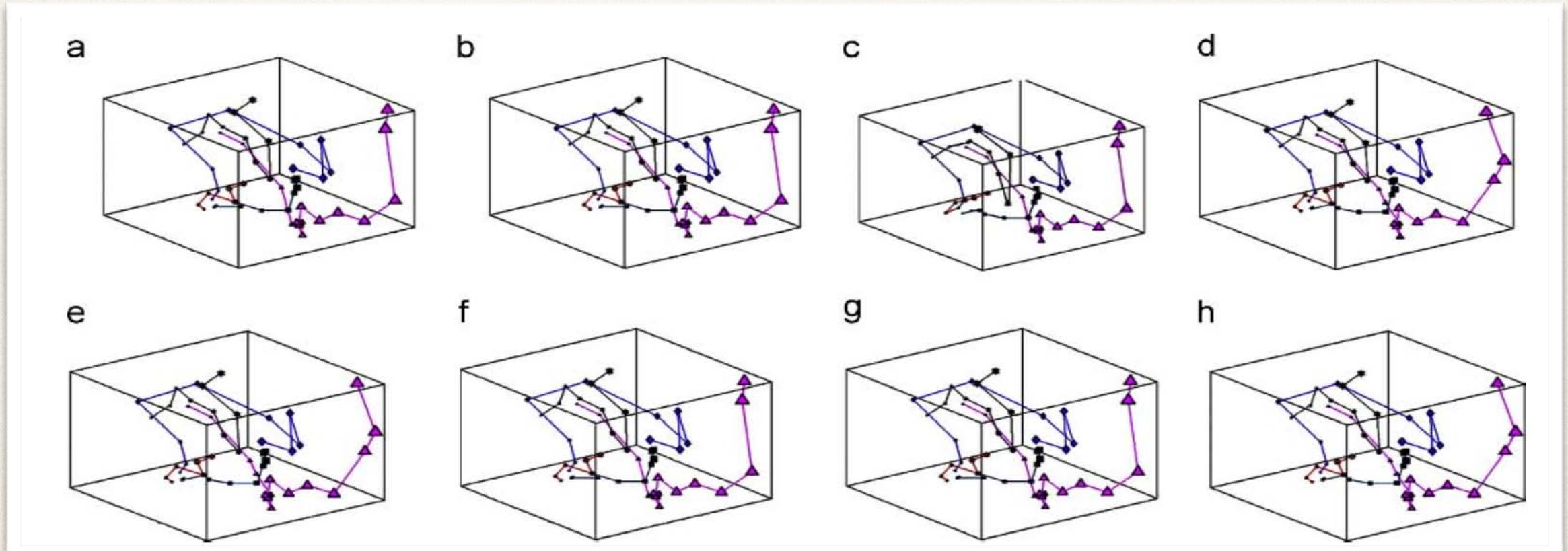
Liu, (2008) *Fuzzy qualitative robot kinematics*, T-FS, vol. 16(6), pp. 1522–1530.
Chan & Liu (2009) *Fuzzy qualitative human motion analysis*, T-FS, vol. 17(4), pp. 851–862.

Appearance-based models

Kinematic Model (Fuzzy)



- Proposed model: Qualitative Normalised Template (QNT)

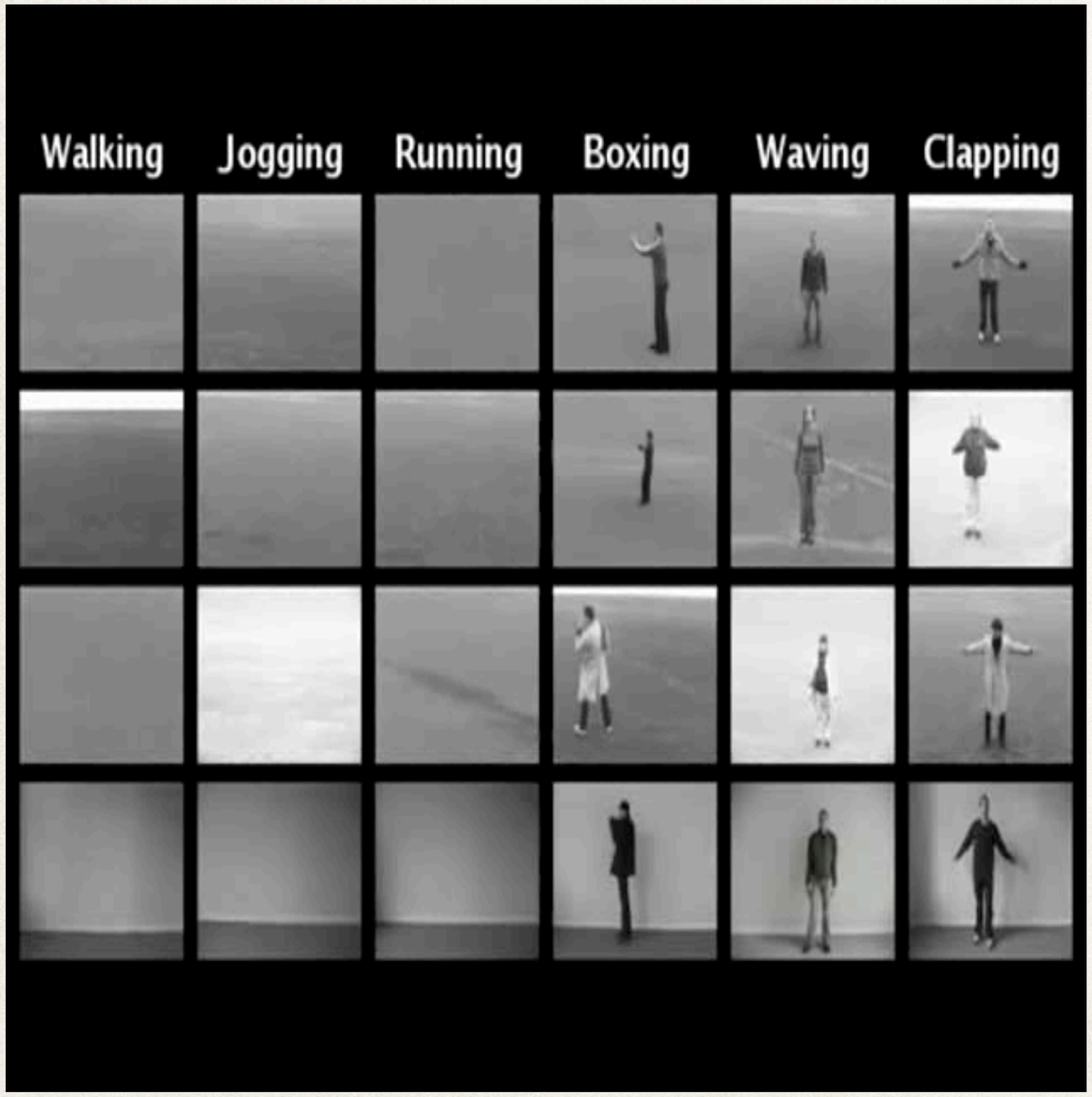


Chan & Liu (2009) ***Fuzzy qualitative human motion analysis*****, T-FS, vol. 17(4), pp. 851–862.**

Appearance-based models

Kinematic Model (Fuzzy) - Results on KTH Dataset

- **KTH Database** – compromise of 25 adults, 6 types of activities in 4 profile view



Appearance-based models

Kinematic Model (Fuzzy) - Results on Weizmann Dataset

- **Weizmann Action Dataset** – compromise of 10 adults, 10 activities in planar view

Appearance-based models

KTH + Weizmann Dataset

- KTH Dataset

Method	QNT (Chan & Liu - TFS 2009)	HMM	FHMM	FVQ (Iosifidis et al - TCSVT 2013)	CV (Sapienza et al - IJCV 2014)
Precision	85%	54%	62%	93.52%	**96.76%**

- Weizmann Dataset

Method	QNT (Chan & Liu - TFS 2009)	HMM	FHMM	CV
Precision	**100%**	75%	84%	**100%**

Appearance-based models

Limitation of 2D/3D model

- Capturing accurate 2D/3D models is **difficult** and **expensive**.
- This is why researcher avoid 2D/3D modeling and instead opt for representing actions at **holistic** or **local level**.

Appearance-based models

Holistic-level

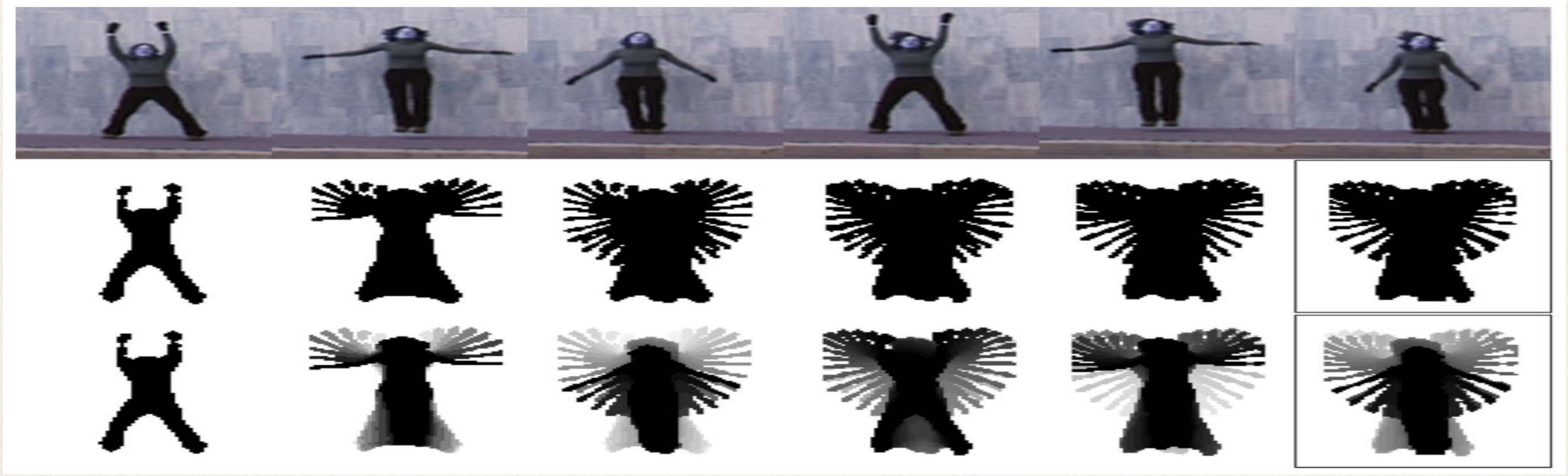
- One of the most simplest method : ***image differencing*** (or simply known as the foreground segmentation)

- Better background / foreground separation methods exists:
  - Modelling color variation at each pixel with Gaussian mixture
  - Motion layer separation for scenes with non-static background

Appearance-based models

Holistic-level

- Influential work: *Motion Energy Image (MEI) and Motion History Image (MHI)* introduced by Bobick and Davis (T-PAMI, 2001)

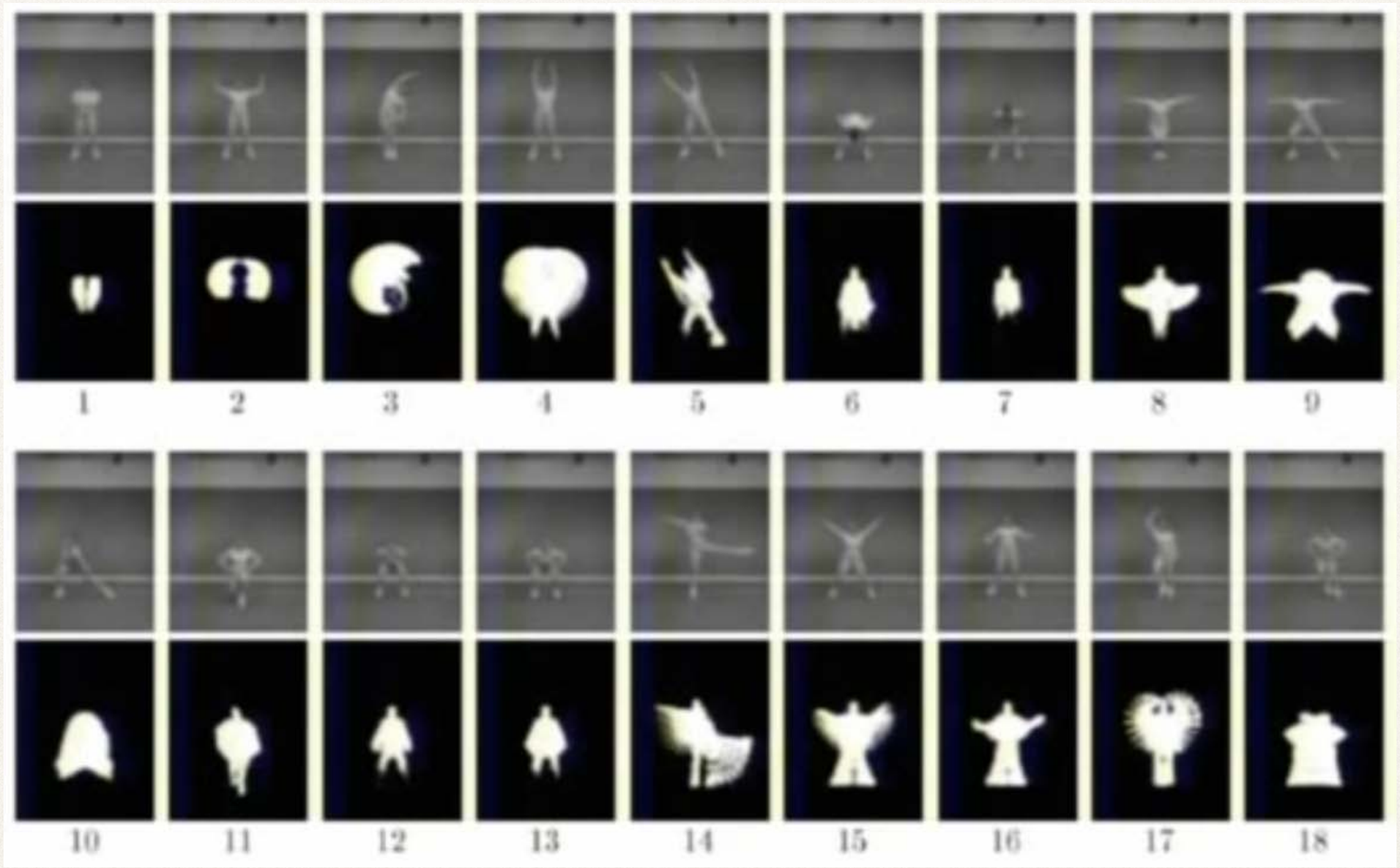


- **Idea**: *Summarize motion* in a video. So that, the MEI template is a binary image describing where the motion happens and defined as:

$$\boldsymbol{E}_\tau(x,y,t) = \bigcup_{i=0}^{\tau-1} \boldsymbol{D}(x,y,t-i)$$

Appearance-based models

Holistic-level (Results) - Aerobics Dataset

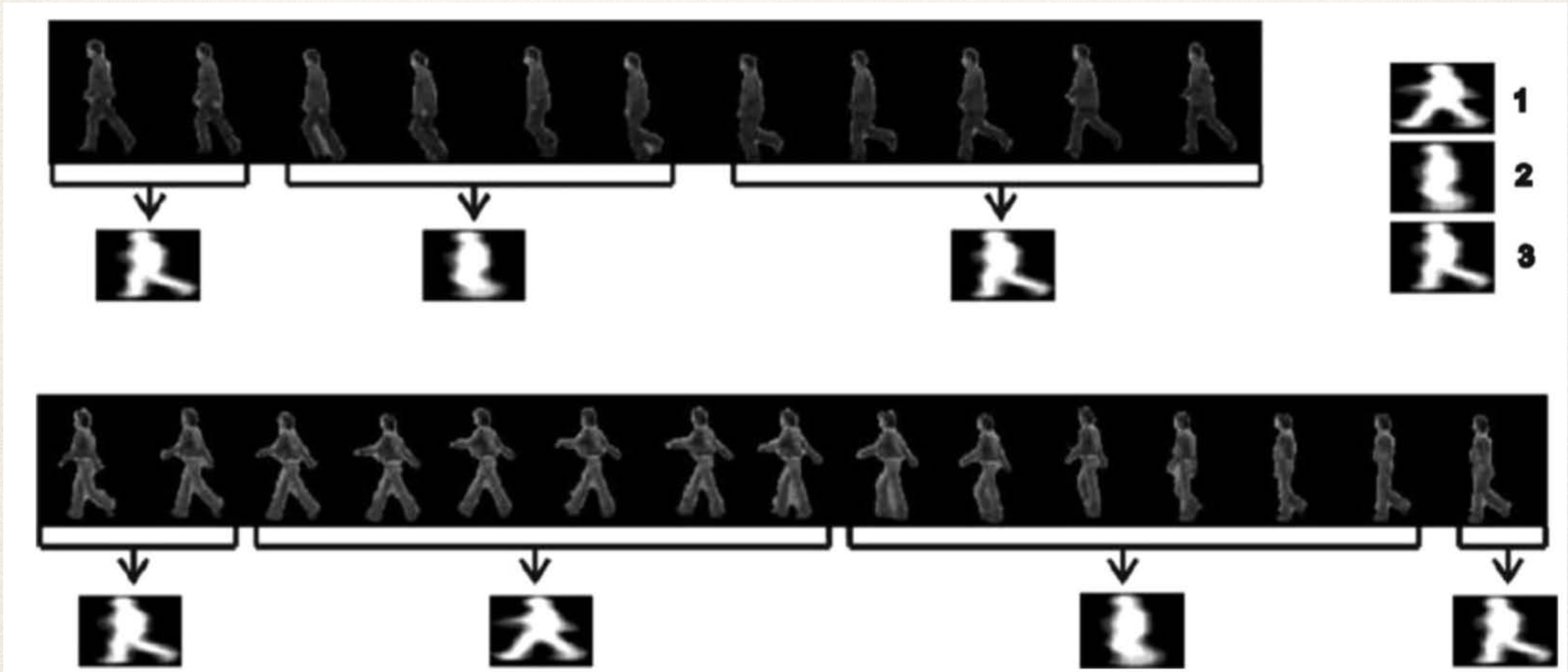


Nearest neighbor classifier: 66%

Appearance-based models

Holistic-level (Fuzzy)

- **Dyneme**: the basic movement patterns of a continuous action using *fuzzy c-means* was introduced by Gkalelis et al. (T-CSVT, 2008)



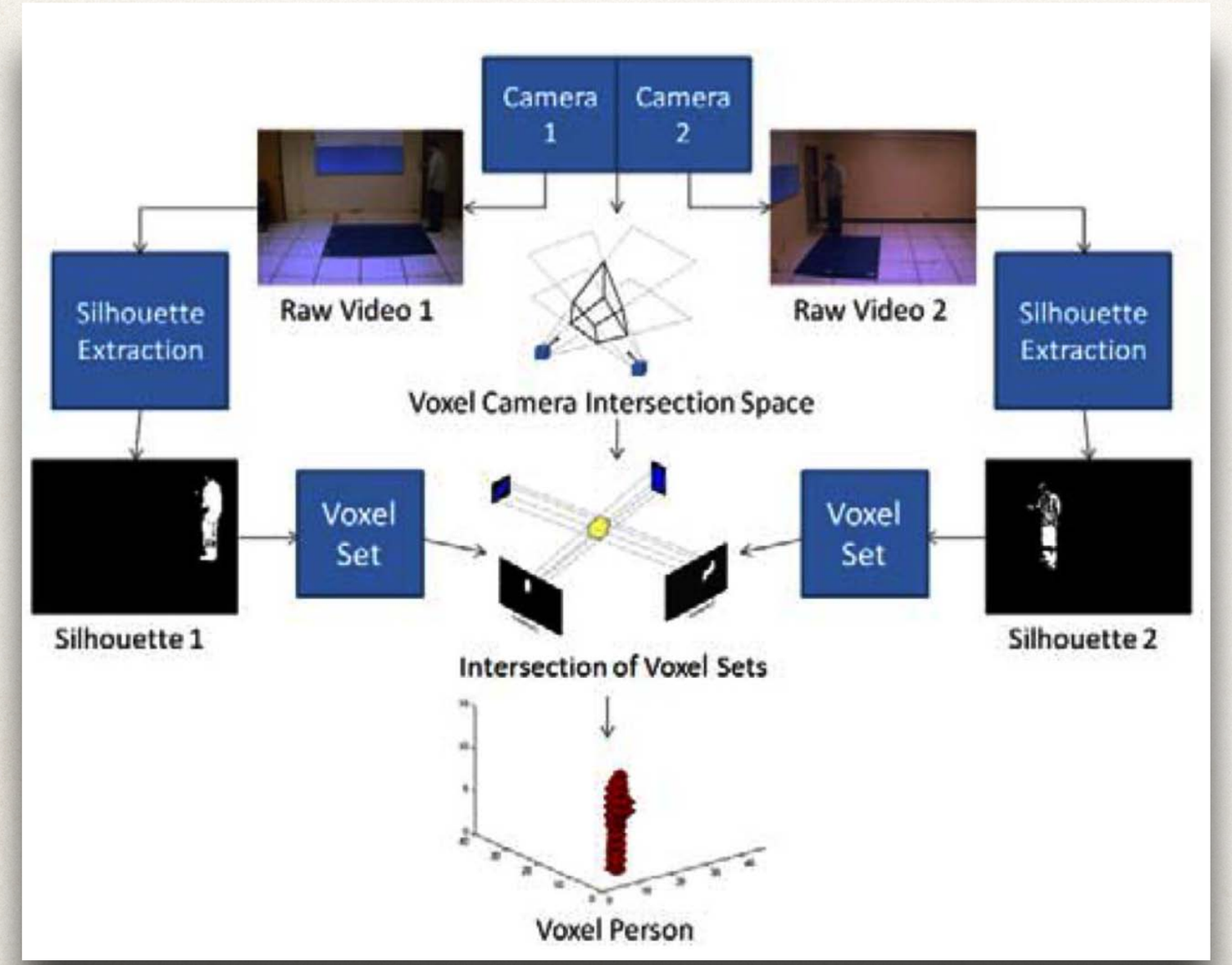
Appearance-based models

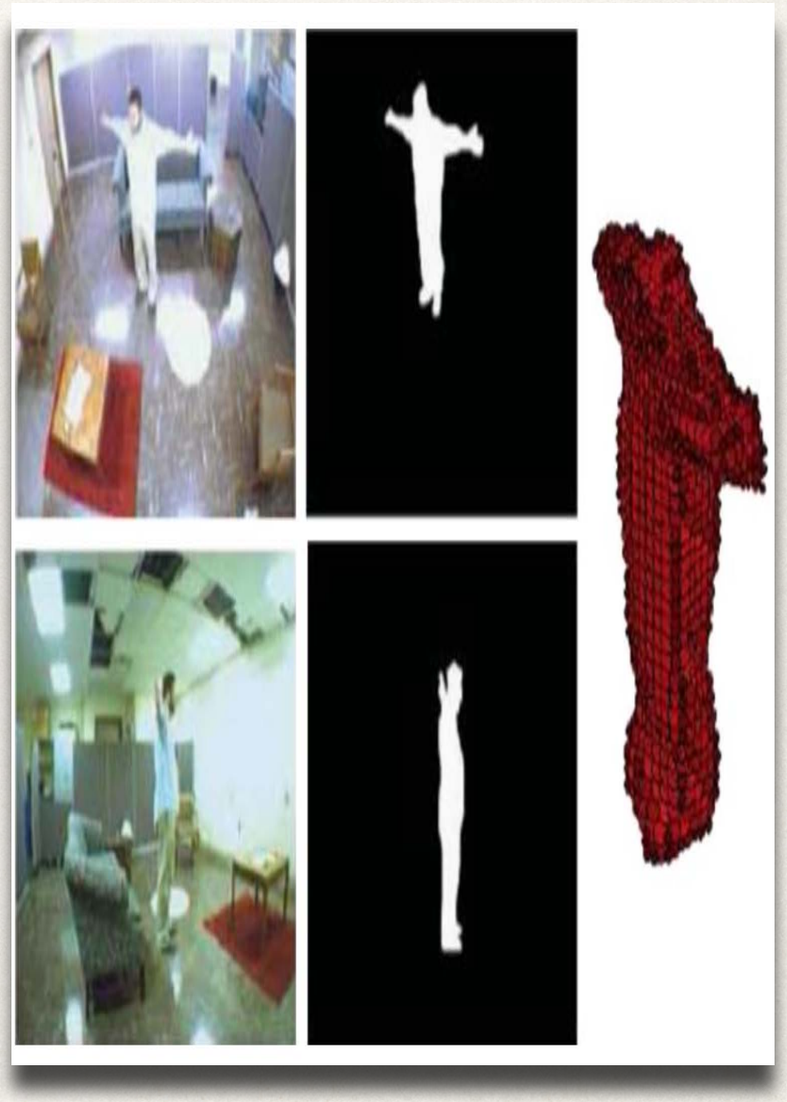
Holistic-level

- **Volumetric MEI templates** was introduced by *Blank et al* (ICCV, 2005). The main idea is to represent an action by a 3D shape induced from its silhouettes in the space-time:

Appearance-based models

Holistic-level (Fuzzy)



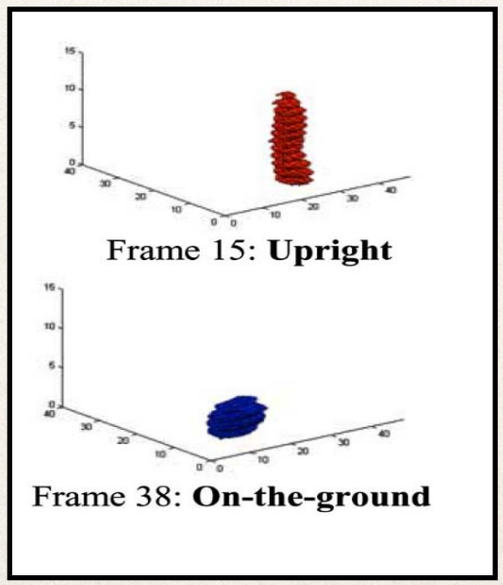


- Anderson et al (2009) *Modeling Human Activity From Voxel Person using Fuzzy Logic*, T-FS, vol. 17(1), pp. 39-49.
- Anderson et al (2009) *Linguistic summarization of video for fall detection using voxel person and fuzzy logic*, CVIU, vol. 113(1), pp. 80-89

Appearance-based models

Holistic-level (Fuzzy) - Assistive Living

- Modeling and monitoring human activity from video, in particular, ***elderly falls***.
- Fuzzy rules for state modeling is built: 1) centroid, 2) eigenheight and 3) similarity the voxel person primary orientation and ground plane normal.



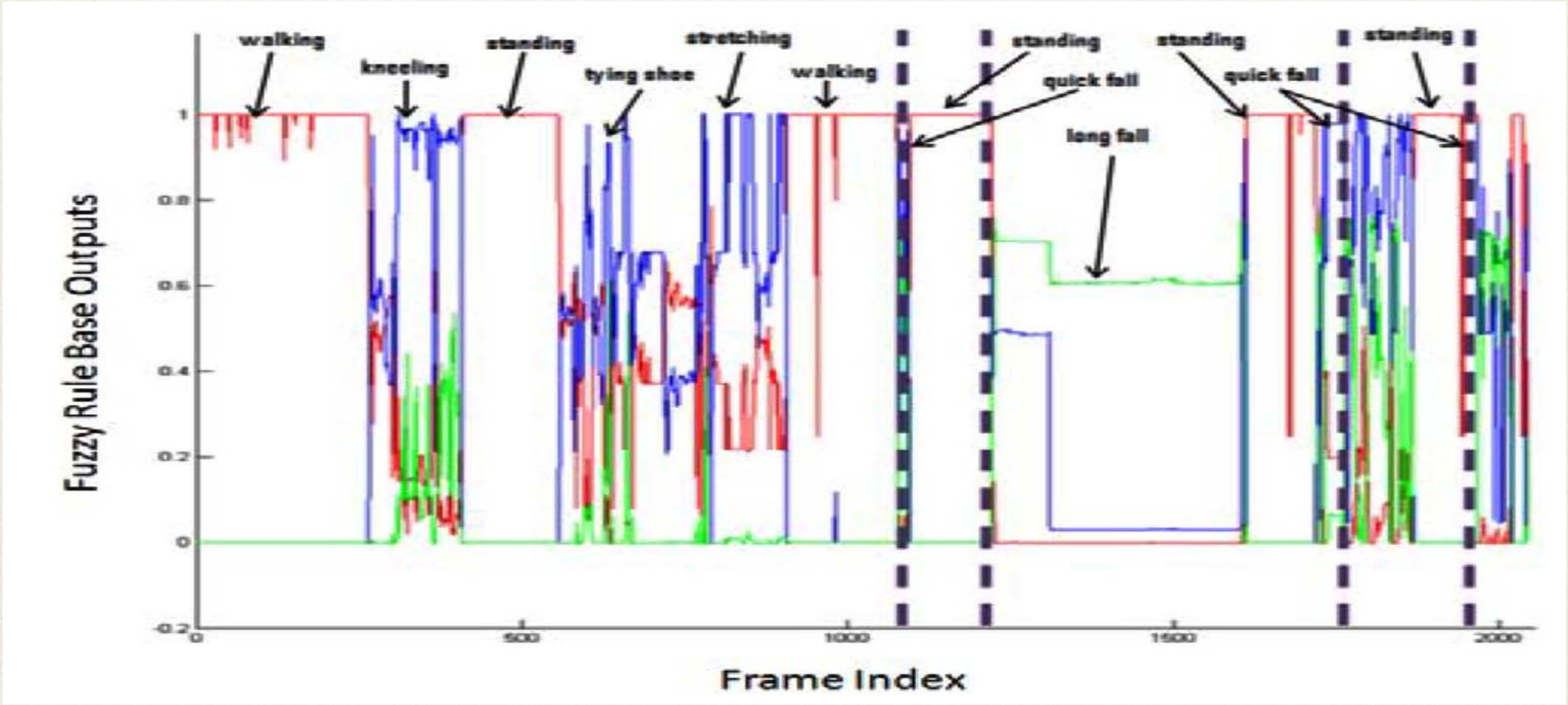
Rule		Centroid	Eigen Height	Normal Similarity		Upright	In Between	On the Ground
1		H	H	H		L	V	V
2		M	H	H		L	L	V
3		L	H	H		V	L	L
4		H	M	H		V	H	V
5		M	M	H		V	H	L
6		L	M	H		V	H	H
7		M	L	H		V	L	H
8		L	L	H		V	V	M
9		H	H	M		L	V	V
10	If	M	H	M	Then	L	L	V
11		L	H	M		L	H	V
12		H	M	M		L	H	V
13		M	M	M		L	H	V
14		L	M	M		V	H	L
15		L	M	L		V	L	H
16		L	L	M		V	L	M
17		H	H	L		H	V	V
18		M	H	L		M	V	V
19		L	H	L		L	L	V
20		H	M	L		M	L	V
21		M	M	L		L	L	V
22		L	M	L		L	H	V
23		M	L	L		V	H	L
24		L	L	L		V	L	H

Anderson et al (2009) Modeling Human Activity From Voxel Person using Fuzzy Logic, TFS, vol. 17(1), pp. 39-49.
Anderson et al (2009) Linguistic summarization of video for fall detection using voxel person and fuzzy logic, CVIU, vol. 113(1), pp. 80-89

Appearance-based models

Holistic-level (Fuzzy) - Assistive Living (Result)

- 11 minutes recorded video with 2042 frames



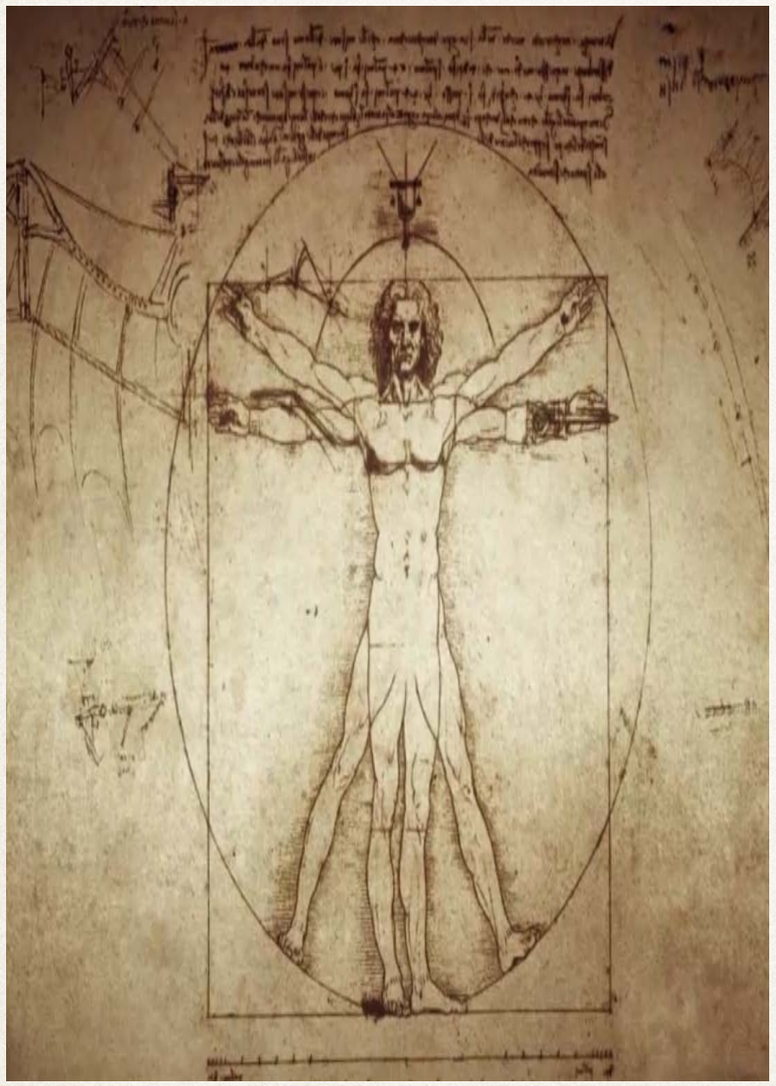
- Anderson et al (2009) *Modeling Human Activity From Voxel Person using Fuzzy Logic*, T-FS, vol. 17(1), pp. 39-49.
- Anderson et al (2009) *Linguistic summarization of video for fall detection using voxel person and fuzzy logic*, CVIU, vol. 113(1), pp. 80-89

Summary

- Holistic approaches flooded the research in action recognition roughly between ***1997-2007***.
- **Pros:**
  - Simple and fast
  - Works in controlled-settings (environment)

- **Cons:**
  - Prone to errors of background subtractions
  - Too rigid to capture possible variations of actions (e.g. view point, appearance, occlusion)
  - Does not capture *fine details*.

Tutorial Overview



The Vitruvian Man,
Leornardo da Vinci, 1490

- Motivation
  - Historical review
  - Applications and challenges
- Appearance-based methods
- **Motion-based methods**
- Deep-based methods
- Datasets

Motion-based Methods

Optical Flow

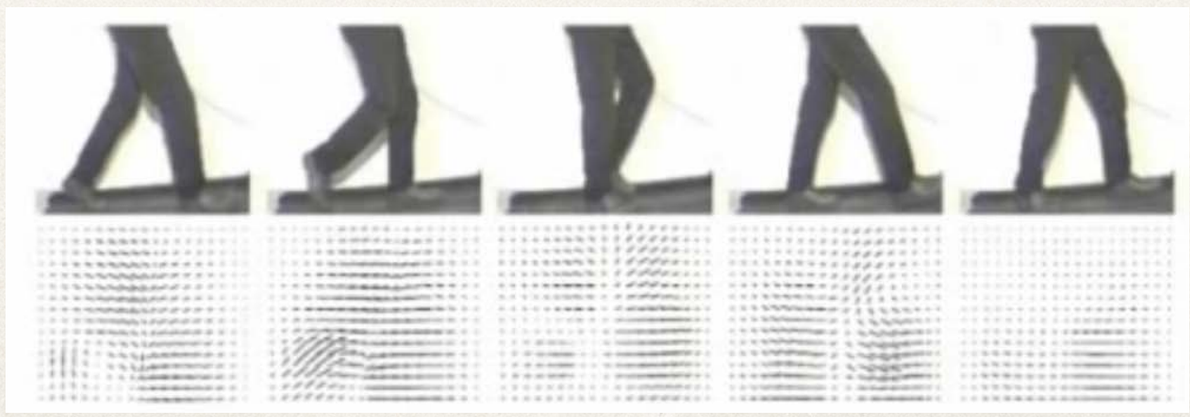
- Motion estimation: ***Optical Flow (Lucas-Kanade, 1981)***
- Classic problem of computer vision (Gibson, 1955)
- Idea: *estimate motion field by estimating pixel-wise correspondence between frames.*
- Assumption:
  - Illumination of the image/image sequence is constant
  - Motion is small

Motion-based Methods

Shape/Appearance vs. Motion

- Shape and appearance in images depends on many factors:
  - clothing, illumination contrast, image resolution etc.

- Estimated motion field is invariant to shape (in theory), and can be used directly to describe human actions

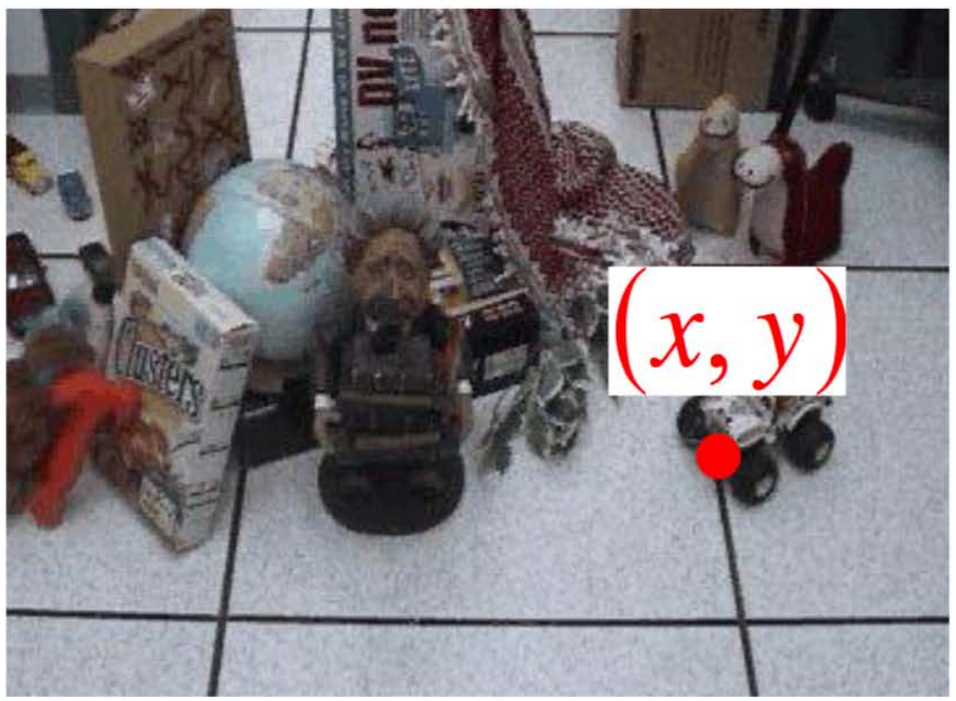


Motion-based Methods

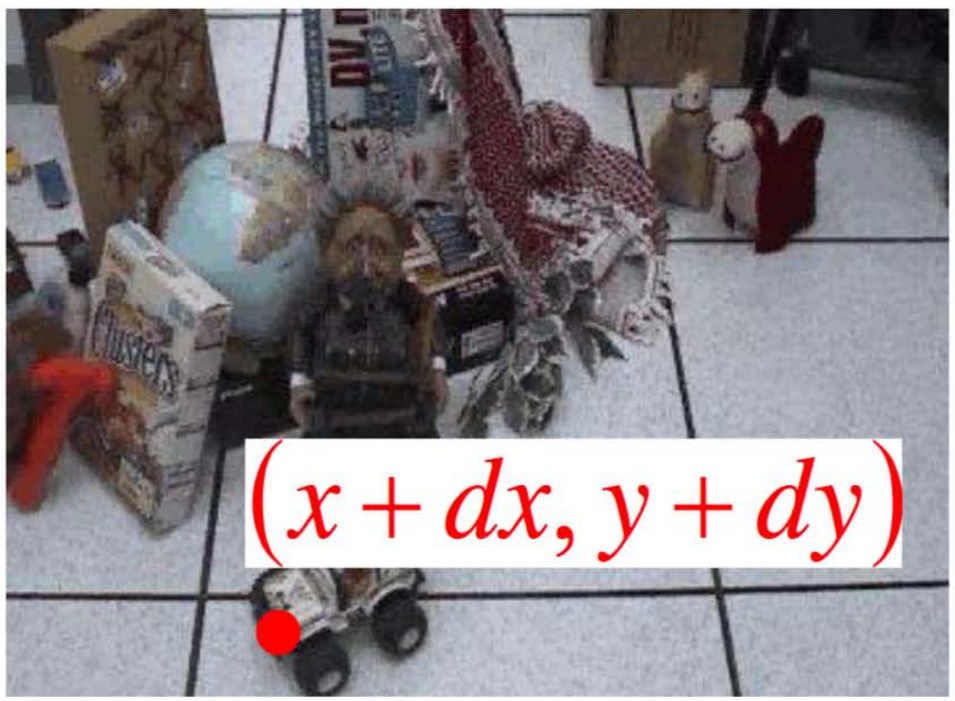
Optical Flow (Procedure)

- Assume the image intensity I is constant

Time = t



Time = $t+dt$



$$I(x, y, t) = I(x+dx, y+dy, t+dt)$$

Motion-based Methods

Optical Flow (Procedure)

- Assumption 1: Brightness is constant.

$$H(x, y) = I(x+u, y+v)$$

- Assumption 2: Motion is small.

$$I(x+u, y+v) = I(x, y)+\frac{\partial I}{\partial x}u+\frac{\partial I}{\partial y}v+\text{higher order terms}$$

$$\approx I(x, y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v$$

Motion-based Methods

Optical Flow (Procedure)

$$I(x, y, t) = I(x + dx, y + dy, t + dt)$$

First order Taylor Expansion

$$= I(x, y, t) + \frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt$$

Simplify notations:

$$I_x dx + I_y dy + I_t dt = 0$$

Divide by *dt* and denote:

$$u = \frac{dx}{dt} \qquad v = \frac{dy}{dt}$$

$$I_x u + I_y v = -I_t$$

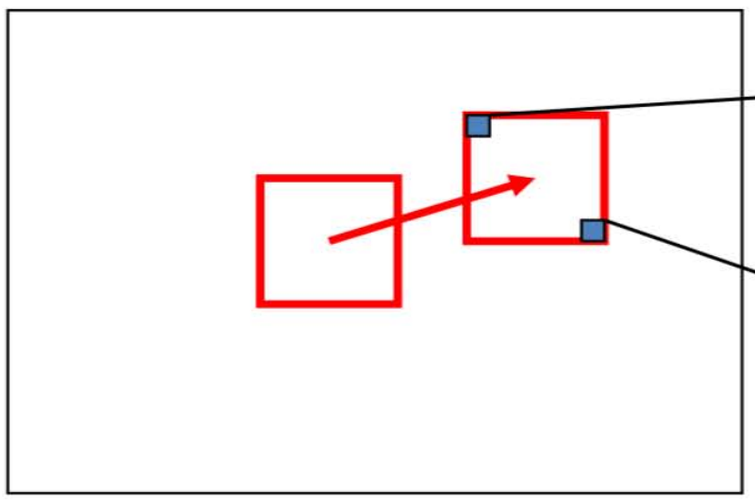
Problem I: One equation, two unknowns

Motion-based Methods

Optical Flow (Procedure)

$$I_x u + I_y v = -I_t \quad \Longrightarrow \quad \begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = -I_t$$

Assume constant (u,v) in small neighborhood



$$\begin{bmatrix} I_{x1} & I_{y1} \\ I_{x2} & I_{y2} \\ \vdots & \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = -\begin{bmatrix} I_{t1} \\ I_{t2} \\ \vdots \end{bmatrix}$$

$$A\vec{u} = b$$

Motion-based Methods

Optical Flow (Results)

- Beauchemin & Barron (1995), *The computation of optical flow*, ACM Comput. Surv. vol. 27(3), pp. 433–466.
- Bhattacharyya et al. (2009), *High-speed target tracking by fuzzy hostility-induced segmentation of optical flow field*, Appl. Soft Comput. vol. 9(1), pp. 126–134

Motion-based Methods

Optical Flow (Results)

Motion-based Methods

Optical Flow (Results)

Application: *Intrusion Detection*

Application: *Multi-persons tracking*

Motion-based Methods

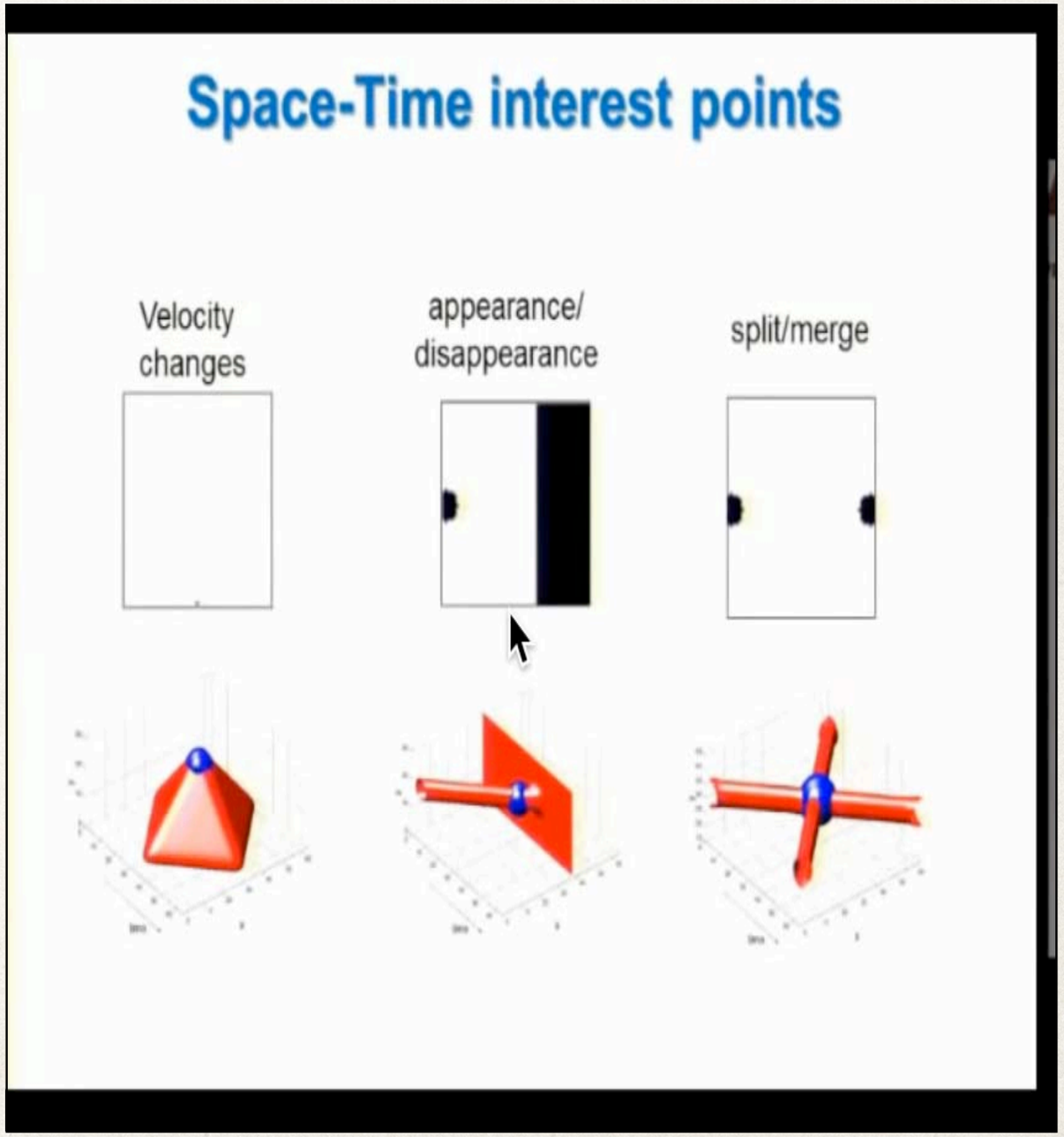
Optical Flow (Results)

Application: Cameras *Tracking*

Motion-based Methods

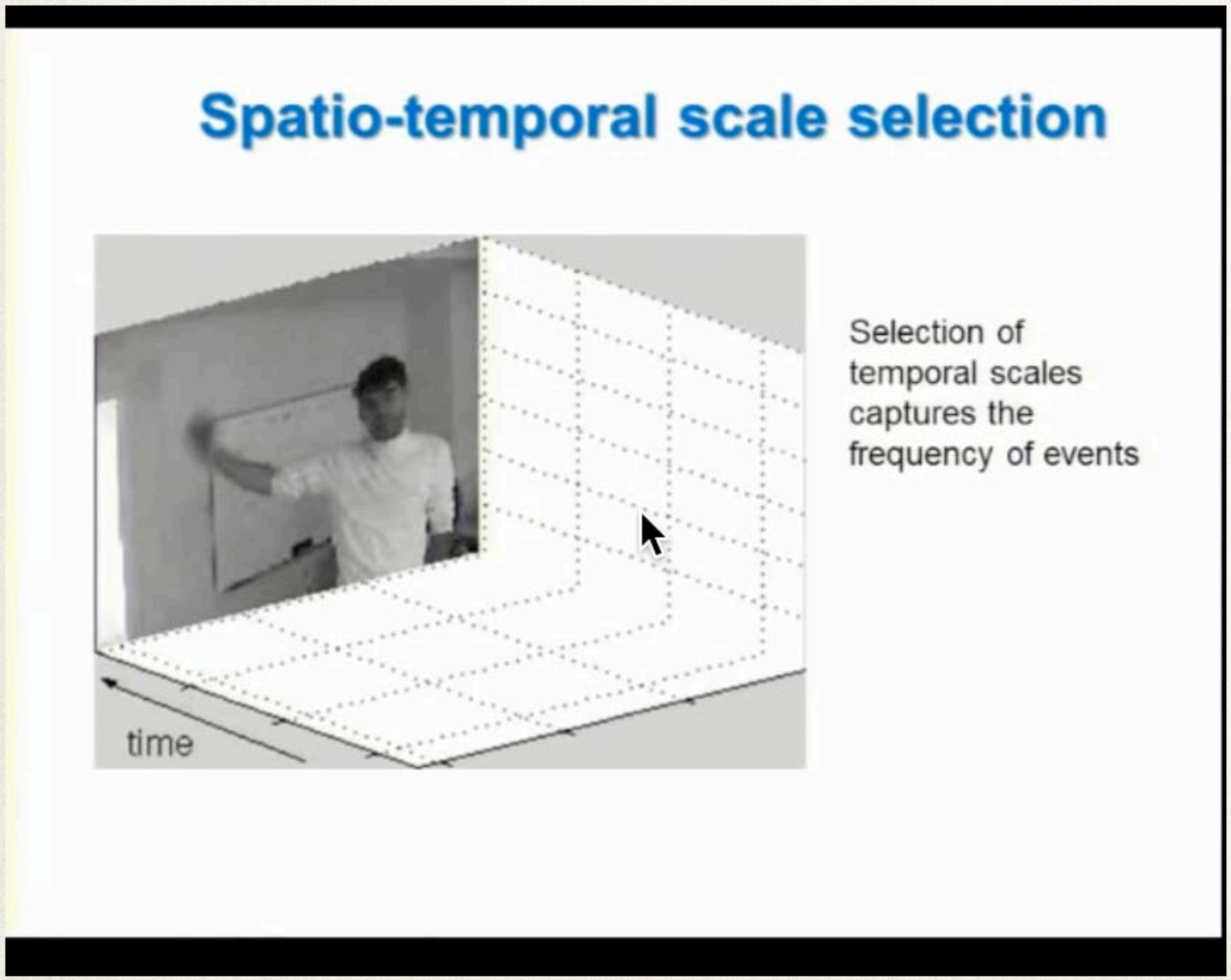
Space-time Interest Points (STIPs)

- One of the seminal work - Space-Time Interest Points (STIPs) by Laptev (IJCV, 2005)
- Laptev extends ***Harris corner detector to 3D-Harris detector***. The **idea** of the 2D Harris corner detector is to find spatial locations in an image with significant changes in two orthogonal directions. The 3D-Harris detector identifies points with large spatial variations and non-constant motions.

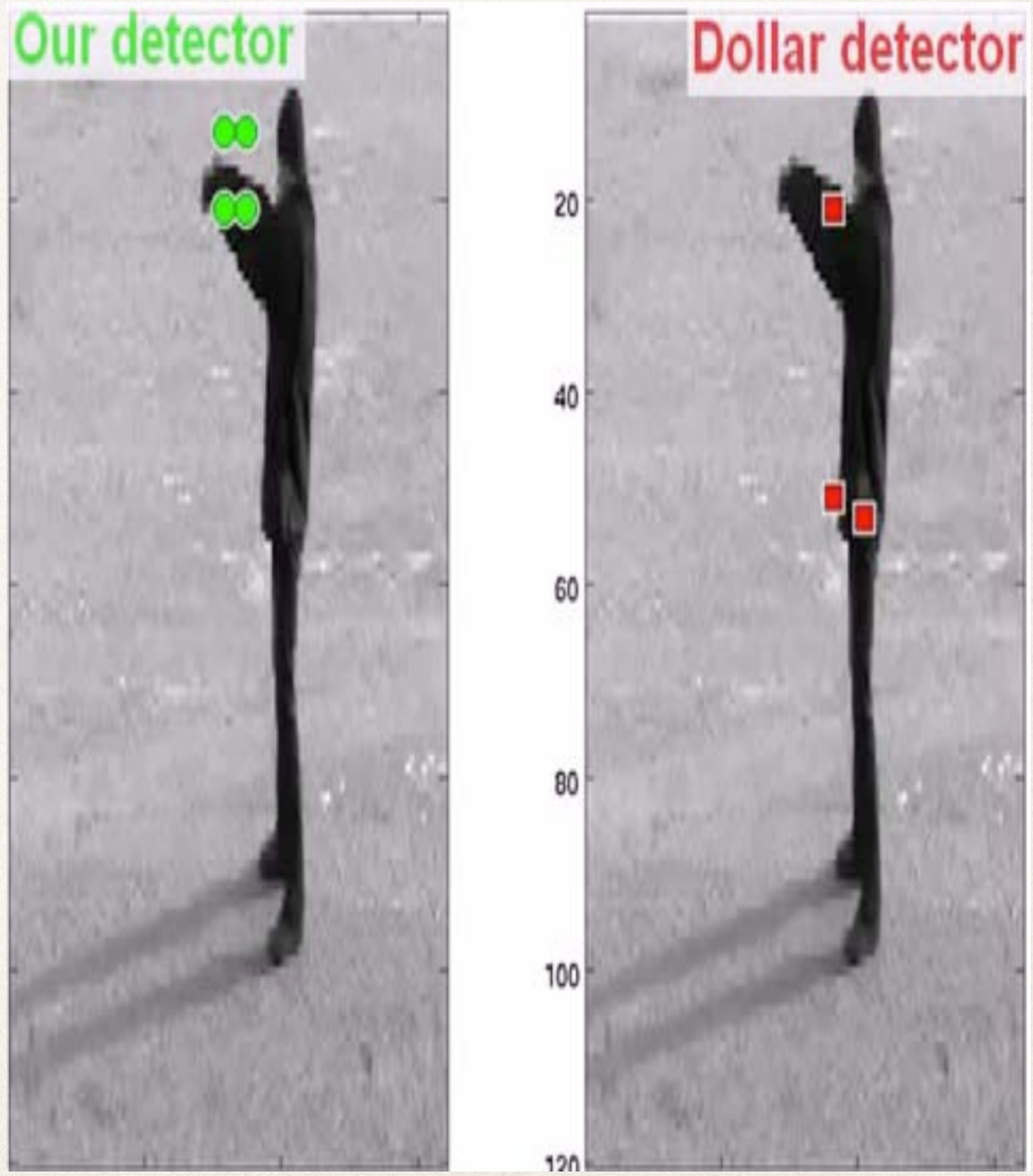


Motion-based Methods

Space-time Interest Points (STIPs)



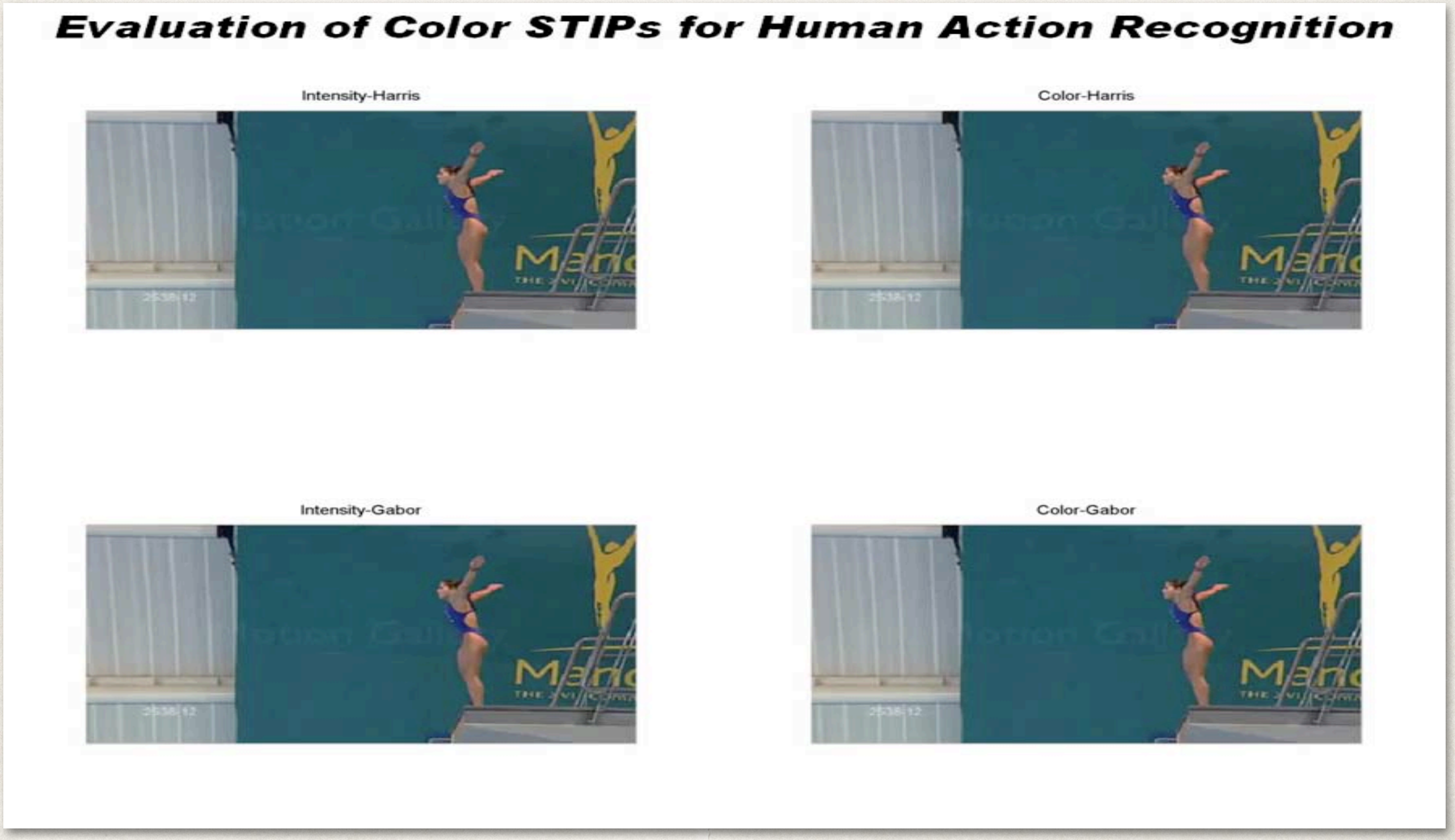
Laptev (IJCV, 2005)



Bregonzio et al (CVPR, 2009)

Motion-based Methods

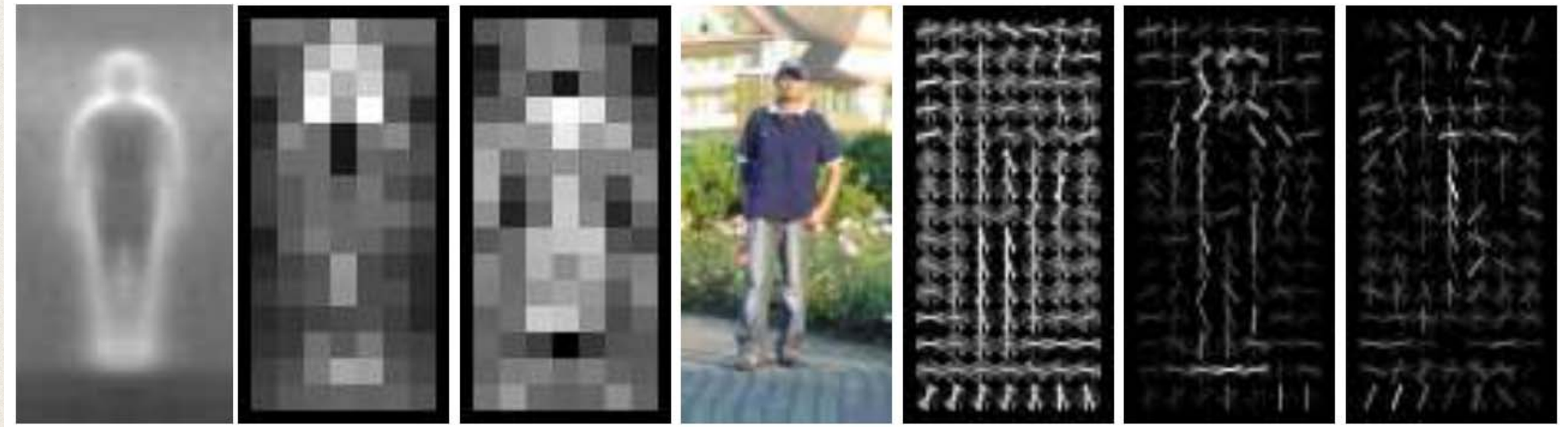
Space-time Interest Points (STIPs)



Motion-based Methods

Others

- Dallas and Triggs (CVPR, 2005) used Histogram of Oriented Gradient (HOG)

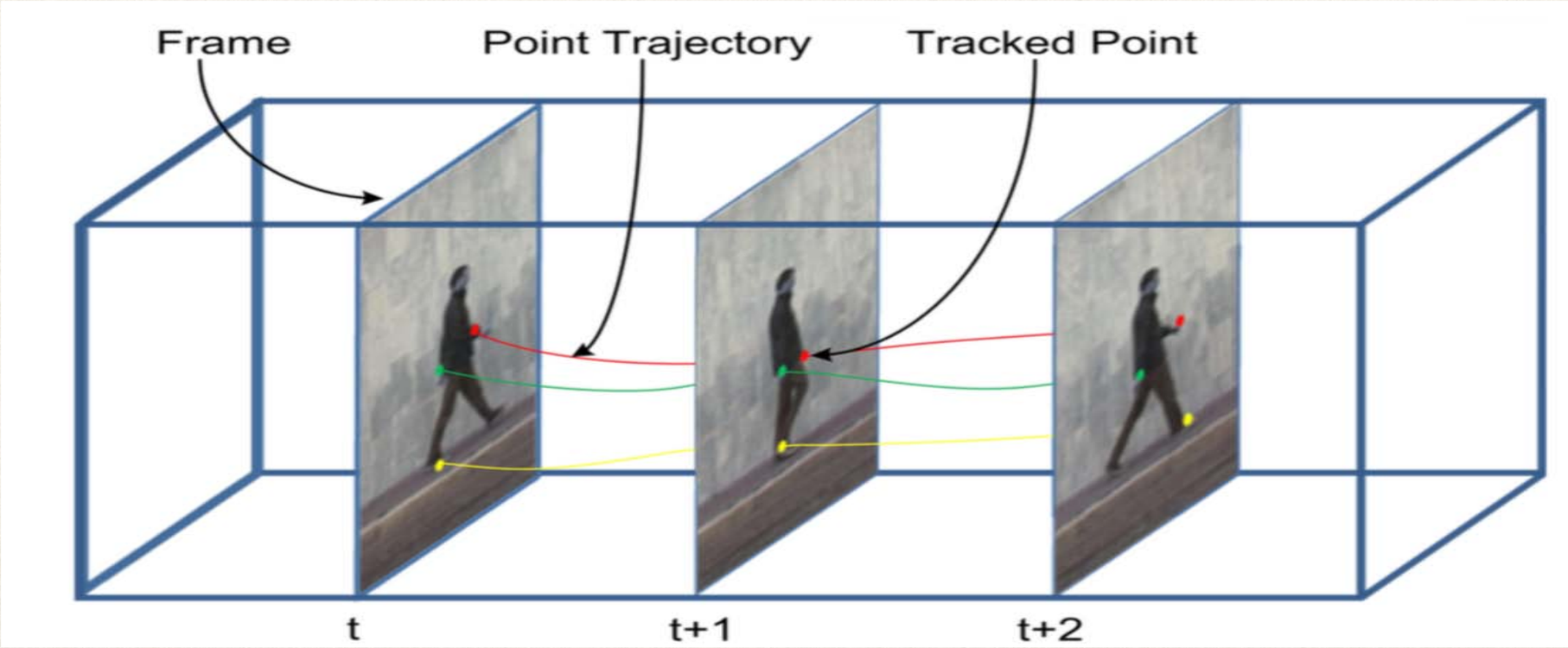


Motion-based Methods

Trajectory solution

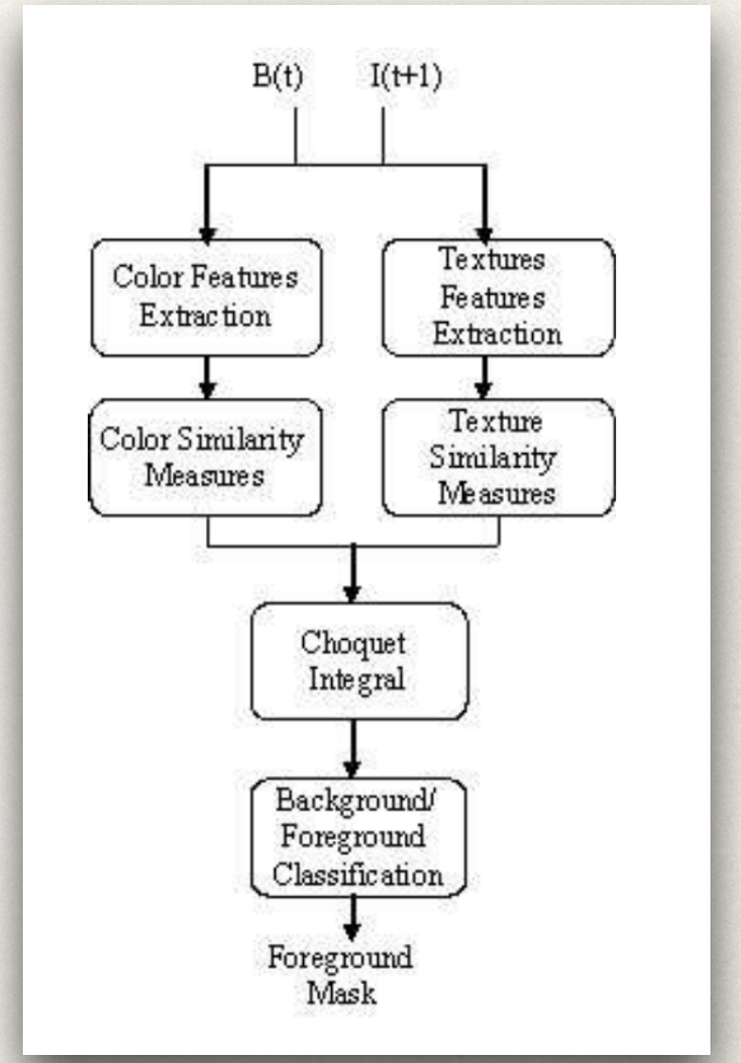
- Extracting local features from trajectories gains its popularity mostly from the work of Messing et al. (ICCV, 2009) and Matikainen et al. (ICCV, 2009). Interestingly, both studies use a form of velocity of trajectories as local features.

- A trajectory is a properly tracked feature over time.



Motion-based Methods

Trajectory solution (Fuzzy integral)



- El Baf et al. (2008), *A fuzzy approach for background subtraction*, in: ICIP, pp. 2648–2651.
- El Baf et al. (2008), *Fuzzy integral for moving object detection*, in: FUZZ-IEEE Systems, pp. 1729–1736.

Motion-based Methods

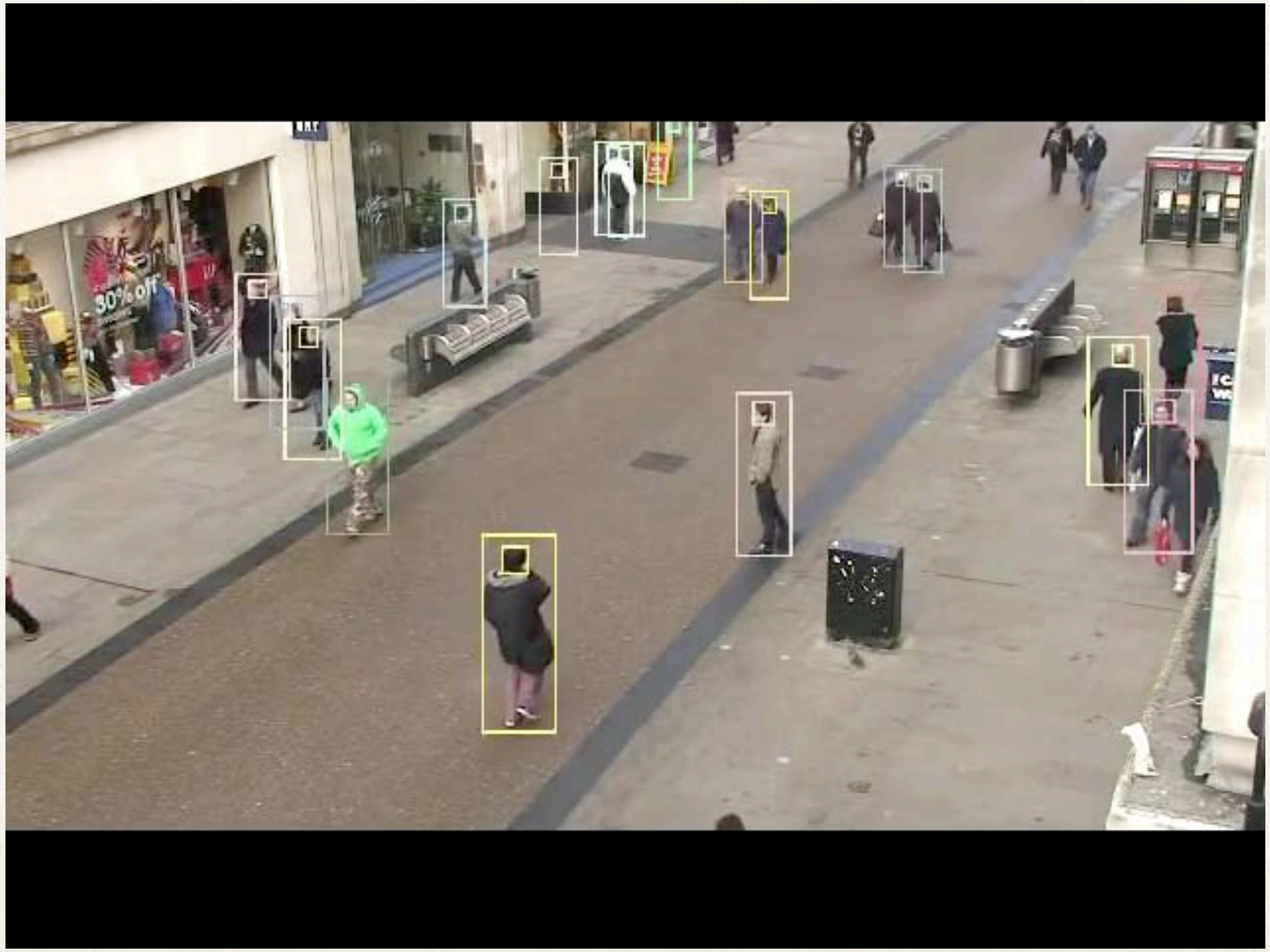
Trajectory solution (Prediction)

Training

Prediction

Motion-based Methods

Trajectory solution (Prediction)



Benfold & Reid (CVPR, 2011)

Tutorial Overview

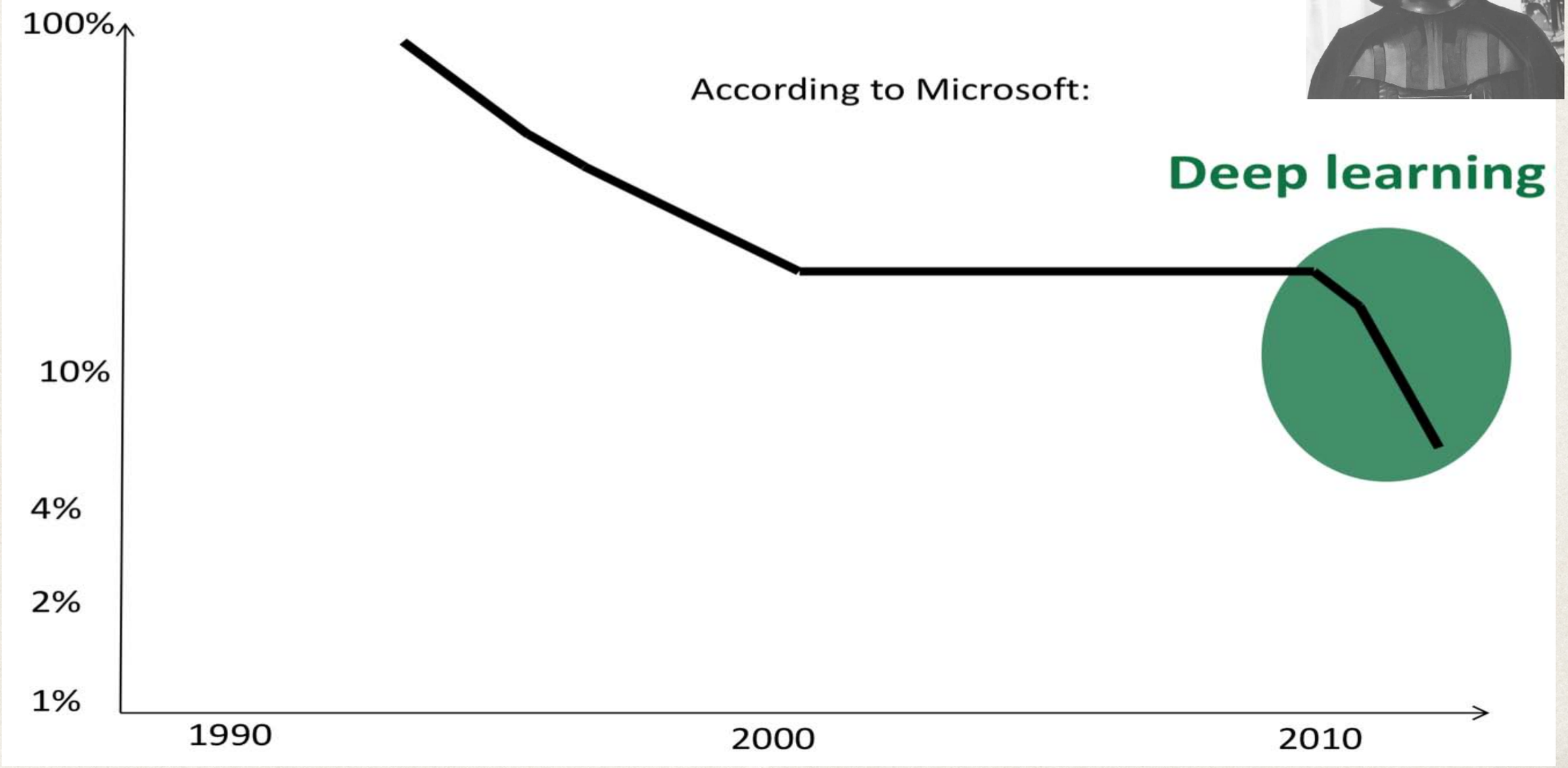
The Vitruvian Man,
Leornardo da Vinci, 1490
Florence, Tuscany, Italy

- Motivation
  - Historical review
  - Applications and challenges
- Appearance-based methods
- Motion-based methods
- **Deep-based methods**
- Datasets

Deep-based model

Neural Nets STRIKE BACK (again)

2010-2012: Breakthrough in speech recognition

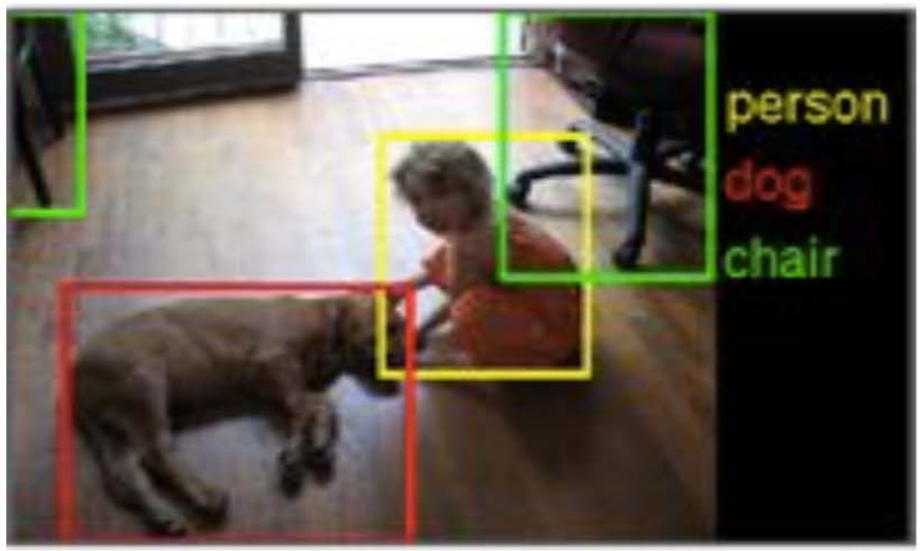


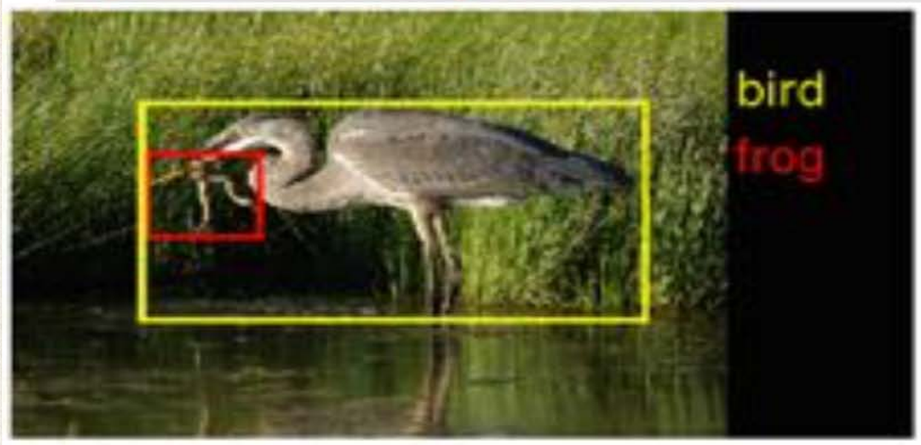
Deep-based model

Neural Nets STRIKE BACK (again)

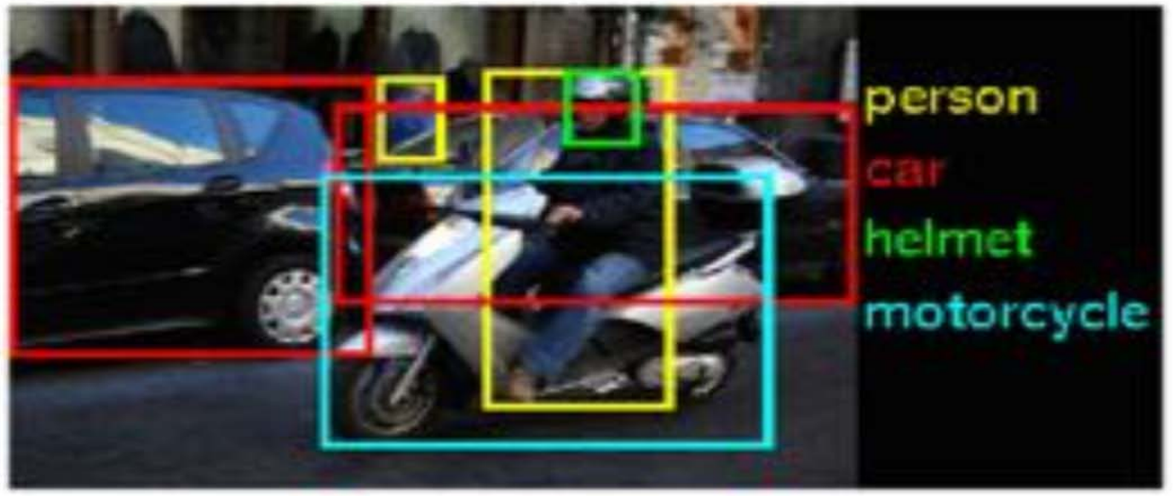
2012-2015: Breakthrough in computer vision

- GPUs + 10x more data



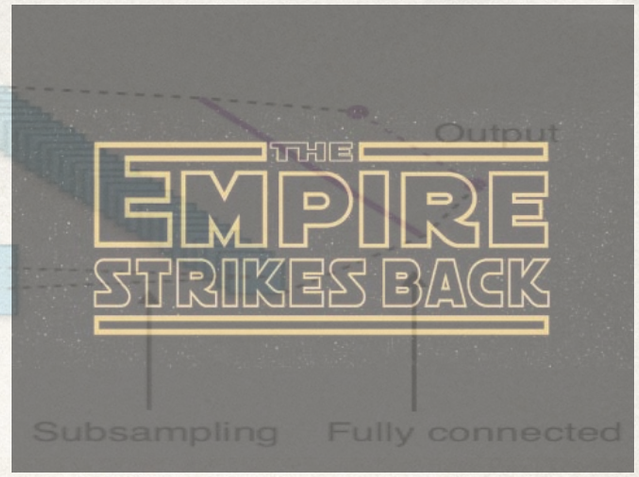


- 1000 object categories,
- Facebook: millions of faces
- 2015: human-level performance



Deep-based model

Neural Nets STRIKE BACK (again)

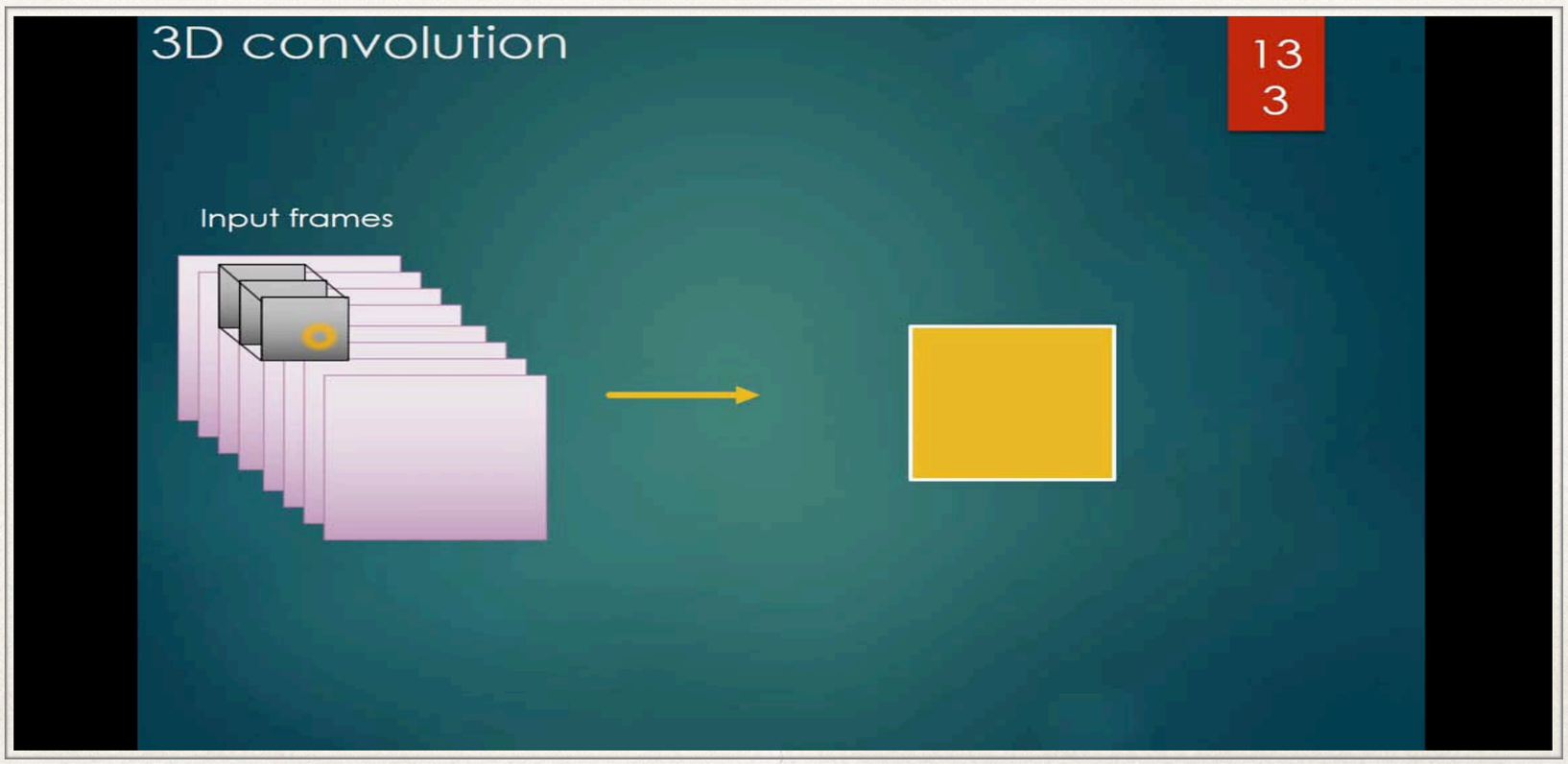


- We are witnessing a **significant** advancement in numerous learning tasks thanks to data driven approaches.
- In particular, deep neural networks such as **Convolutional Neural Networks (CNN)** (Lecun et al., 1998) have become the method of choice in learning image contents (Krizhevsky et al., 2012; Chatfield et al., 2014; Sutskever et al., 2014; Szegedy et al., 2015).
- Generally speaking, the problem of learning is to determine a complicated decision function from the available data. In deep architectures, this is achieved by **composing multiple level of nonlinear operations**. Searching the parameter space of deep architectures is not an easy job given the non-convexity of the decision surface. Learning algorithms based on the gradient descent approach along the computational power of new hardware have been shown to be successful when ***large amount of annotated data is available*** (Wang et al., 2015b; Srivastava et al., 2015b; He et al., 2015).

Deep-based model

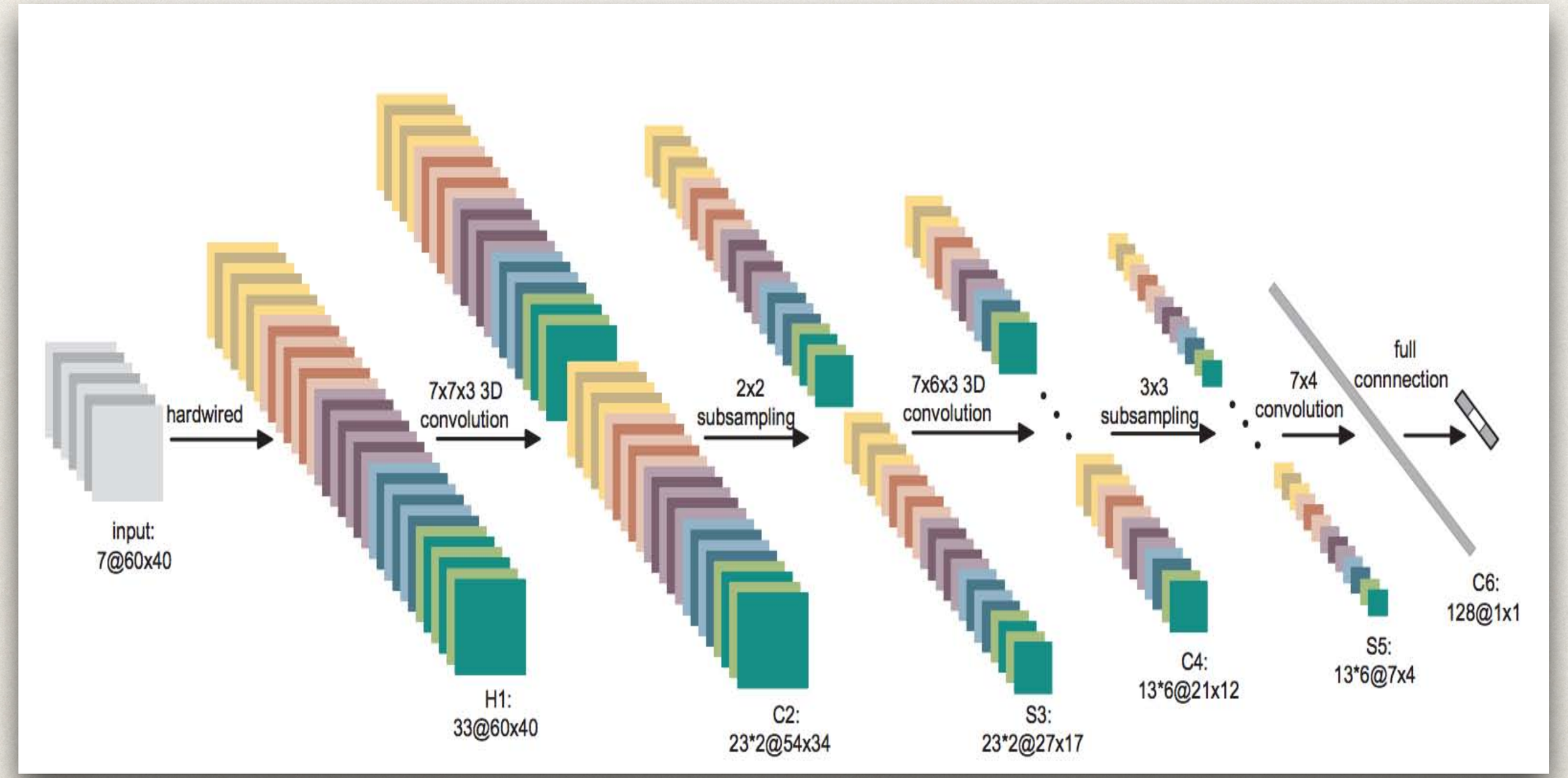
3D CNN

- Analyzing filters learned by CNN architectures suggests that the very first layers learn low level features (*e.g.*, *Gabor*-like filters) while top layers learn high level semantics (Zeiler and Fergus, ECCV, 2014).
- ***3D convolutional*** **networks** are introduced in Ji et al. (TPAMI, 2012). extract features from both spatial and temporal dimensions, hence is expected to capture spatiotemporal information and motions encoded in adjacent frames.



Deep-based model

3D CNN

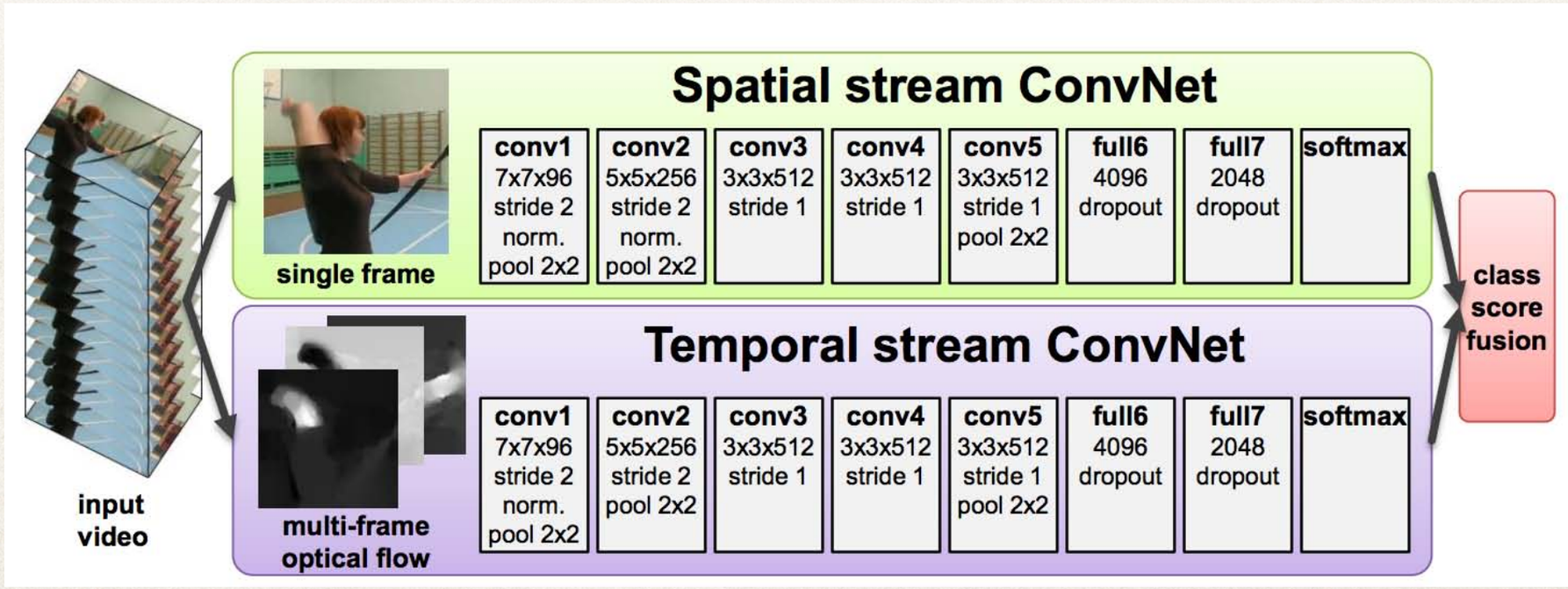


Gray Grad_x Grad_y OF_x OF_y

Deep-based model

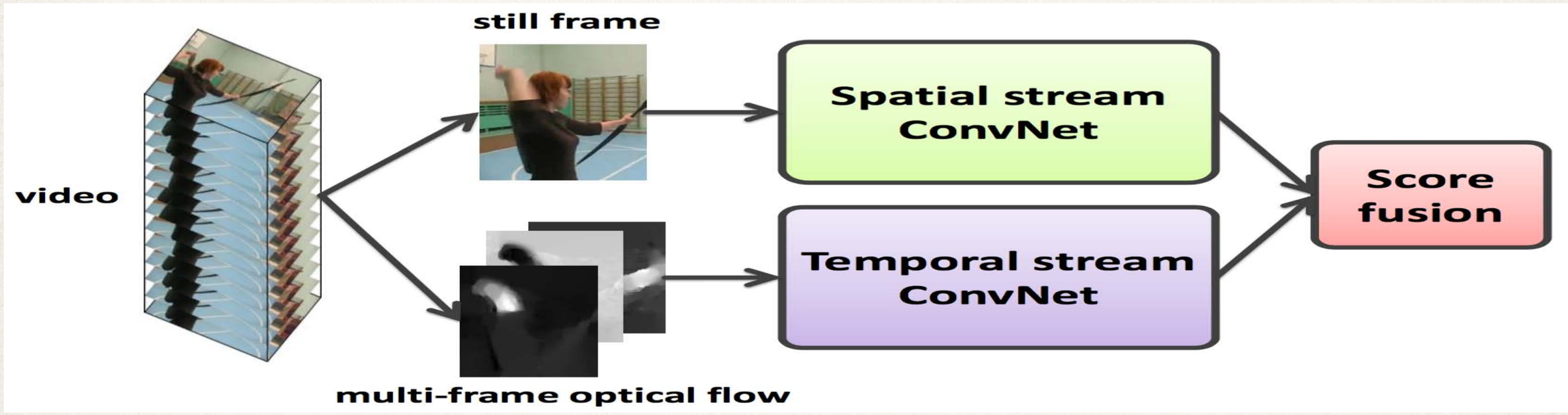
Multi-stream CNN

- In visual perception, the ***Ventral Stream*** of our visual cortex processes object attributes such as appearance, color and identity. The motion an object and its location is handled separately through the ***Dosaral Stream*** Goodale and Milner (Essential Sources in the Scientific Study of Consciousness, 2003). A class of deep neural networks opted for separating appearance based information from motion related ones for action recognition Simonyan and Zisserman (NIPS, 2014).
- Simonyan and Zisserman (NIPS, 2014) introduced one of the **first multiple-stream deep convolutional networks** where the structure of two parallel networks are selected as VGG-16 of Chatfield et al. (BMVC, 2014) for action recognition. The so called spatial stream network accepts raw video frames while the temporal stream network gets optical flow fields as input.



Deep-based model

Multi-stream CNN



Idea:

- Video decomposed into spatial & temporal components: **still frames** & **optical flow**.
- Separate recognition stream for each component.

Deep-based model

Multi-stream CNN

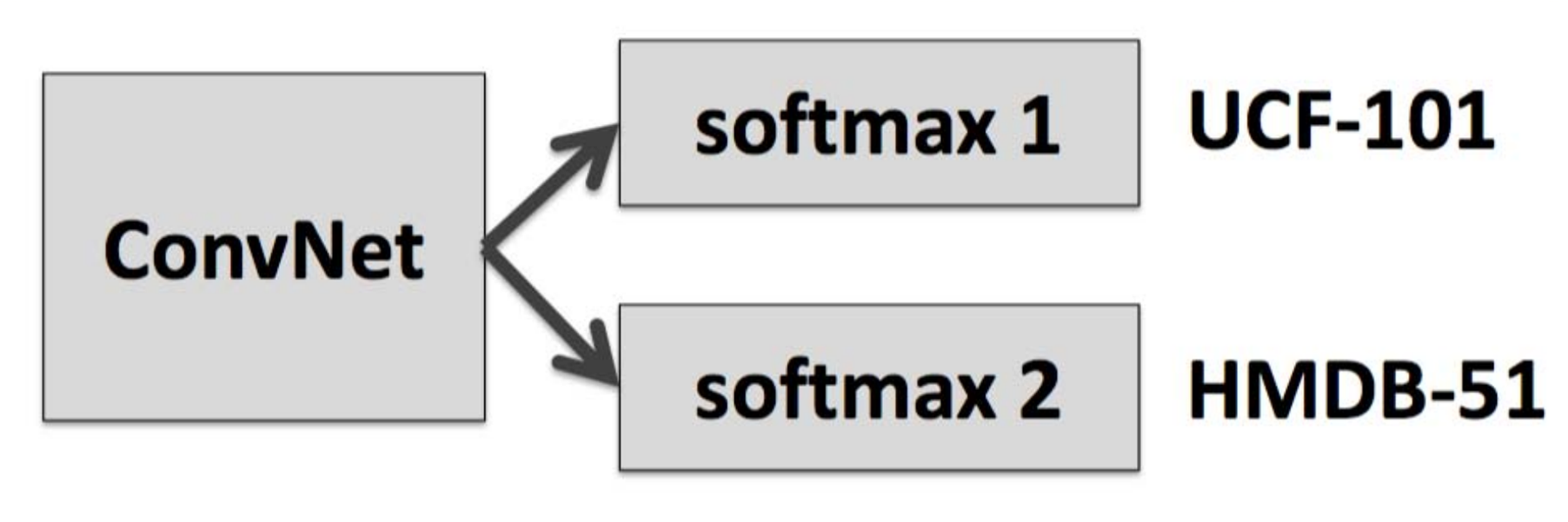
conv1	conv2	conv3	conv4	conv5	full6	full7	full8
7x7x96 stride 2 norm. pool 2x2	5x5x256 stride 2 pool 2x2	3x3x512	3x3x512	3x3x512 pool 2x2	4096 dropout	2048 dropout	softmax

Idea:

- Video decomposed into spatial & temporal components: still frames & optical flow.
- Separate recognition stream for each component.

Deep-based model

Multi-stream CNN



Idea:

- Video decomposed into spatial & temporal components: still frames & optical flow.
- Separate recognition stream for each component.
- Streams combined by late **fusion of soft-max scores** (averaging or linear SVM).
- Most previous approaches: stack frames into a 3-D input volume (Ji et al., TPAMI, 2012).

Deep-based model

Multi-stream CNN

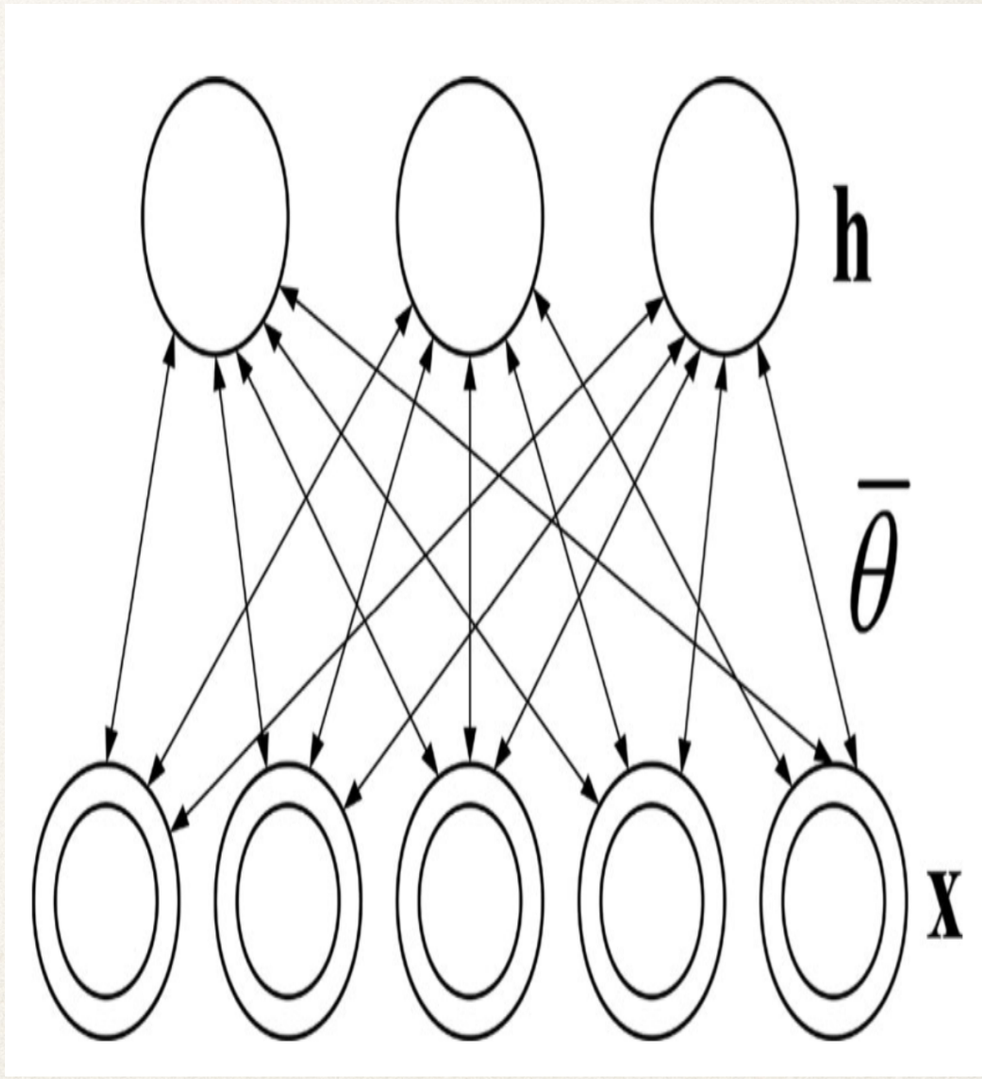
Comparison with the state of the art (mean accuracy over 3 splits, %)

Model	UCF-101	HMDB-51
Spatial Stream ConvNet	73.0	40.5
Temporal Stream ConvNet	83.7	54.6
Two-stream ConvNet (SVM fusion)	**88.0**	**59.4**
Spatio-temporal HMAX [Kuehne et al., ICCV '11]	-	22.8
Spatio-temporal ConvNet [Karpathy et al., CVPR '14]	65.4	-
Two-stream ConvNet & LSTM (split 1) [Donahue et al., arXiv '14]	82.9	-
Hand-crafted feat. & Fisher vector [Wang and Schmid, ICCV '13]	85.9	57.2
Hand-crafted feat. & higher-dim encoding [Peng et al., arXiv '14]	**87.9**	61.1
Hand-crafted feat. & deep Fisher encoding [Peng et al., ECCV '14]	-	**66.8**

Deep-based model

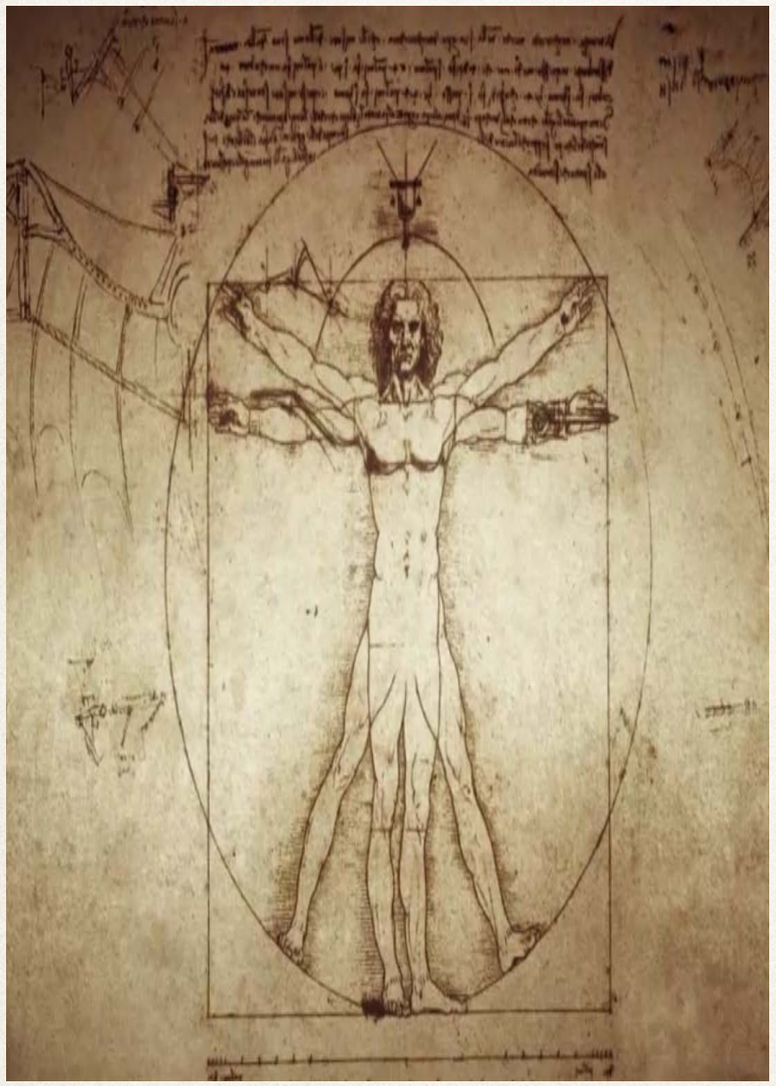
Fuzzy Restricted Boltzmann Machine (Chen et al, T-FS 2015)

- The proposed FRBM is illustrated on the left, in which the connection weighs and biases are **fuzzy parameters** denoted by θ.
- The optimization in learning process turns into a **fuzzy maximum likelihood problem**. However, this kind of problem is **quite intractable** because the fuzzy objective function is non-linear and the membership function is difficult to compute, since the computation of its alpha-cuts become NP-hard problems.
- Therefore, the work transforms the problem into regular maximum likelihood problem by **defuzzifying** the fuzzy free energy function. And center of area (centroid) method is employed to defuzzify.



FRBM

Tutorial Overview

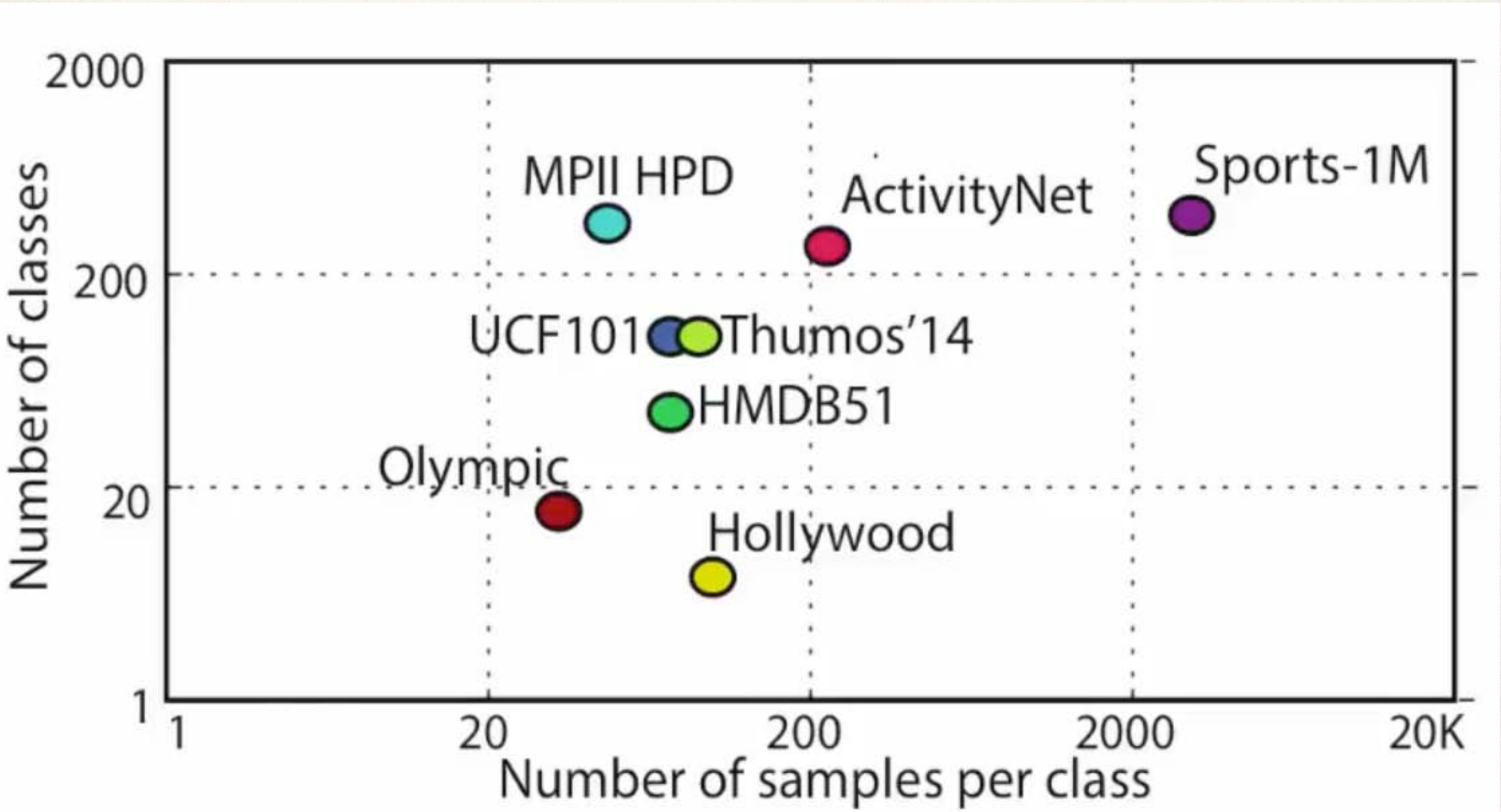


The Vitruvian Man,
Leornardo da Vince, 1490

- Motivation
  - Historical review
  - Applications and challenges
- Appearance-based methods
- Motion-based methods
- Deep-based methods
- **Datasets**

Dataset

HMA Benchmark



Dataset

Hollywood

Dataset

ActivityNet

Dataset

Sports1M

Paper	Method	Dataset						
		HMDB51	UCF101	UCF50	UCF-Sports	Holly-wood2*	Olympic Sports*	Sports1M
Wang et al. (2011)	Dense Traj (Traj + HOG+HOF+MBH)				88.2	58.3		
Kliper-Gross et al. (2012)	Motion Interchange Patterns	29.2		68.5				
Sadanand and Corso (2012)	General	26.9						
	Video Wise			76.4				
	Group Wise			57.9				
Oneata et al. (2013)	MBH + SIFT + Sqrt + L2 Normalization	54.8		90		63.3	82.1	
Yang et al. (2013)					88			
Wang and Schmid (2013)	Without Human Detector	55.9		90.5		63	90.2	
	With Human Detector	57.2		91.2		64.3	91.1	
Jain et al. (2013)	Traj + HoG + HoF + MBH + DCS on w-flow	52.1				62.5		
Peng et al. (2014b)	Stacked FVs + FV	66.8						
Peng et al. (2014a)	Hybrid-BoW	61.1	87.9	92.3				
Kantorov and Laptev (2014)	MPEG-Flow : VLAD encodings of	46.3						
Gaidon et al. (2014)	SDT tree ATEP	41.3				54.4	85.5	
Simonyan and Zisserman (2014)	Two-stream(VGGNet-16)	59.4	88.0					
Karpathy et al. (2014)	Transfer Learning on Sports 1M		65.4					
	Clip Hit @ 1 - Slow Fusion							41.9
	Video Hit @ 1 - Slow Fusion							60.9
Wang et al. (2015b)	Two-Stream (ClarifaiNet)		88.0					
	Two-Stream (GoogLeNet)		89. 3					
	Two-Stream (VGGNet-16)		91.4					
Wang et al. (2015a)	TDD + Wang and Schmid (2013)	65.9	91.5					
	TDD (Only)	63.2	90.3					
Ng et al. (2015)	Conv Pooling Hit@1 (Best)							72.4
	LSTM Hit@1 (Best)							73.1
	Conv Pooling (Image + Opt Flow)		88.2					
	LSTM (Image + Opt Flow)		88.6					
Fernando et al. (2015)	Rank Pooling	63.7				73.7		
Donahue et al. (2015)	LRCN- Weighted Avnerage of RBG + Flow		82.9					
Wu et al. (2015)	Adaptive Multi-Stream Fusion		92.6					
Jiang et al. (2015)	TrajShape+TrajMF	48.4	78.5			55.2	80.6	
	TrajShape+TrajMF+ Wang and Schmid (2013)	57.3	87.2			65.4	91	
Lan et al. (2015)	Multi-Skip Feat. Stacking	65.1	89.1	94.4		68.0	91.4	
Hoai and Zisserman (2015)	Proposed SSD + RCS	62.2				72.7		
Misra et al. (2016)	ImageNet pretrain + tuple verification	29.9						
	HMDB + UCF101 Labels Only	30.6						
Wang et al. (2016)	Proposed Only (RBG + Opt Flow Networks)	62	92.4					


POTENTIAL
FUTURE
DIRECTION

Future Direction I

EARLY EVENT PREDICTION - VIDEO

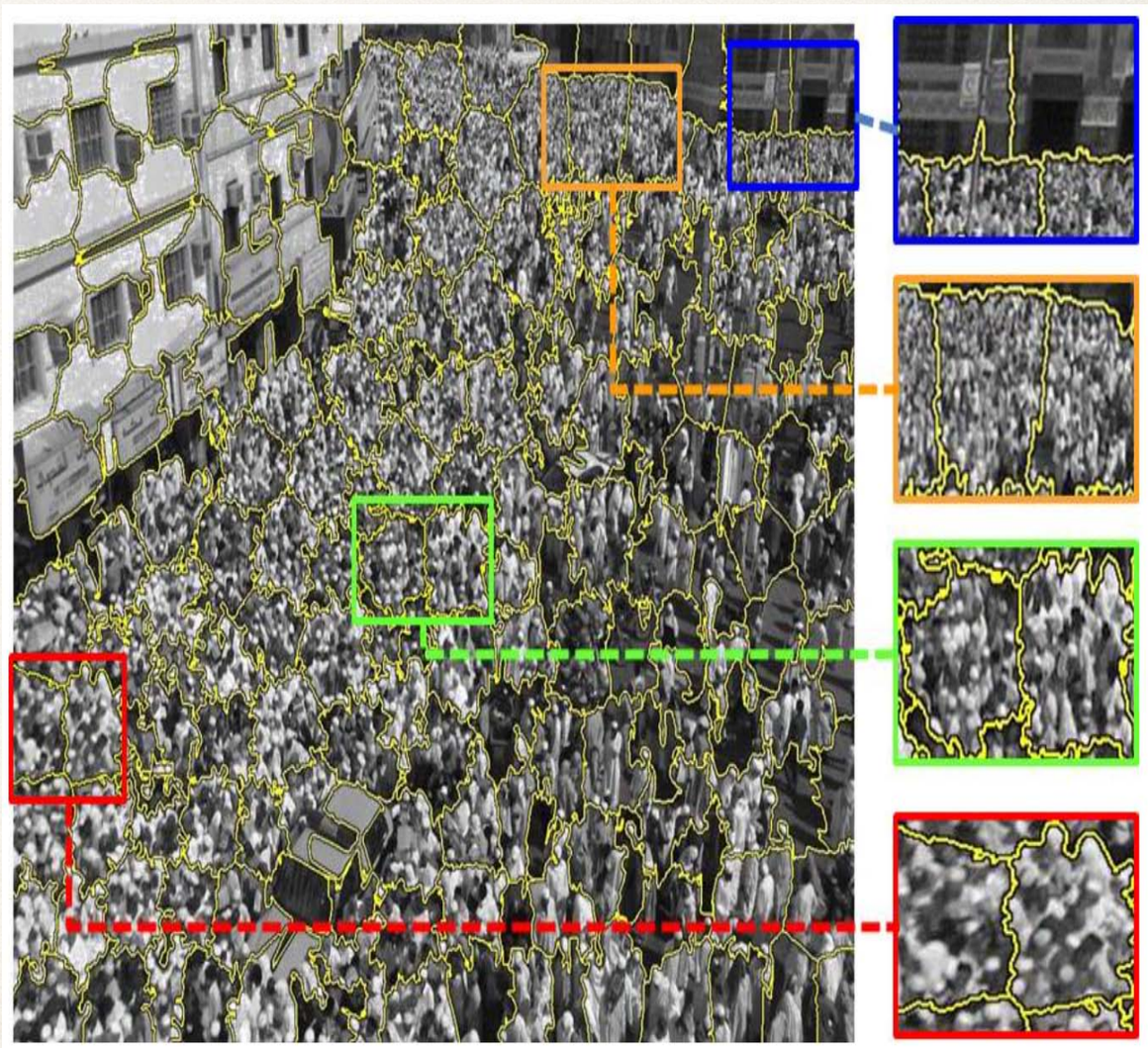
- Early Event Detector

*** Vats et al. (2015) Early Human Actions Detection using BK Sub-triangle Product, FUZZ-IEEE (Best Student Nomination Paper)**

Future Direction II

CROWD MODELING - Video

- Crowd Behaviour Understanding



- **Ryan et al (2010)** ***Crowd Counting Using Group Tracking and Local Features*****, in AVSS**
- **Kok & Chan (Accepted)** ***GrCS: Granular Computing Based Crowd Segmentation*****, T-Cyb.**

Future Direction III

EARLY EVENT PREDICTION - SINGLE IMAGE

What will the man do next?

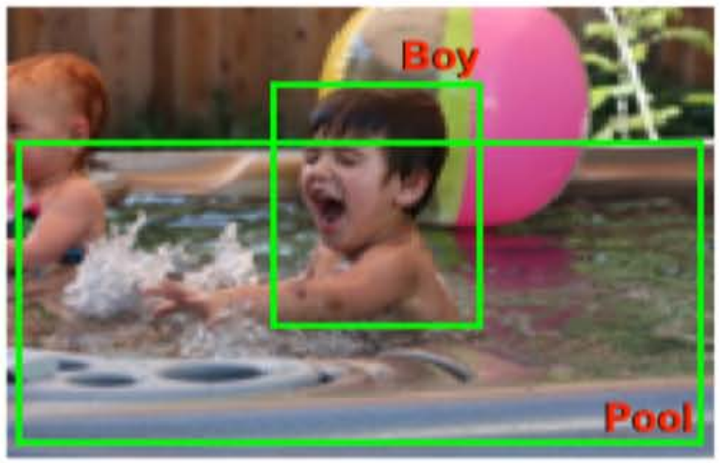
Although predicting the future is difficult, you can find several clues in the image below if you look carefully. *"The young couple seems to be at an open house, the real estate agent is holding paperwork, and the man's arm is starting to move. You might wager that this couple has bought a house, and therefore, to finalize the deal, the man will soon shake hands".*

Vondrick et al. (2015) -
***Anticipating the Future by Watching Unlabeled Video*, arXiv.**

Future Direction IX

Image Captioning/Visual Question Answer

Image:



Our proposed: A little boy is playing in the pool.

Reference : A boy with a beach ball behind him playing in a pool.

Our proposed: A person is climbing a rock wall.

Reference : A distant person is climbing up a very sheer mountain.

http://cloudcv.org/vqa/

Future Direction V

USE OF "BIG" DATA

- **Deep Learning** (avoid hand-crafted features)

  Why ???

- Facebook: Photo uploads total **2 million per minutes**
- Instagram: **49K** photos uploaded per minutes
- TV-channels recording since 60's
- Youtube: **300 hours** uploaded per minutes
- CCTV: 30M surveillance cameras in US => **700 videos hours/day**

Concluding Remarks

- In this tutorial, the team has represented the fuzzy computer vision.
- This domain is still limited by fuzzy researchers as compare to stochastic/ machine learning based solutions - ICCV, CVPR, ECCV….
- Given the wide range of applications and more to come, it is rather surprising that our involvement in this topic is limited. ***Encourage…..***
- Jim, Derek and myself - hosted special session since 2013, hopefully next year in Naples……

Material of This Tutorial

WCCI 2016

ELSEVIER

Pattern Recognition

Volume 48, Issue 5, May 2015, Pages 1773–1796

Fuzzy human motion analysis: A review

Chern Hong Lim[1], Ekta Vats[1], Chee Seng Chan

Centre of Image and Signal Processing, Faculty of Computer Science & Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

Received 2 June 2014, Revised 3 September 2014, Accepted 24 November 2014, Available online 9 December 2014

CrossMark

Show less

doi:10.1016/j.patcog.2014.11.016

Highlights

- A survey of fuzzy set oriented methods for human motion analysis is presented.
- This is the first time such a survey is presented in the fuzzy set literature.
- Categorization of existing approaches into three broad levels is performed.
- Insights and suggestions for future research are discussed.

* **Lim et al. (2015)** ***Fuzzy Human Motion Analysis: A Review,*** **Pattern Recognition, vol. 48(5), pp. 1773-1796**

(First fuzzy review paper in human motion analysis with 252 references)

THANK YOU !!!!!!