Intelligent Multimedia Systems

Master AI, 2012, Lecture 1

Lecturers: Theo Gevers

Lab: Intelligent Systems Lab Amsterdam (ISLA)

Email: th.gevers@uva.nl

http://staff.science.uva.nl/~gevers

Lectures

- 29-10-2012, Monday, 15:00-17:00, Science Park A1.04 - Introduction
- 05-11-2011, Monday, 15:00-17:00, Science Park A1.04 - Image and Video Formation
- 12-11-2011, Monday, 15:00-17:00, Science Park A1.04 - Color Invariance and Image Processing
- 19-11-2011, Monday, 15:00-17:00, Science Park A1.04 - Feature Extraction and Object Recognition
- 26-11-2011, Monday, 15:00-17:00, Science Park A1.04 - Learning and Tracking
- 03-12-2011, Monday, 15:00-17:00, Science Park A1.04 - Visual Attention and Affective Computing
- 10-12-2011, Monday, 15:00-17:00, Science Park A1.04 - Human Behavior Analysis
- 18-12-2011, Tuesday, 15:00-18:00, Science Park, C1.10 - Examination

Lab Session

- 02-11-2011, Friday, 13:00-17:00, Science Park G1.23 - Fares Alnajar, Ronan Sicre and Emrah Tasli
- 09-11-2011, Friday, 13:00-17:00, Science Park G1.23 - Fares Alnajar, Ronan Sicre and Emrah Tasli
- 16-11-2011, Friday, 13:00-17:00, Science Park G1.23 - Fares Alnajar, Ronan Sicre and Emrah Tasl
- 23-11-2011, Friday, 13:00-17:00, Science Park G1.23 - Fares Alnajar, Ronan Sicre and Emrah Tasl
- 30-11-2011, Friday, 13:00-17:00, Science Park G1.23 - Fares Alnajar, Ronan Sicre and Emrah Tasl
- 07-12-2011, Friday, 13:00-17:00, Science Park G1.23 - Fares Alnajar, Ronan Sicre and Emrah Tasl
- 14-12-2011, Friday, 13:00-17:00, Science Park G1.23 - Fares Alnajar, Ronan Sicre and Emrah Tasl

Tutorials/Exercises

- 31-10-2011, Wednesday, 13:00-15:00, Science Park G4.15 - Free
- 07-11-2011, Wednesday, 13:00-15:00, Science Park G4.15 - Theo Gevers
- 14-11-2011, Wednesday, 13:00-15:00, Science Park G4.15 - Theo Gevers
- 21-11-2011, Wednesday, 13:00-15:00, Science Park G4.15 - Theo Gevers
- 28-11-2011, Wednesday, 13:00-15:00, Science Park G4.15 - Theo Gevers
- 05-12-2011, Wednesday, 13:00-15:00, Science Park G4.15 - Theo Gevers
- 12-12-2011, Wednesday, 13:00-15:00, Science Park G4.15 - Preparation

Grades

The evaluation of the course consists of 1 written examination and 1 lab session assignment.

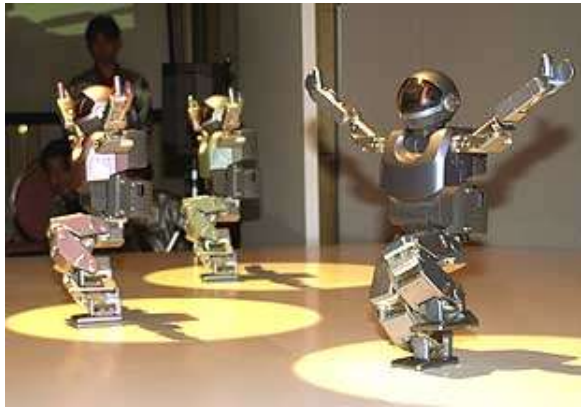
Both count for 50% of the final mark.

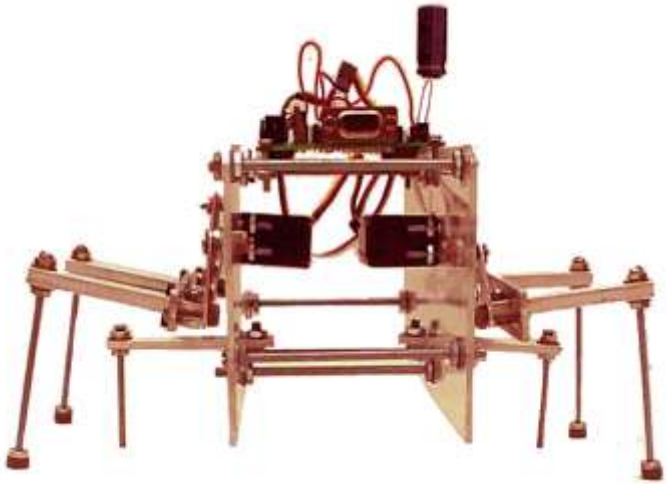
Today's Class

- Introduction
- What are Multimedia Intelligent Systems
- Computer vision and machine learning
- Specifics of this course

Why intelligent systems matter
Robots

A255 robot





Why intelligent systems matter

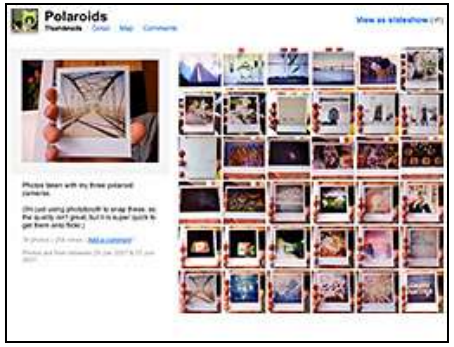
Safety

Health

Security

Comfort

Fun



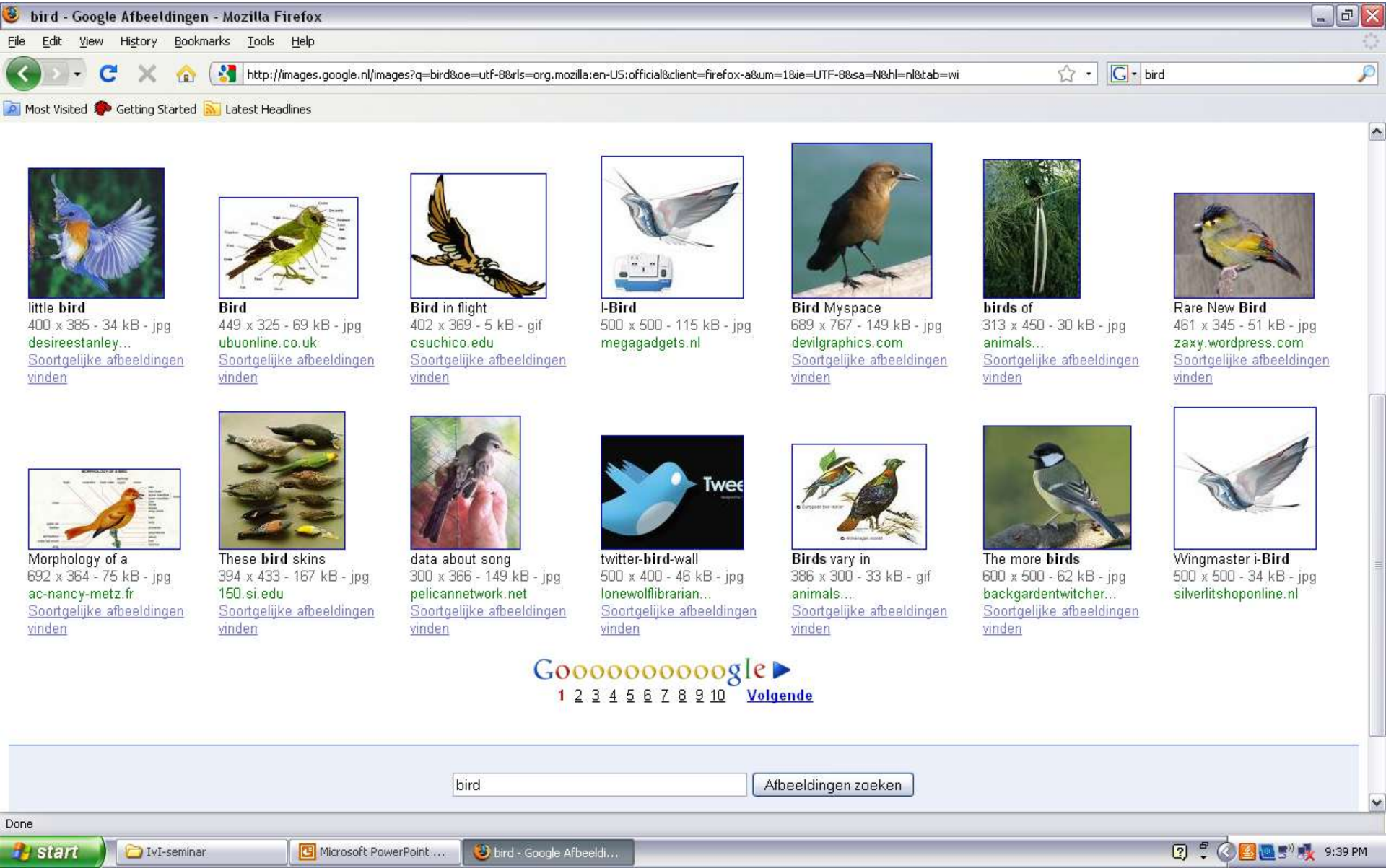
Access

Introduction
Image and Video Access

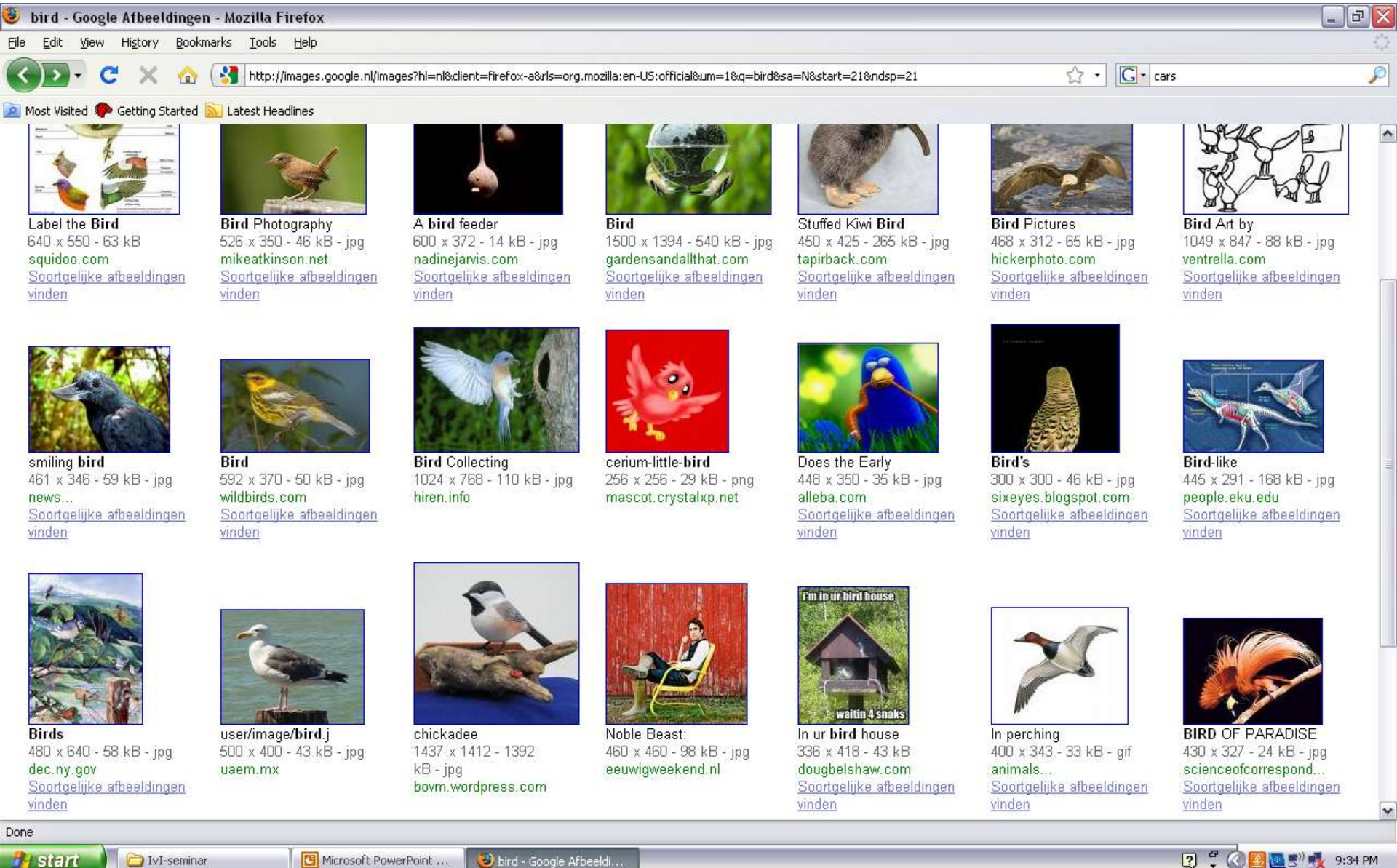
- Today, there are billions of images on the Internet and in collections such as FaceBook and Flickr.

- Suppose I want to find pictures of birds, humans, cars, boats or videos of explosion, violence etc

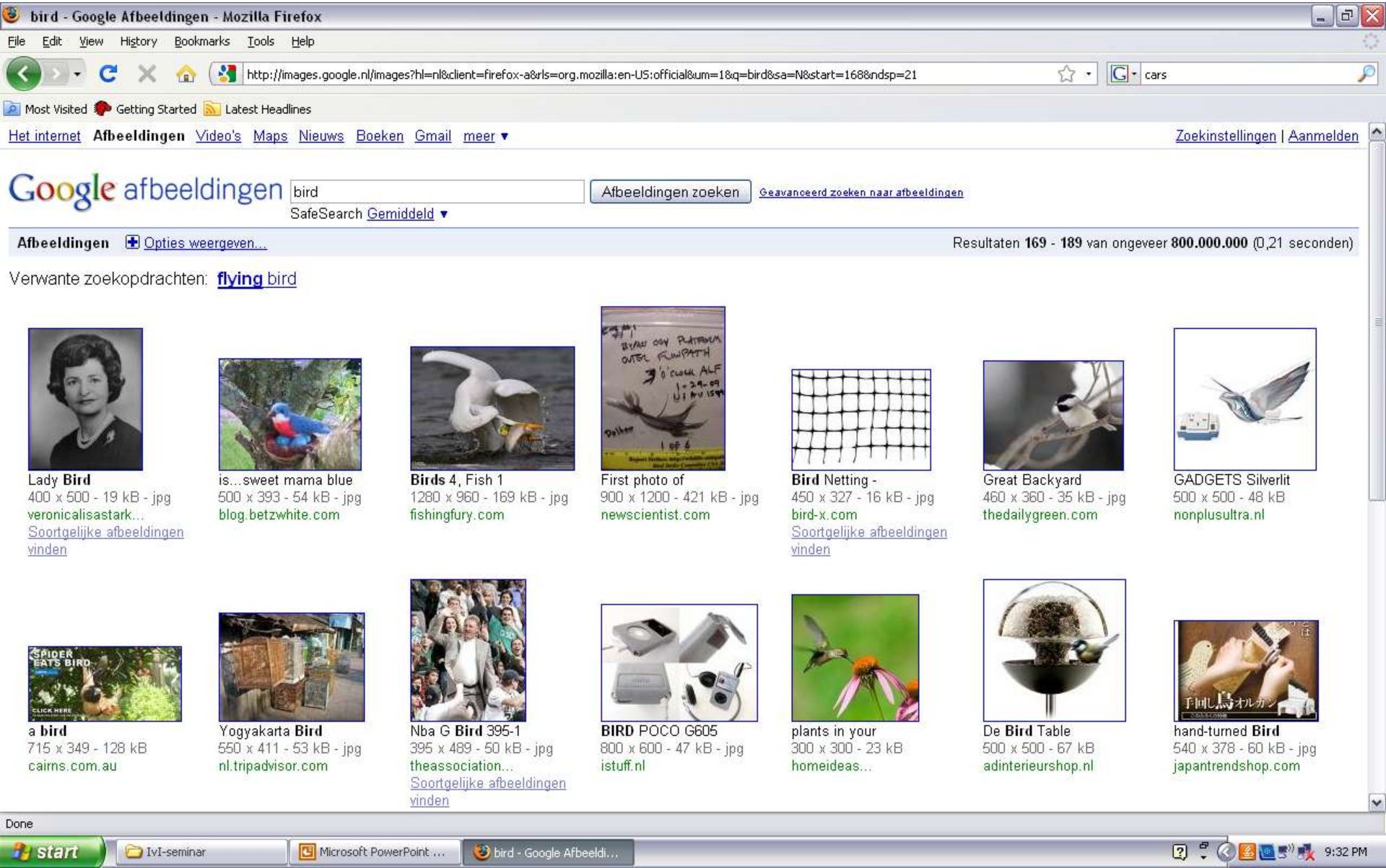
Google Image Search – Bird(1)



Google Image Search – Bird(2)



Google Image Search – Bird(3)



Flickr Search for tag *monkey*








- Even with humans doing the labeling, the data is extremely noisy -- context, polysemy, photo sets
- Tags are not enough either!

Video Retrieval

Given a shot from a video…
… is some semantic *concept* present in that shot?

Example concepts:

- Airplane
- Building
- **Car**
- Crowd
- **Desert**
- **Explosion**
- **Outdoor**
- People
- **Vehicle**
- **Violence**

Object/Scene Categories


Aircraft


Animal


Boat


Building


Bus


Car


Chart


Corp. leader


Court


Crowd


Desert


Entertainment


Explosion


Face


Flag USA


Gov. leader


Map


Meeting


Military


Mountain


Natural disaster


Office


Outdoor


People


People marching


Police / security


Prisoner


Screen


Sky


Sports


Studio


Truck

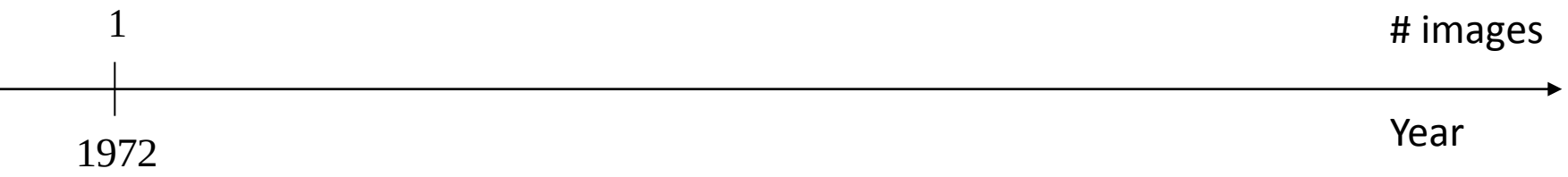

Urban


Vegetation

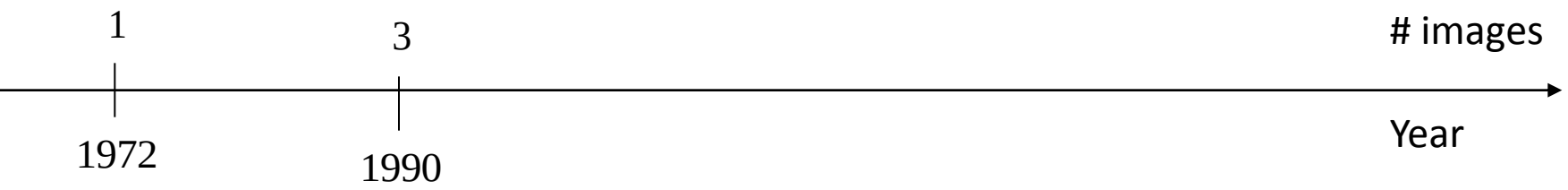

Vehicle


Violence

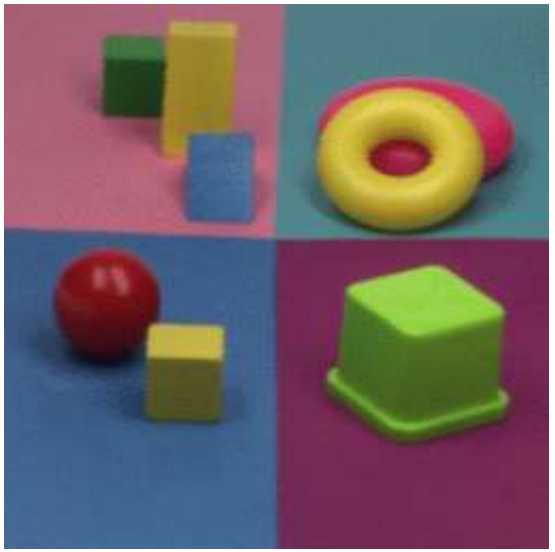
Datasets in perspective



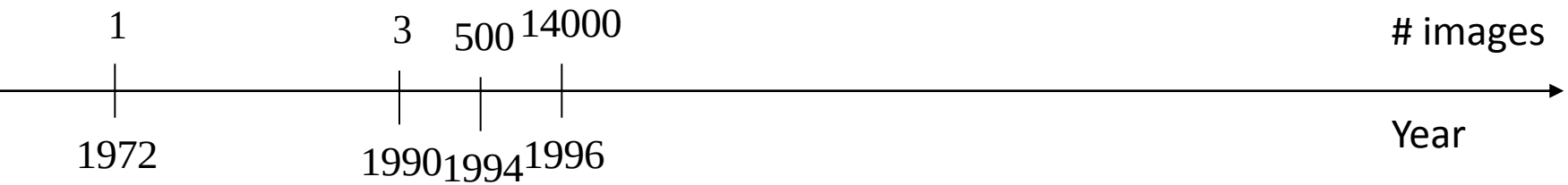
Lena (1972)



“trui” and toy (1990)

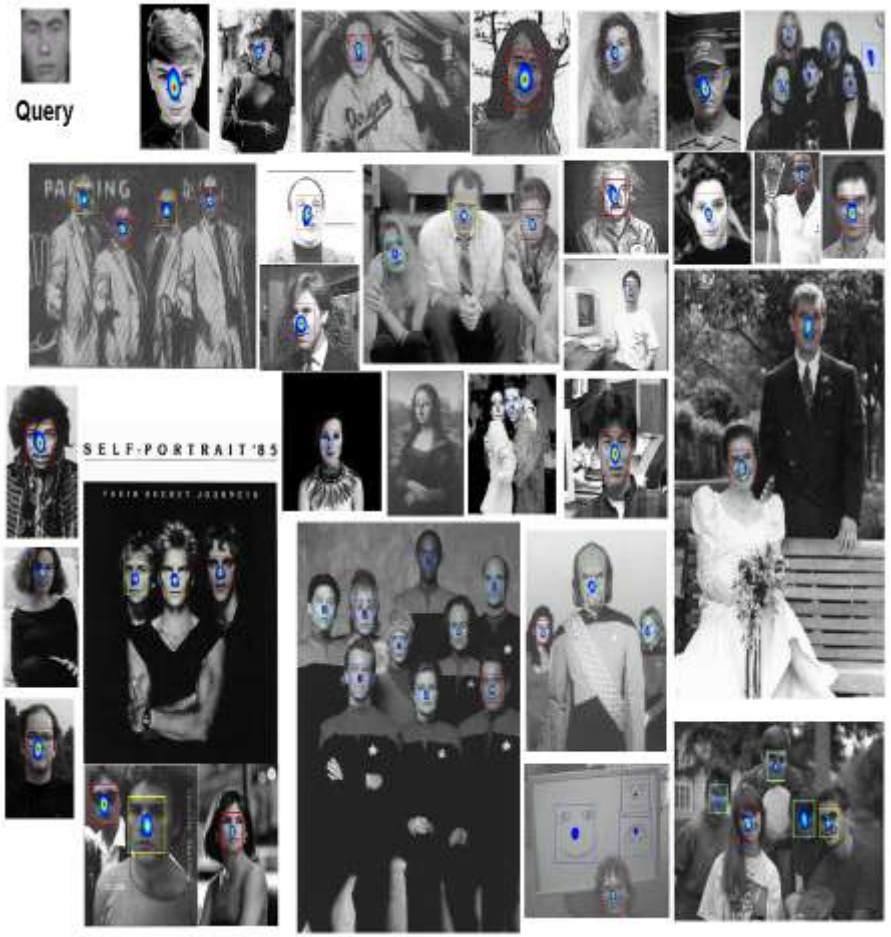


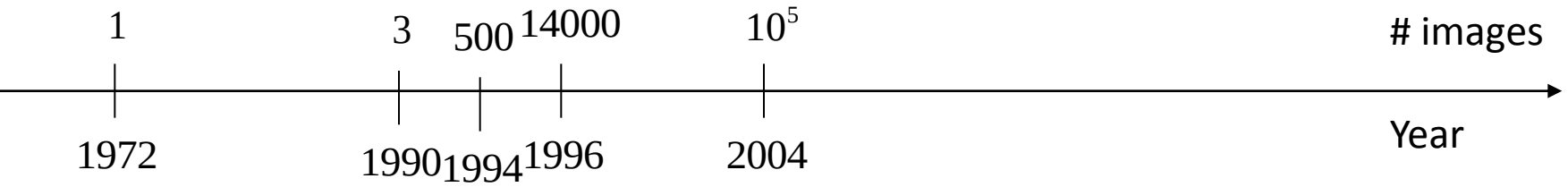

1
3
500
images
1972
19901994
Year



DARPA Faces (1996)

In 1996 DARPA released
14000 images,
from over 1000 individuals.



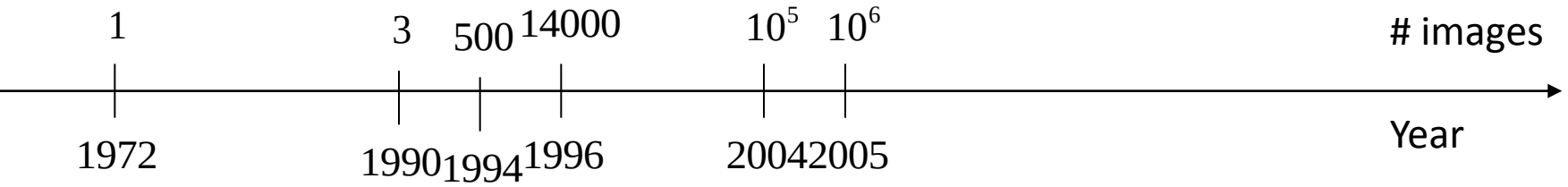


Caltech 101 and 256 (100,000)

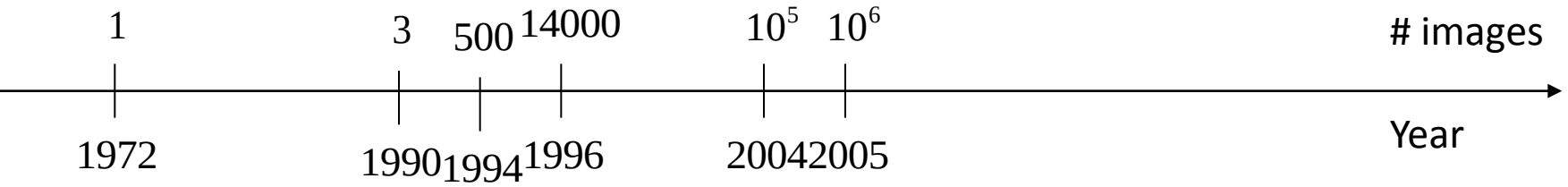
Fei-Fei, Fergus, Perona, 2004

Griffin, Holub, Perona, 2007

101 categories. About 40 to 800 images per category. Most categories have about 50 images. Collected in September 2003



Amsterdam Library of Object Images (ALOI)

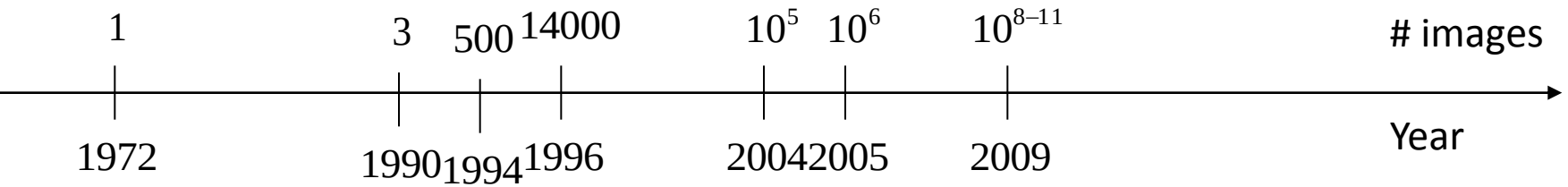


TRECVID and PASCAL VOC competition (2005-2009)

- 86 hours of video from TRECVID 2005
- Shot segmentation available: 43.907 shots
- Ground truth available from Mediamill Challenge

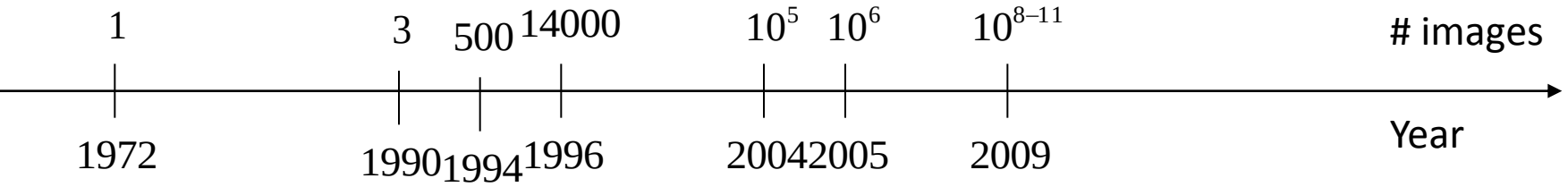
- The goal of VOC challenge is to recognize objects from a number of visual object classes in realistic scenes

- The twenty object classes are:
  - *Person:* person
  - *Animal:* bird, cat, cow, dog, horse, sheep
  - *Vehicle:* aeroplane, bicycle, boat, bus, car, motorbike, train
  - *Indoor:* bottle, chair, dining table, potted plant, sofa, tv/monitor.



> Google over 80,000,000 images

Google™ Image Search



> 1 billion images!

Ask Images

Google Image Search

webshots beta

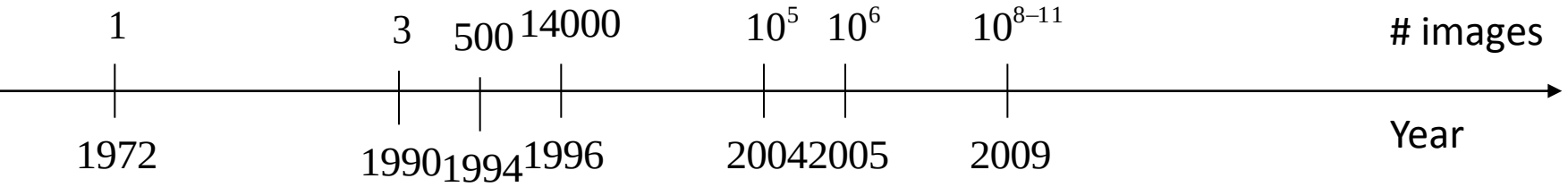
picsearch

flickr GAMMA

Cydral Image & Site Search

altavista

facebook



> 1 billion images!

Google Image Search

altavista

facebook

Number of images seen by all humanity:

1020

106,456,367,669 humans1 * 60 years * 3 images/second * 60 * 60 * 16 * 365 =

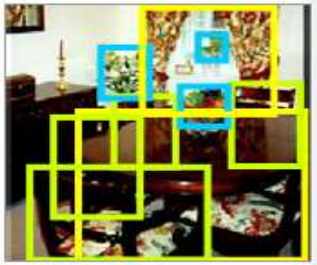
1 from http://www.prb.org/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx

Number of images seen during the first 10 years:

(3 images/second * 60 * 60 * 16 * 365 * 10 = 630,720,000)

Labeling

Labeling to get a Ph.D.



Just labeling

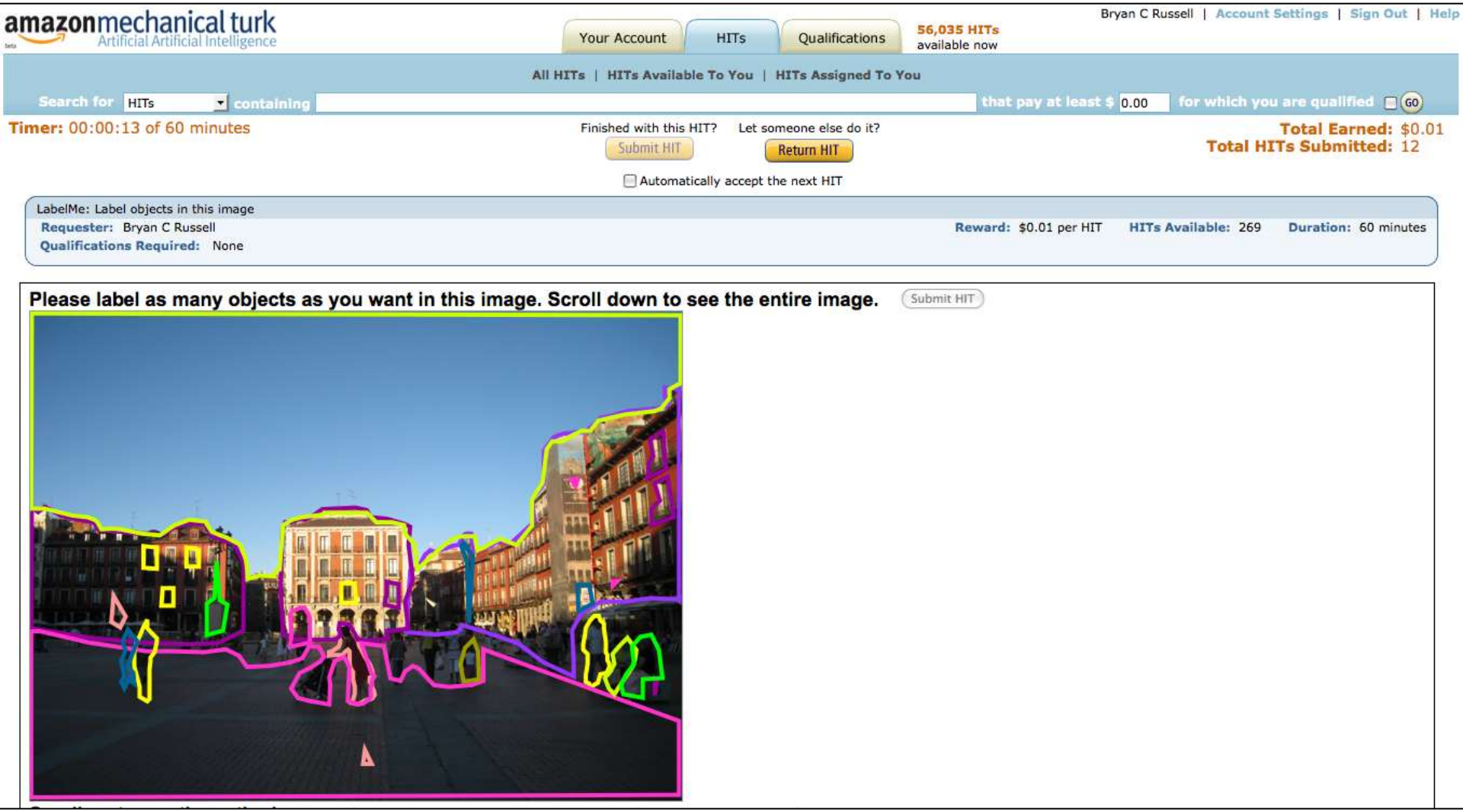
LabelMe

Labeling for fun

Labeling for money

amazonmechanical turk
Artificial Artificial Intelligence

Mechanical Turk

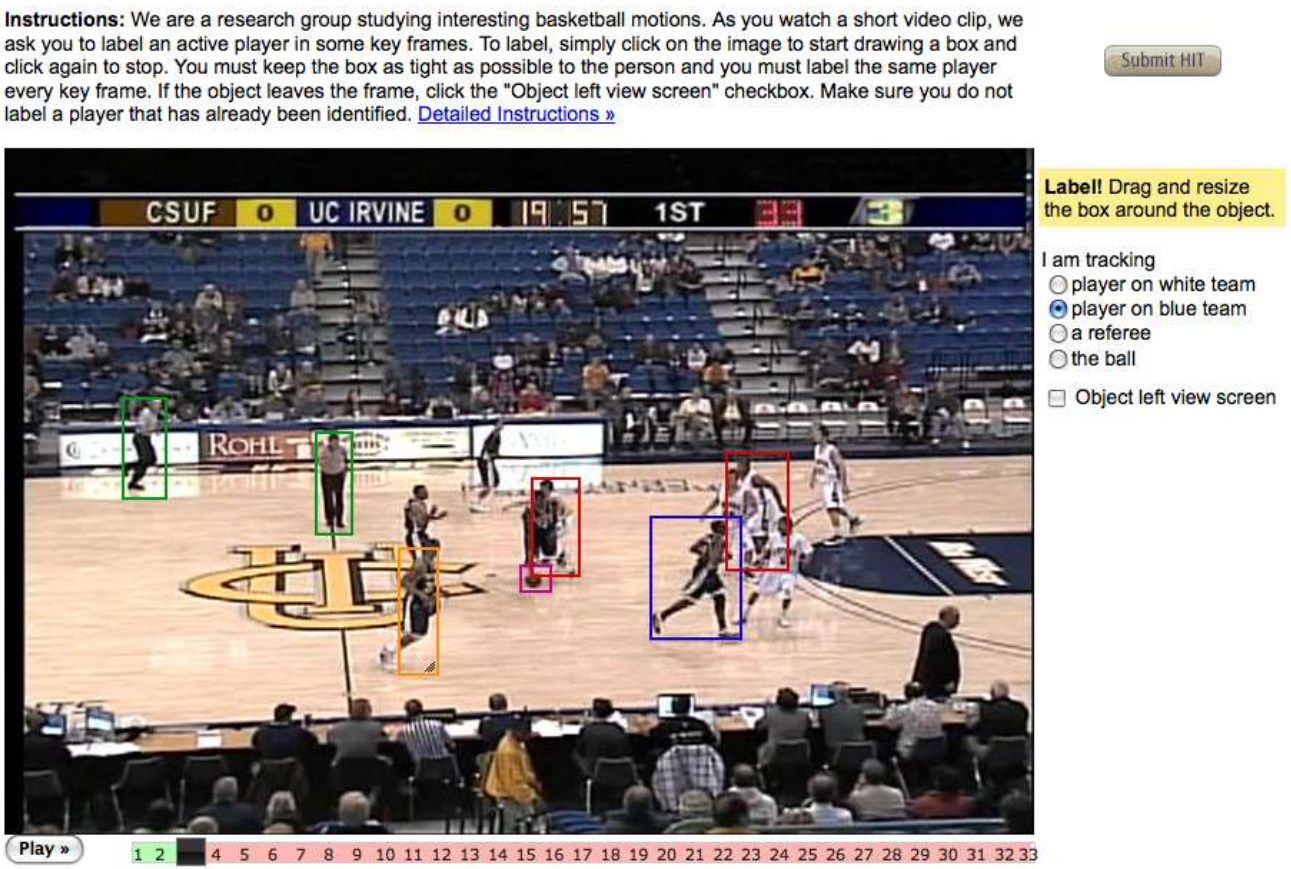


1 cent per image

Task: Label one object in this image

Labeling human activities

Carl Vondrick, Deva Ramanan, Don Patterson



https://workersandbox.mturk.com/mturk/preview?groupId=0YNZVTYH13MZP2ZVKS30

Image and Video Formation

Image search challenges

Challenges

- Viewpoint variation

- Illumination change

- Orientation and scale

Challenges

- Viewpoint variation

- Illumination change

- Orientation and scale

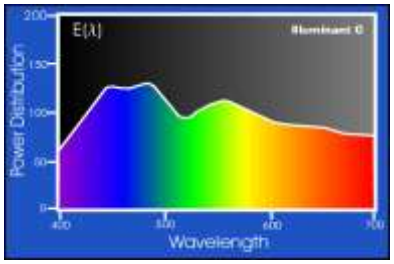
- Occlusion

- Clutter

- Appearance change

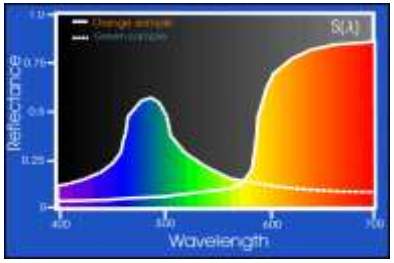
What makes an image

Light source



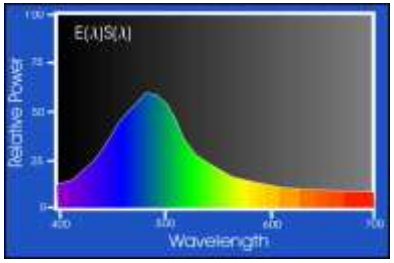
$e(\lambda)$

Object



$\rho(\lambda)$

Sensor



$e(\lambda)\rho(\lambda)$

$$R = \int_{\lambda} e(\lambda)\rho(\lambda) f_R(\lambda) d\lambda, \quad G = \int_{\lambda} e(\lambda)\rho(\lambda) f_G(\lambda) d\lambda, \quad B = \int_{\lambda} e(\lambda)\rho(\lambda) f_B(\lambda) d\lambda$$

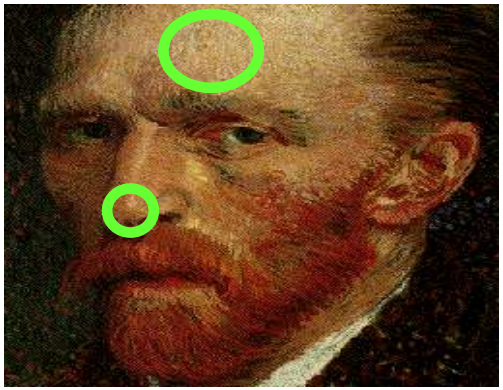
Image Formation

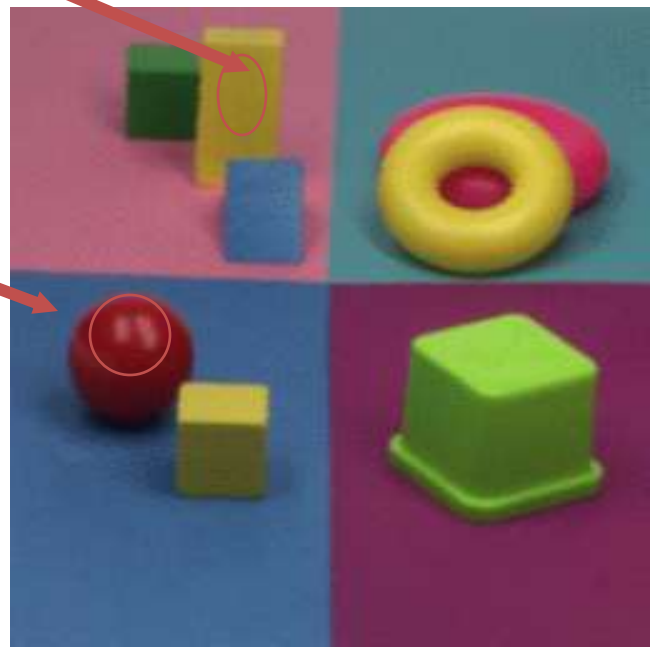
Reflectance model [Shafer]

body = $m_b(\vec{n},\vec{s})\int_{\lambda} f_C(\lambda)e(\lambda)c_b(\lambda)d\lambda$

surface = $m_s(\vec{n},\vec{s},\vec{v})\int_{\lambda} f_C(\lambda)e(\lambda)c_s(\lambda)d\lambda$

for {R,G,B} giving an R-, B-, G-sensor response



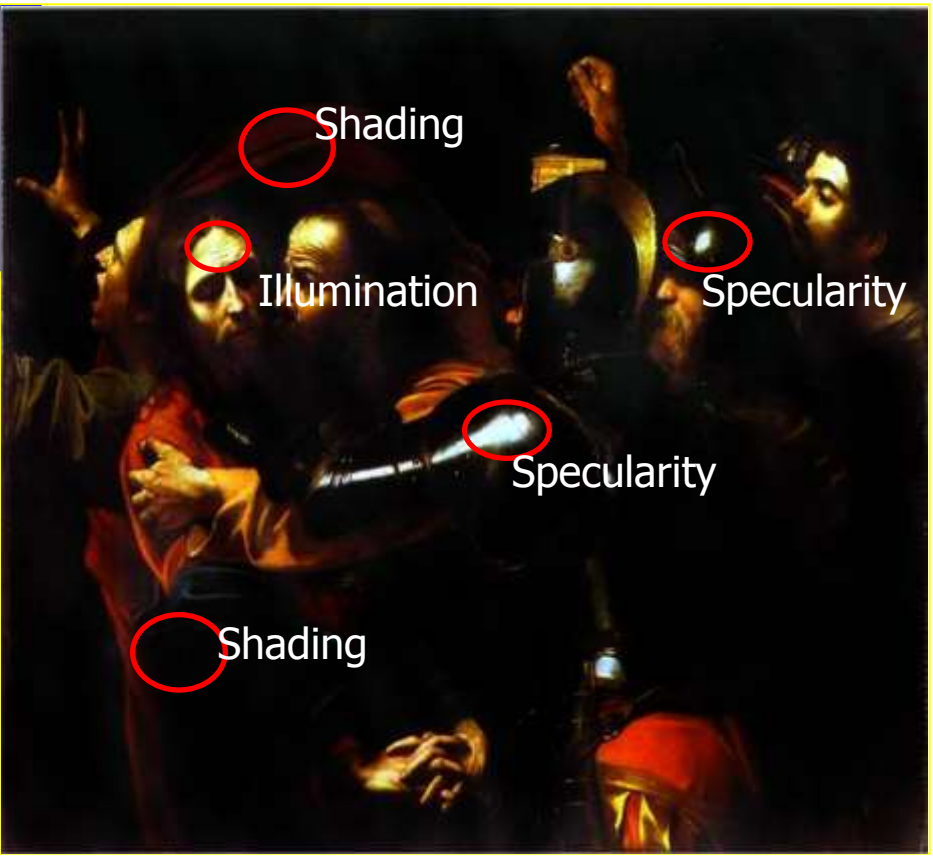


Reflection Model

$$C = m_b(\vec{n},\vec{s})\int_{\lambda} f_C(\lambda)e(\lambda)c_b(\lambda)d\lambda + m_s(\vec{n},\vec{s},\vec{v})\int_{\lambda} f_C(\lambda)e(\lambda)c_s(\lambda)d\lambda$$

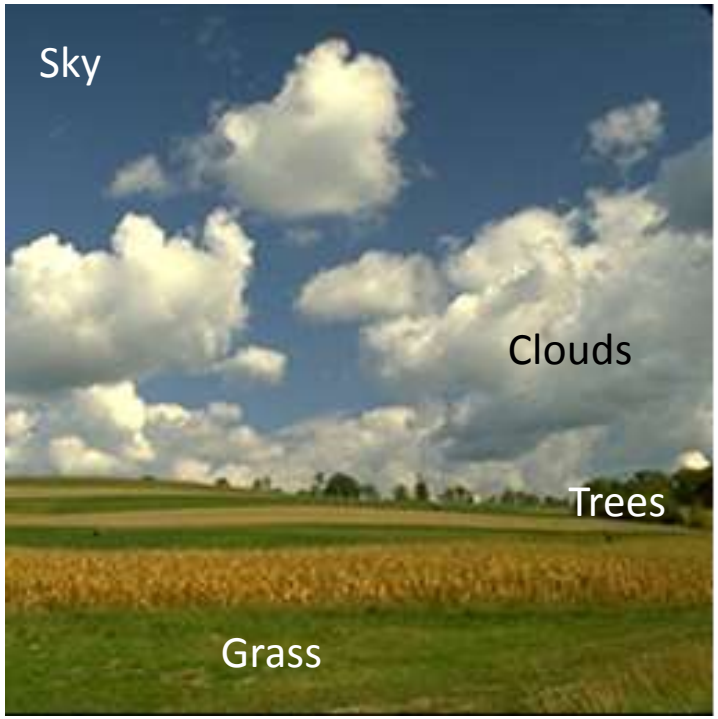
$c_b(\lambda)$	surface albedo	viewpoint invariant
$e(\lambda)$	illumination	scene dependent
$\vec{n}$	object surface normal	object shape variant
$\vec{s}$	illumination direction	scene dependent
$\vec{v}$	viewer's direction	viewpoint variant
$f_C(\lambda)$	sensor sensitivity	scene dependent

Image semantics (low-level)



Categorization (high-level)

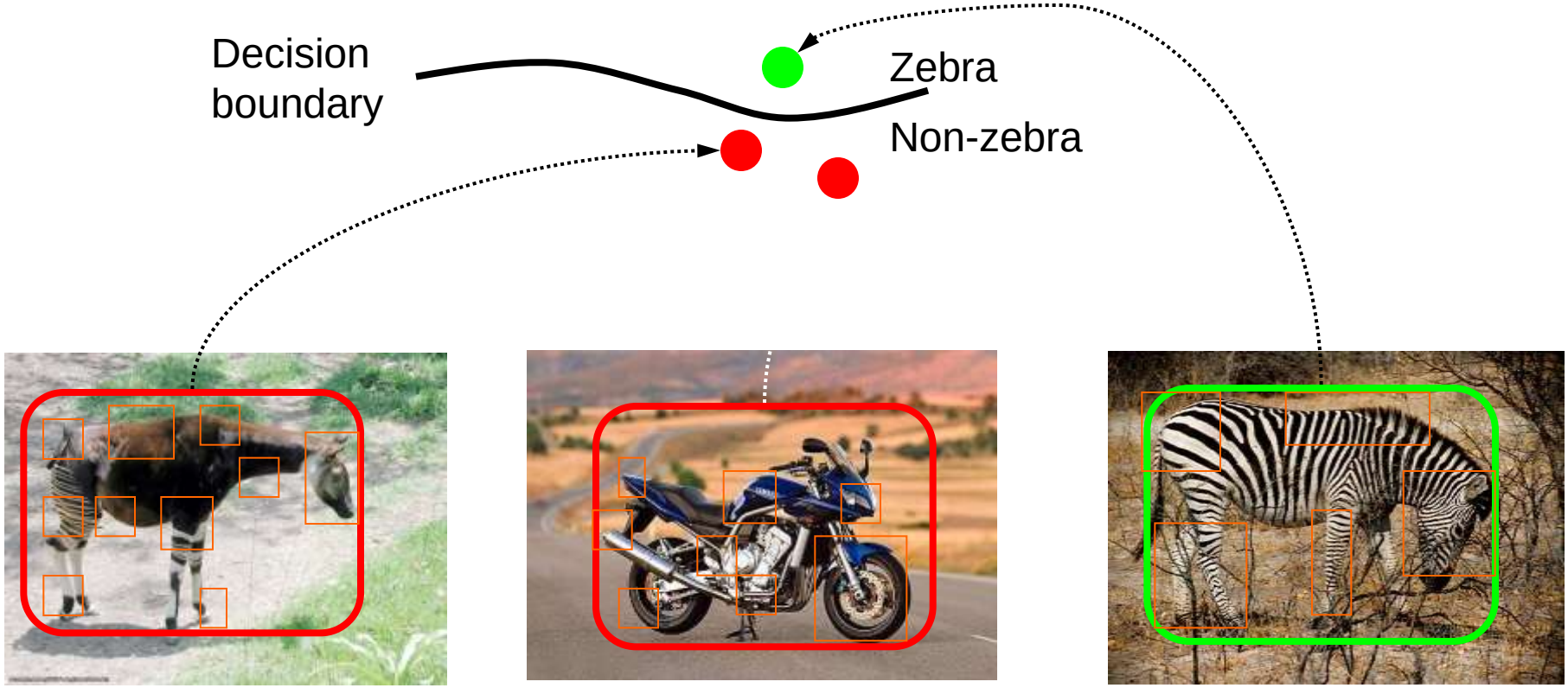
Outdoor/Landscape/vegatation…



Outdoor/city/street…

Machine learning

- Direct modeling of $\frac{p(zebra \mid image)}{p(no\ zebra \mid image)}$

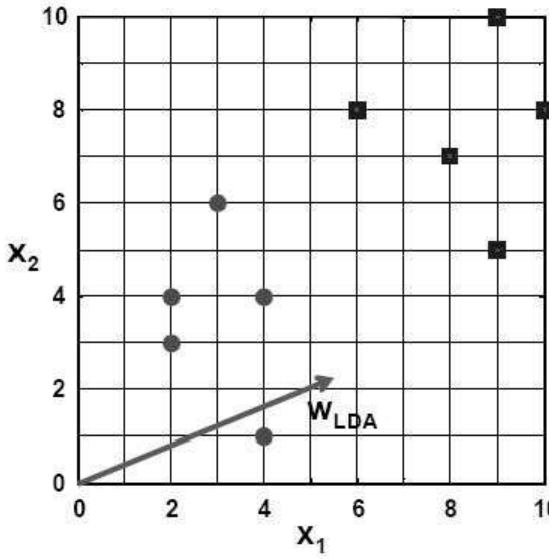


LDA – 2 Classes Example

- **Compute the Linear Discriminant projection for the following two-dimensional dataset**
  - $X1=(x_1,x_2)=\{(4,1),(2,4),(2,3),(3,6),(4,4)\}$
  - $X2=(x_1,x_2)=\{(9,10),(6,8),(9,5),(8,7),(10,8)\}$
- **SOLUTION (by hand)**
  - The class statistics are:

$$S_1=\begin{bmatrix}0.80 & -0.40\\ -0.40 & 2.60\end{bmatrix};\ S_2=\begin{bmatrix}1.84 & -0.04\\ -0.04 & 2.64\end{bmatrix}$$

$$\mu_1=\begin{bmatrix}3.00 & 3.60\end{bmatrix};\qquad \mu_2=\begin{bmatrix}8.40 & 7.60\end{bmatrix}$$



  - The within- and between-class scatter are

$$S_B=\begin{bmatrix}29.16 & 21.60\\ 21.60 & 16.00\end{bmatrix};\ S_W=\begin{bmatrix}2.64 & -0.44\\ -0.44 & 5.28\end{bmatrix}$$

  - The LDA projection is then obtained as the solution of the generalized eigenvalue problem

$$S_W^{-1}S_B v=\lambda v \Rightarrow \left|S_W^{-1}S_B-\lambda I\right|=0 \Rightarrow \begin{vmatrix}11.89-\lambda & 8.81\\ 5.08 & 3.76-\lambda\end{vmatrix}=0 \Rightarrow \lambda=15.65$$

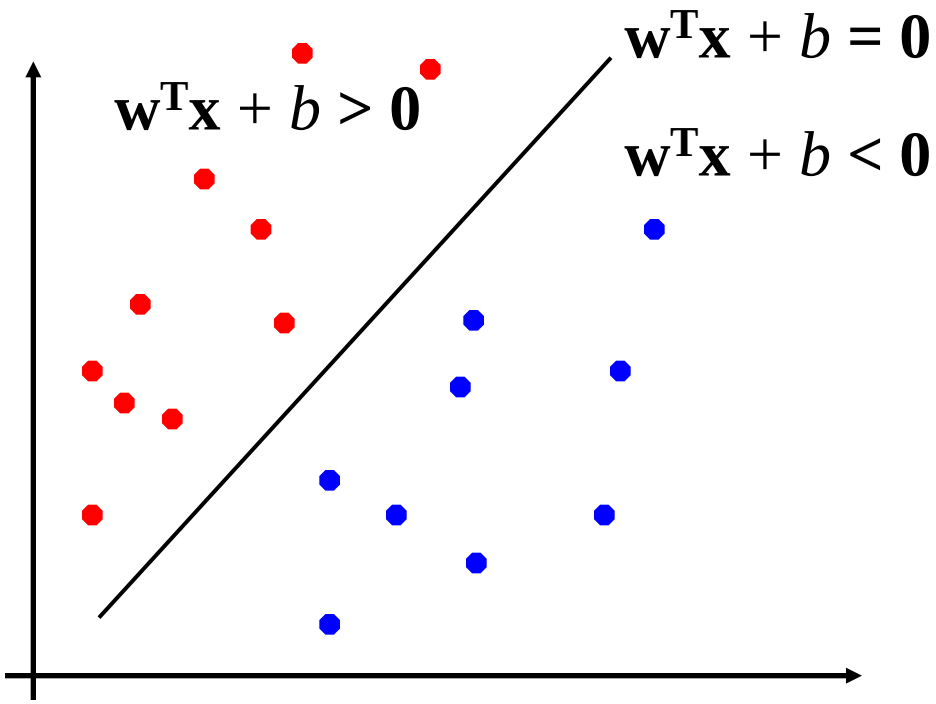
$$\begin{bmatrix}11.89 & 8.81\\ 5.08 & 3.76\end{bmatrix}\begin{bmatrix}v_1\\ v_2\end{bmatrix}=15.65\begin{bmatrix}v_1\\ v_2\end{bmatrix}\Rightarrow\begin{bmatrix}v_1\\ v_2\end{bmatrix}=\begin{bmatrix}0.91\\ 0.39\end{bmatrix}$$

  - Or directly by

$$w^*=S_W^{-1}(\mu_1-\mu_2)=\begin{bmatrix}-0.91 & -0.39\end{bmatrix}^T$$

Linear Separators

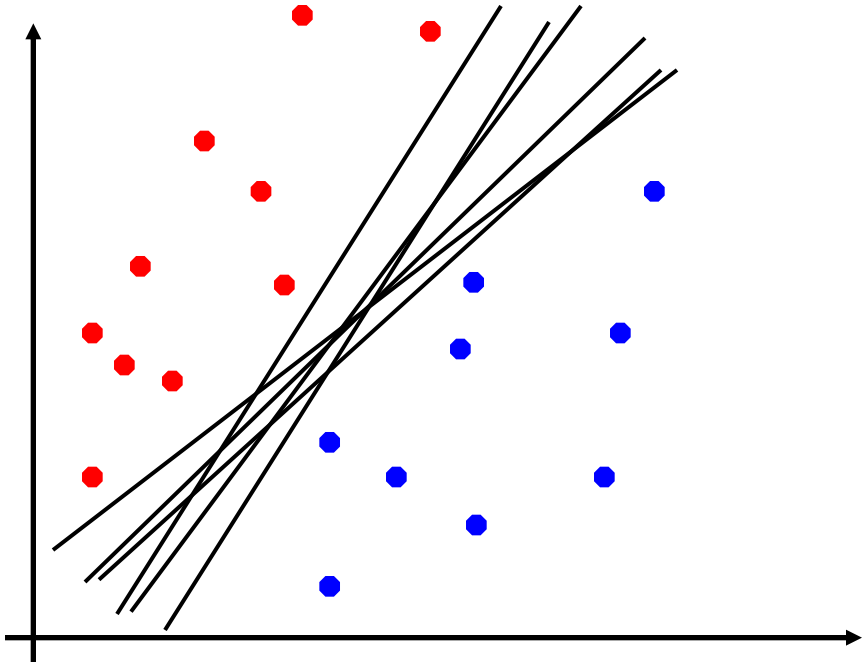
Binary classification can be viewed as the task of separating classes in feature space:



$$f(\mathbf{x}) = \text{sign}(\mathbf{w}^T\mathbf{x} + b)$$

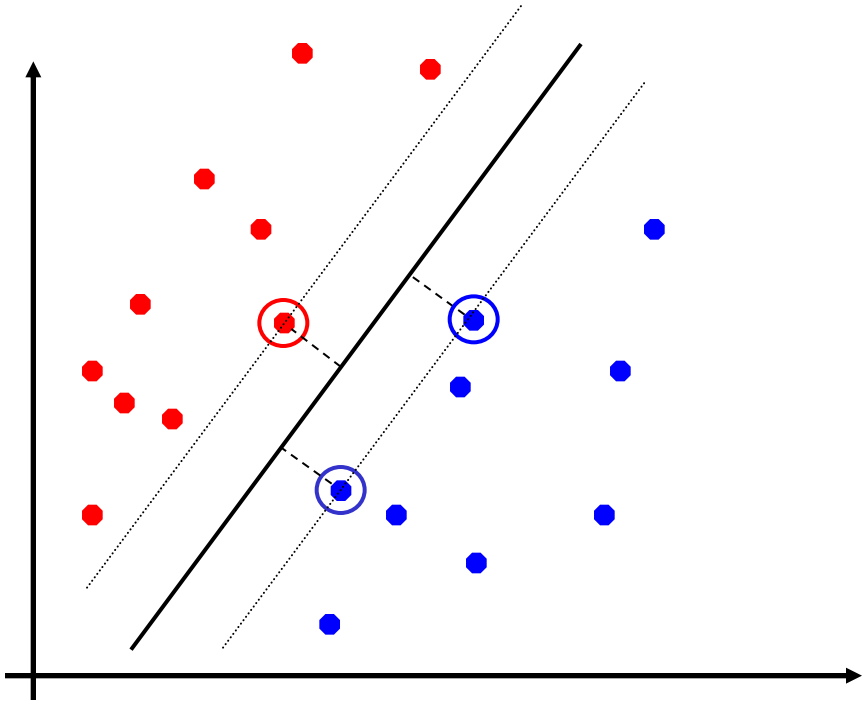
Linear Separators

Which of the linear separators is optimal?

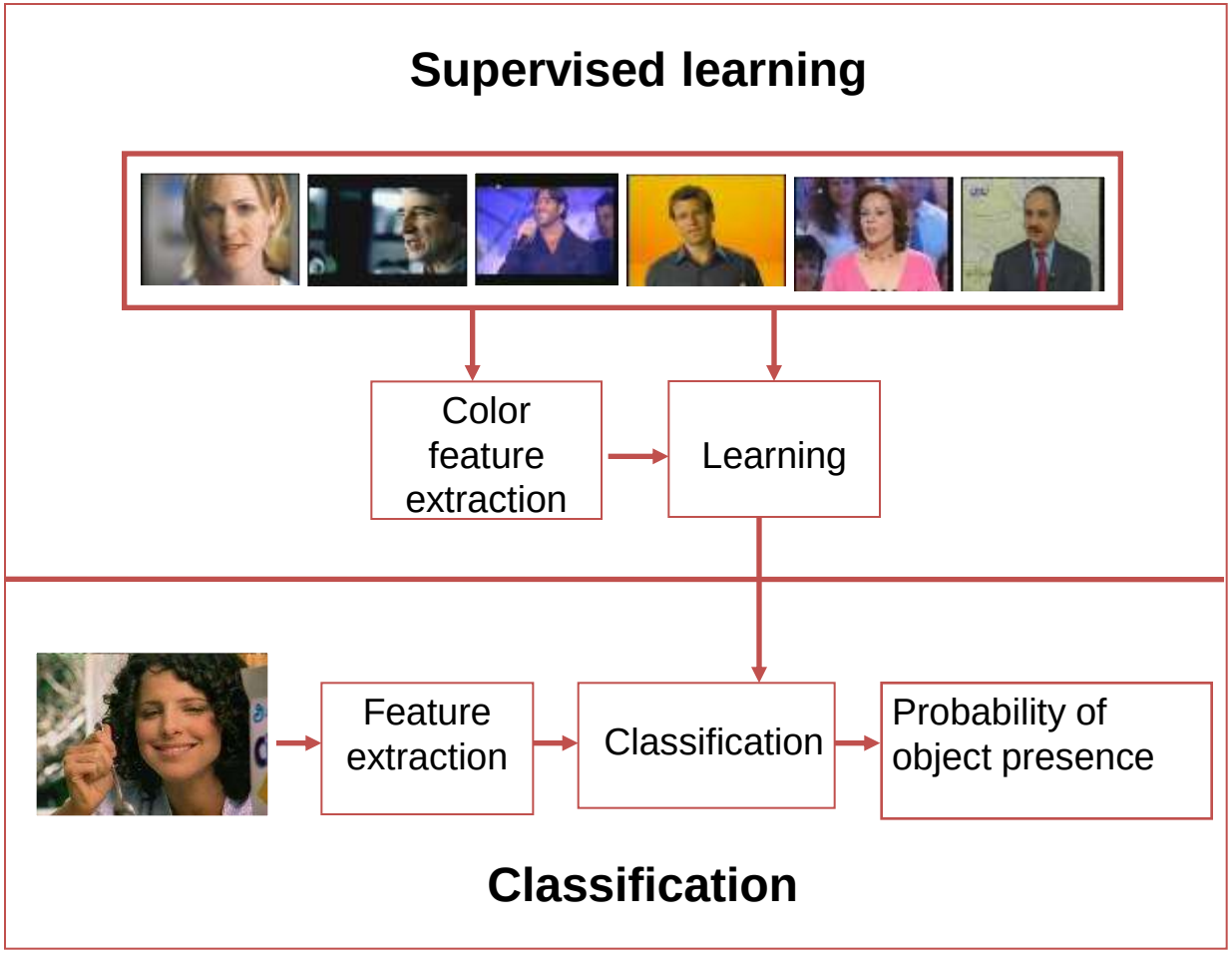


Maximum Margin Classification: SVM

- Maximizing the margin is good according to intuition and PAC theory.
- Implies that only support vectors matter; other training examples are ignorable.



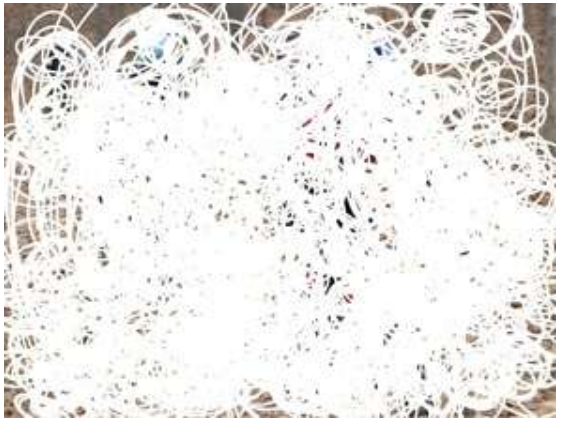
Recognition scheme



Invariance

•Original Image

•Harris Laplacian
impl. by Mikolajczyk (e.g. CVPR06)



•Shape adapted Harris Laplacian
impl. by Mikolajczyk (ICCV07)

•Color salient points
Quasi invariant HSI

•Color salient points
Color boosted OCS

•Structured backgrounds of same color tones and shadowing effects are discarded effectively.

Photometric Analysis

Photometric Analysis (2)

2.

Real-world Datasets

Company	Application	Data type	Complexity
Ilse Media	Image retrieval on internet	Visual/textual	unstructured
ANP Press	Photostock	Visual/category	structured
VNU/Nielsen	Image tracking/monitoring	Visual	unstructured
Akzo Nobel	Coating effect classification	Visual	structured
Forensic lab	XTC-pill image classification	Visual	structured
Philips	People tracking	Visual	structured
SBS	Billboard tracking/replacement	Visual	unstructured

...

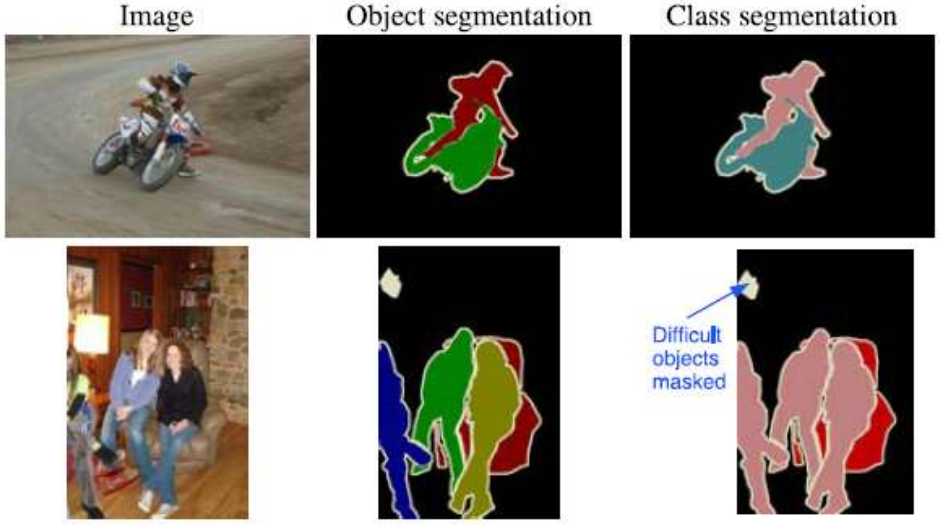
PASCAL VOC

Year	Statistics	New developments	Notes
2005	Only 4 classes: bicycles, cars, motorbikes, people. Train/validation/test: 1578 images containing 2209 annotated objects.	Two competitions: classification and detection	Images were largely taken from exising public datasets, and were not as challenging as the flickr images subsequently used. This dataset is obsolete.
2006	10 classes: bicycle, bus, car, cat, cow, dog, horse, motorbike, person, sheep. Train/validation/test: 2618 images containing 4754 annotated objects.	Images from flickr and from Microsoft Research Cambridge (MSRC) dataset	The MSRC images were easier than flickr as the photos often concentrated on the object of interest. This dataset is obsolete.
2007	20 classes: • *Person*: person • *Animal*: bird, cat, cow, dog, horse, sheep • *Vehicle*: aeroplane, bicycle, boat, bus, car, motorbike, train • *Indoor*: bottle, chair, dining table, potted plant, sofa, tv/monitor Train/validation/test: 9,963 images containing 24,640 annotated objects.	• Number of classes increased from 10 to 20 • Segmentation taster introduced • Person layout taster introduced • Truncation flag added to annotations • Evaluation measure for the classification challenge changed to Average Precision. Previously it had been ROC-AUC.	This year established the 20 classes, and these have been fixed since then. This was the final year that annotation was released for the testing data.
2008	20 classes. The data is split (as usual) around 50% train/val and 50% test. The train/val data has 4,340 images containing 10,363 annotated objects.	• Occlusion flag added to annotations • Test data annotation no longer made public.	

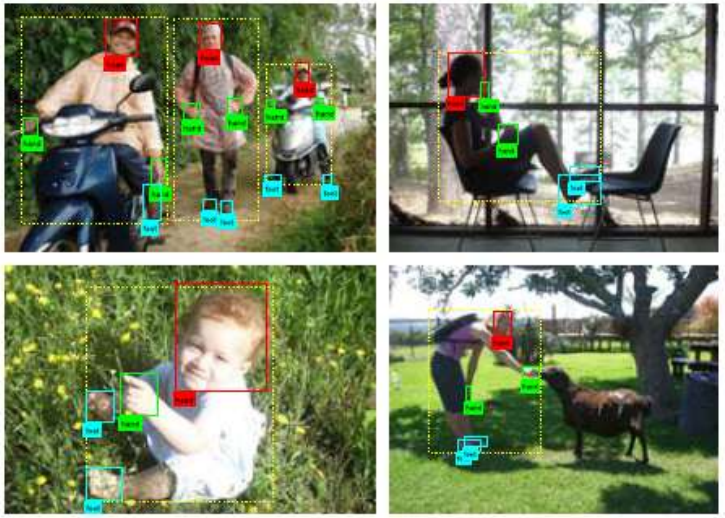
PASCAL VOC

Year	Statistics	New developments	Notes
2009	20 classes. The train/val data has 7,054 images containing 17,218 ROI annotated objects and 3,211 segmentations.	• From now on the data consists of the previous years' images augmented with new images. In earlier years an entirely new data set was released each year. • Augmenting allows the number of images to grow each year, and means that test results can be compared on the previous years' images. • Segmentation becomes a standard challenge (promoted from a taster)	• No difficult flags were provided for the additional images (an omission). • Test data annotation not made public.
2010	20 classes. The train/val data has 10,103 images containing 23,374 ROI annotated objects and 4,203 segmentations.	• Action Classification taster introduced. • Associated challenge on large scale classification introduced based on ImageNet. • Amazon Mechanical Turk used for early stages of the annotation.	• Method of computing AP changed. Now uses all data points rather than TREC style sampling. • Test data annotation not made public.

PASCAL VOC



(a) Segmentation taster



(b) Person layout taster

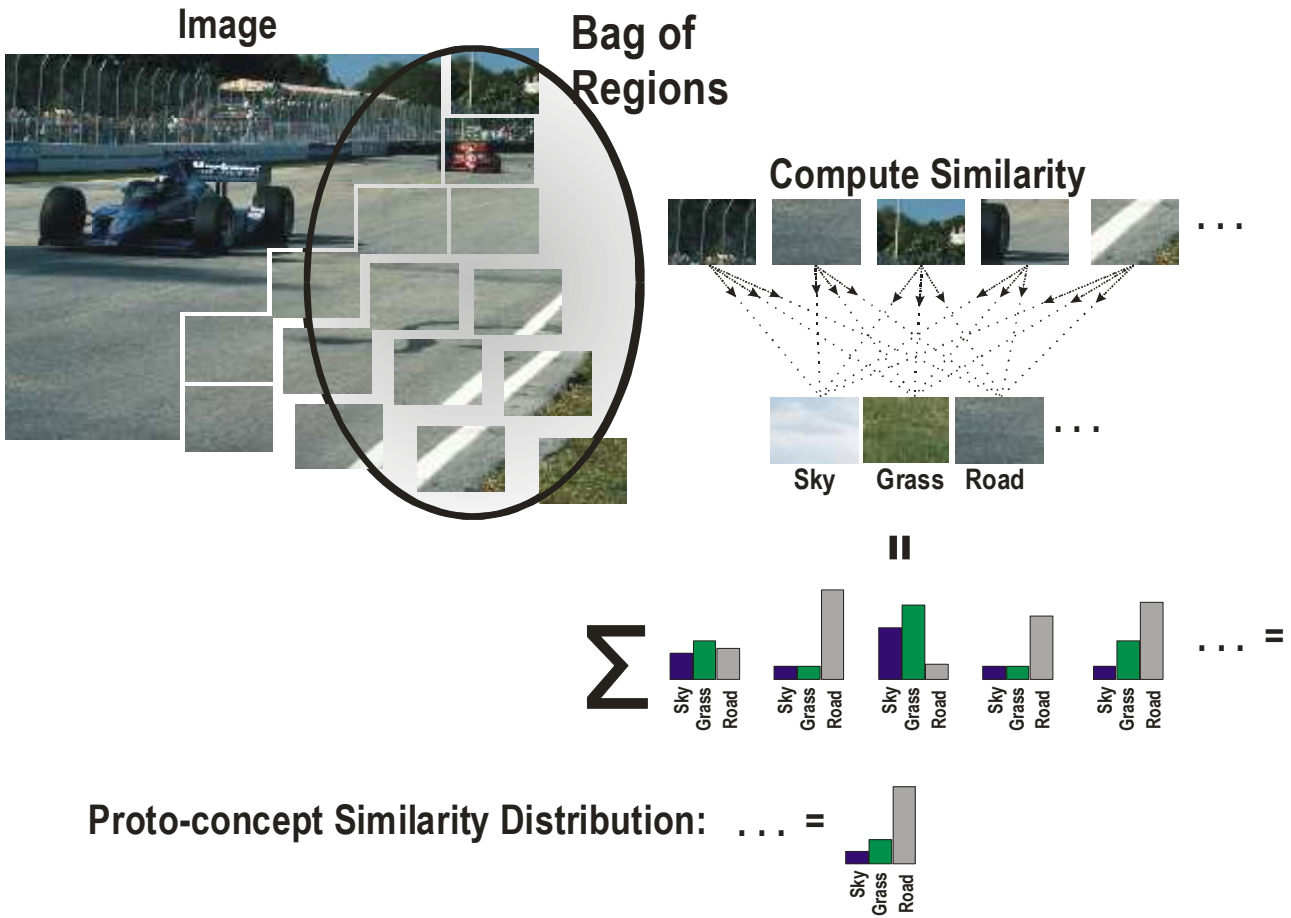
Feature Extraction and Object Recognition

Video Retrieval – Scene

- How to recognize a scene? Context!

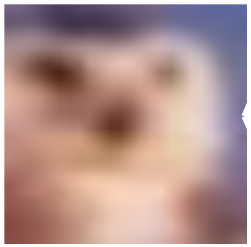
- Use Proto-Concepts to Describe Context
- Machine Learning to link Context to Concepts

Video Retrieval – Scene



SIFT

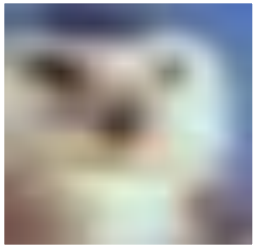
Feature Detection



Shape Description

SIFT

Color Description



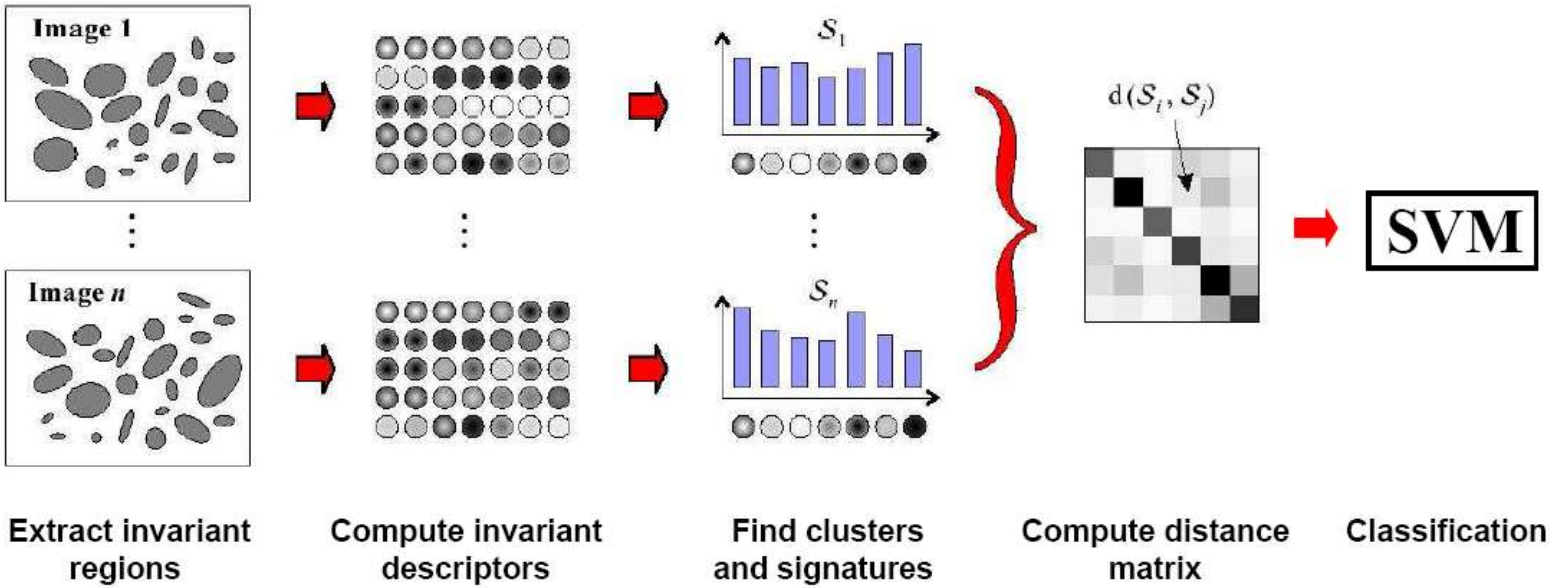
rgb

From: Van de Weijer 2006

Invariance and Learning

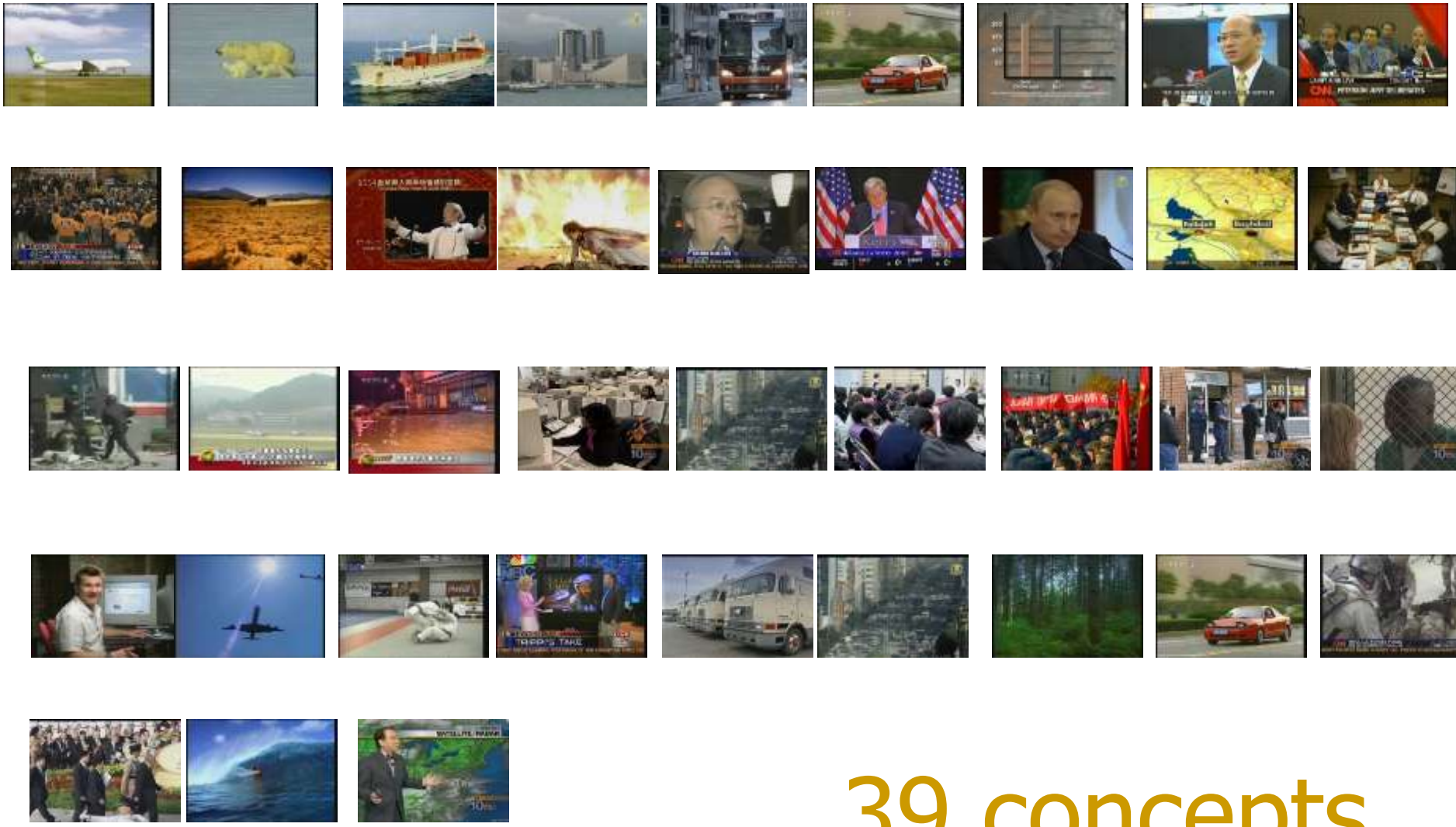
Learn from the content-based image retrieval field: Zhang (2005) achieved state of the art performance using local features:

- invariant region detectors
- SIFT descriptor



Dataset: news broadcasts

- 85 hours of data of TRECVID 2005
- Shot segmentation available: 43.907 shots
- Concept annotations available



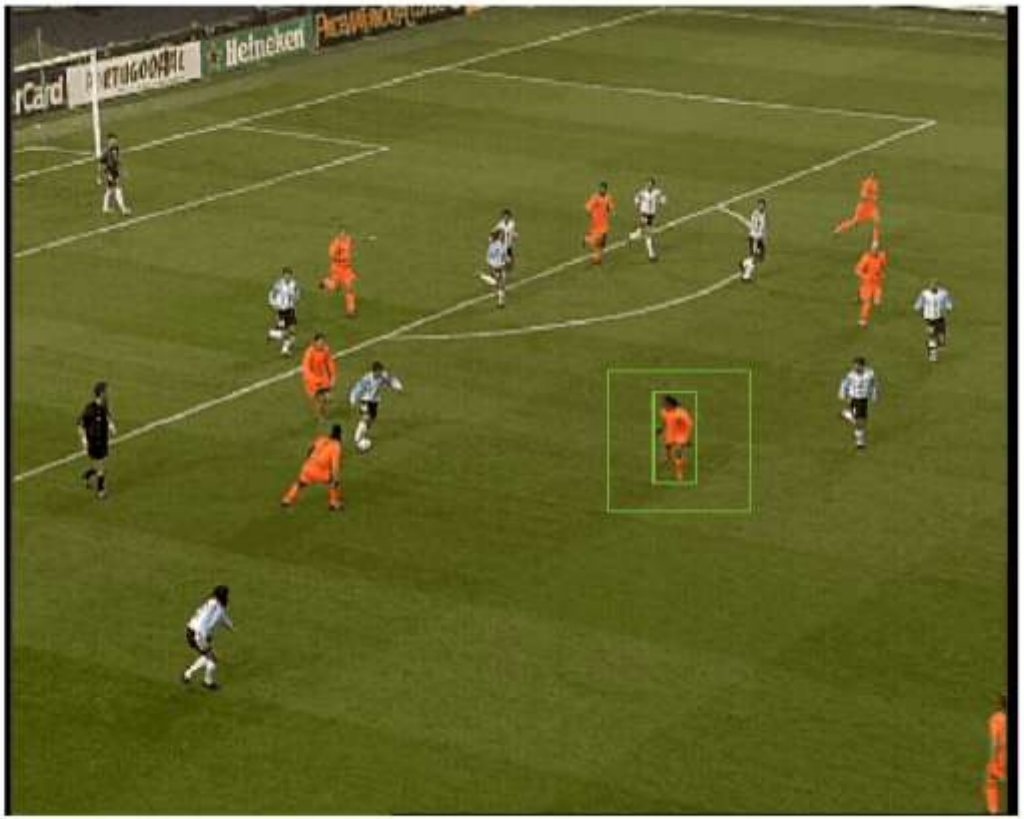
39 concepts

Learning and tracking

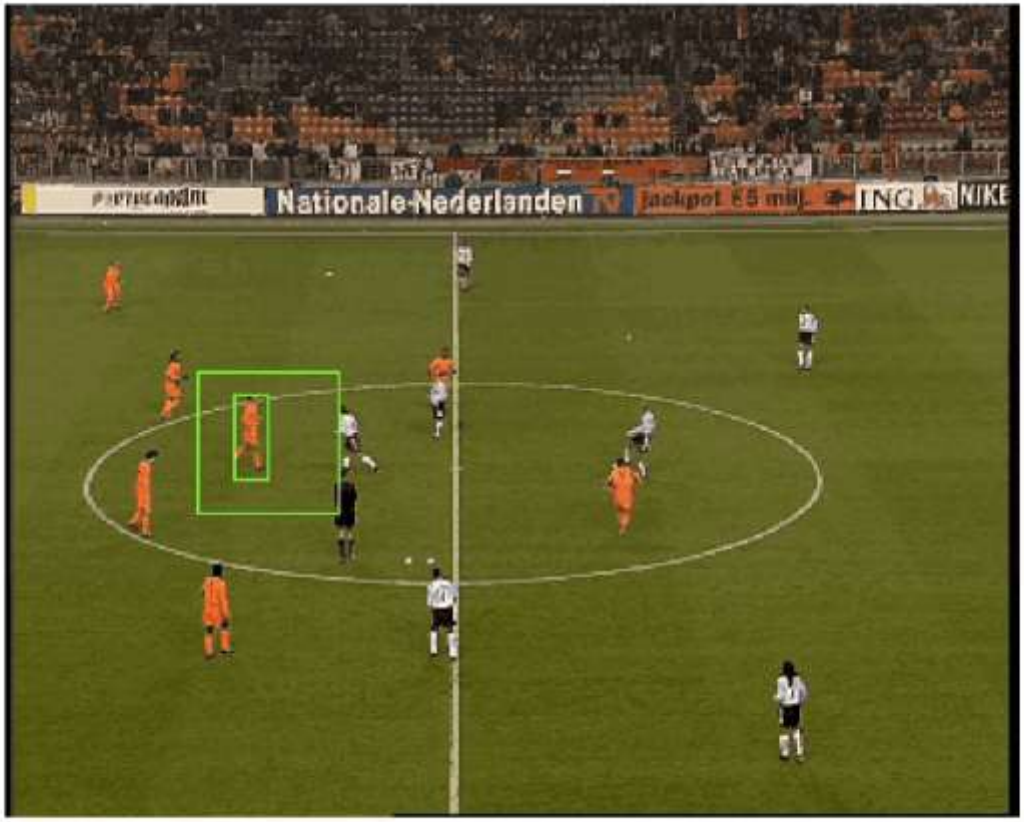
Demo1: Real-time Object Tracking

Demo2: Real-time Object Tracking
Mean-shift

Demo3: Real-time Object Tracking
Foreground/background separation

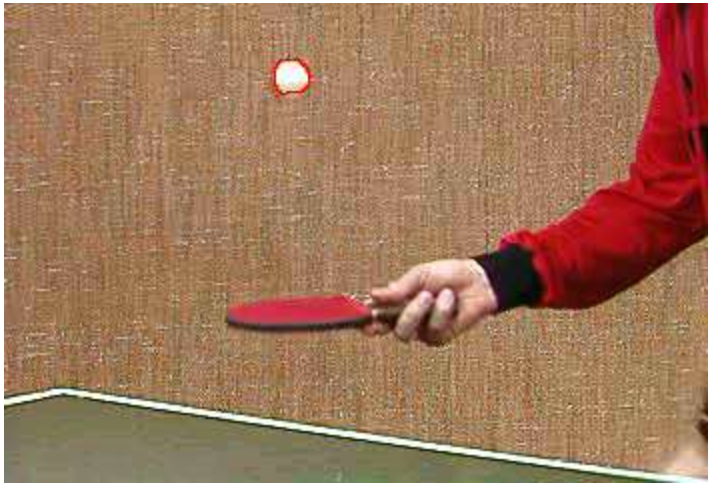


Demo4: Real-time Object Tracking
Foreground/background separation

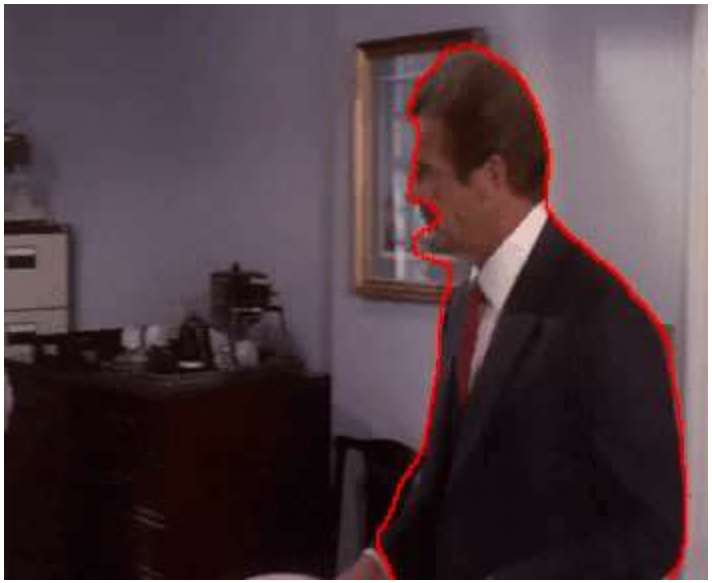


Demo5: Real-time Object Tracking
Deformable Contours

Demo6: Real-time Object Tracking
Deformable Contours



Demo7: Human Recognition and Tracking



Robust to background clutter and changing object appearance

Demo8: Appearance Changes

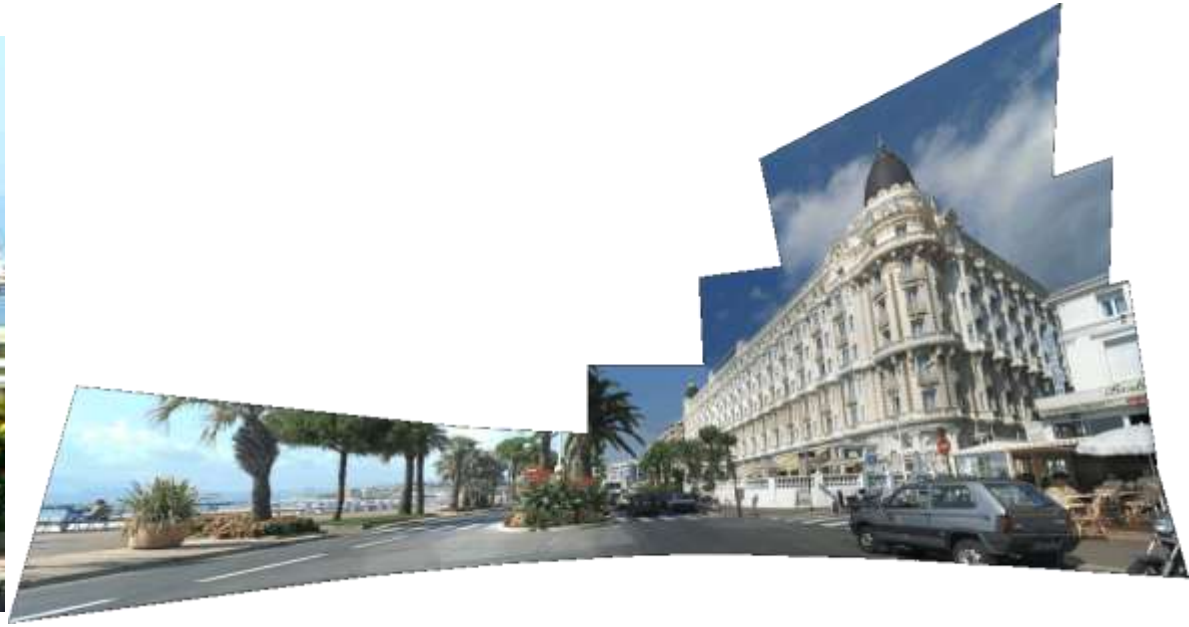
Robust to background clutter and changing object appearance

Demo9: Background Removal

Demo10: Object Detection and Replacement

Demo11: Object Detection and Replacement

Demo13: Mosaicking



Demo14: Mosaicking

Feike Winkelman

Demo15: Mosaicking and Tracking

Techniques:

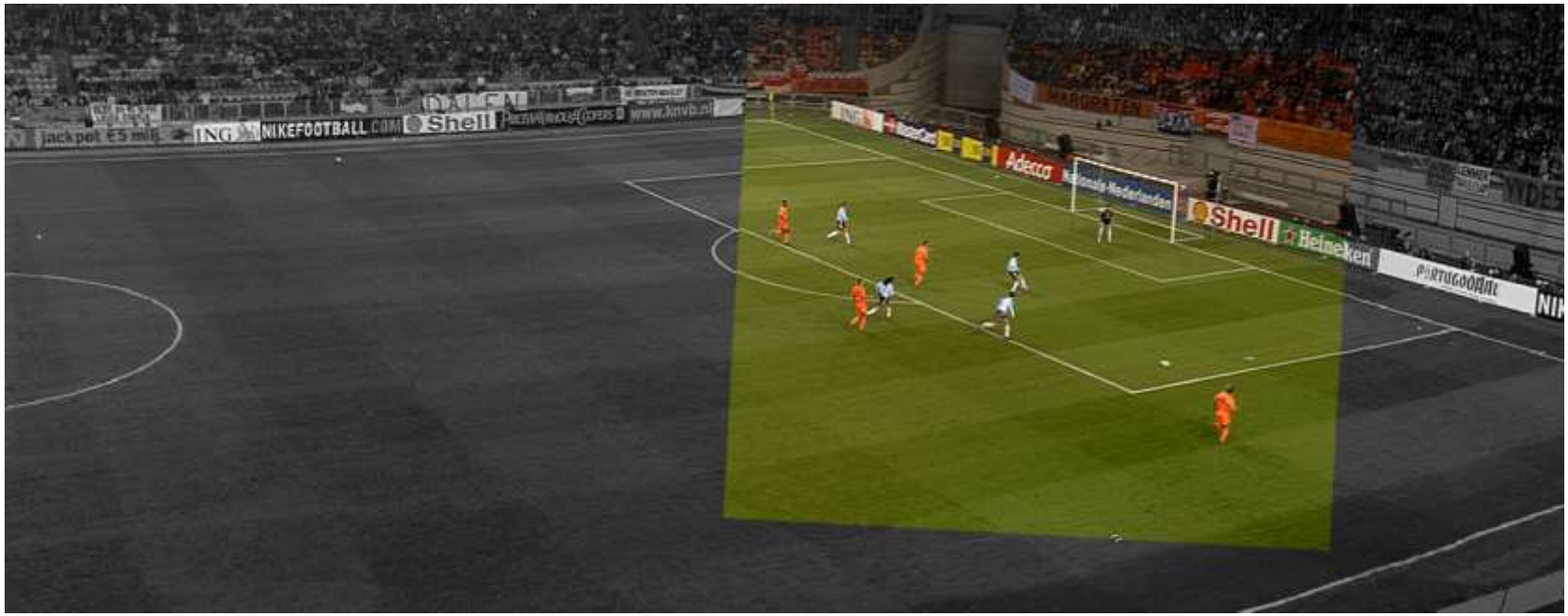
- Mosaics.
- Shot and key-frame detection.
- Analysis of camera-motion.

Sport Statistics

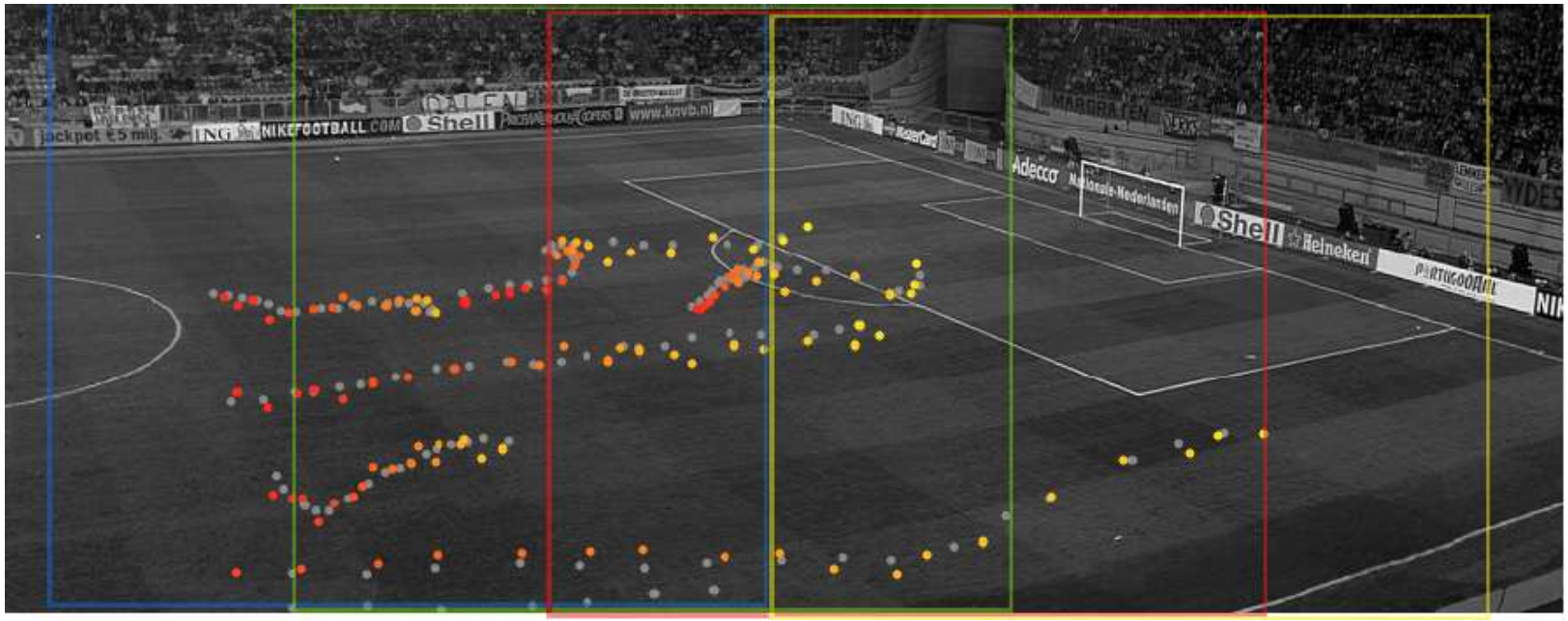
Sport Statistics

Sport Statistics

Sport Statistics



Sport Statistics



•Trajectories of Dutch players, colour-coded in orange. The ground- truth is also shown in grey dots.

•The sub-mosaics for the decomposition are represented by the coloured rectangles.

Sport Statistics

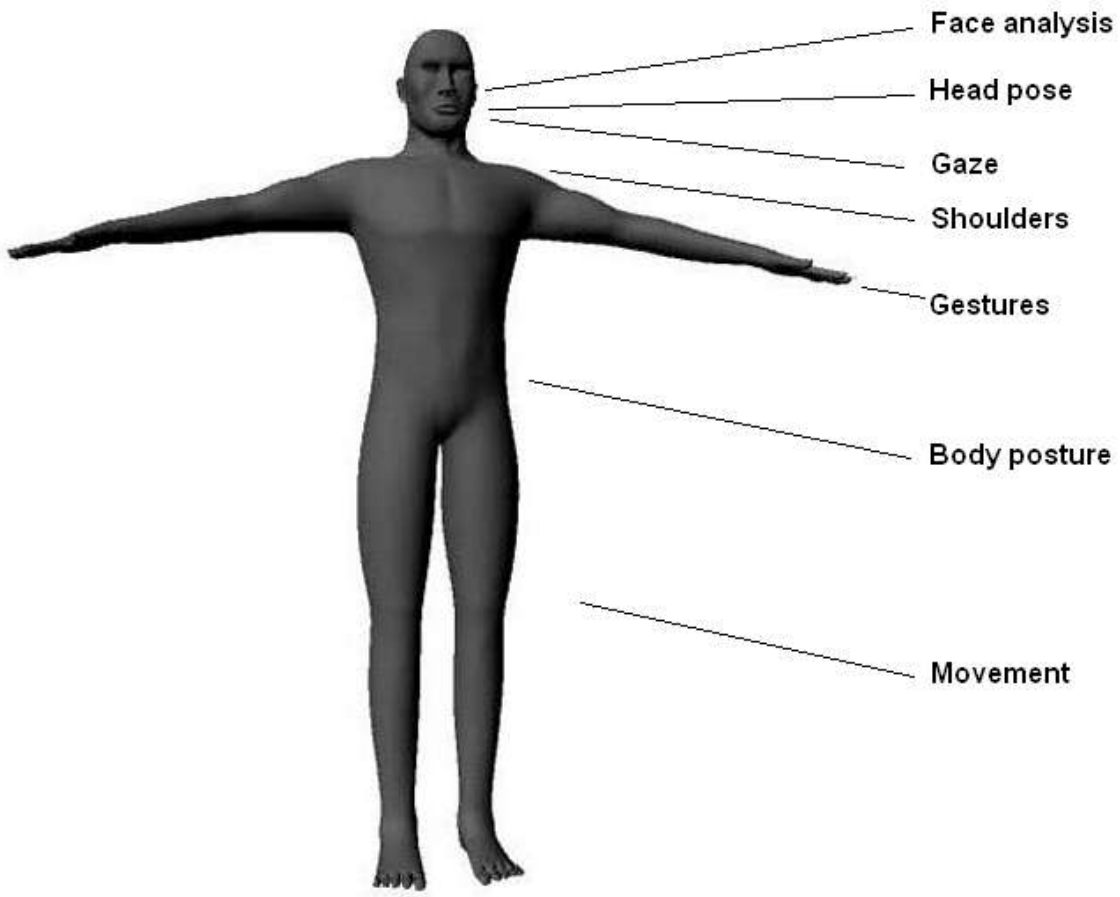
- Several frames projected on the mosaic, according to their recovered registration parameters.
- Showing ‘ghosts’ of players is very illustrative

Sport Statistics

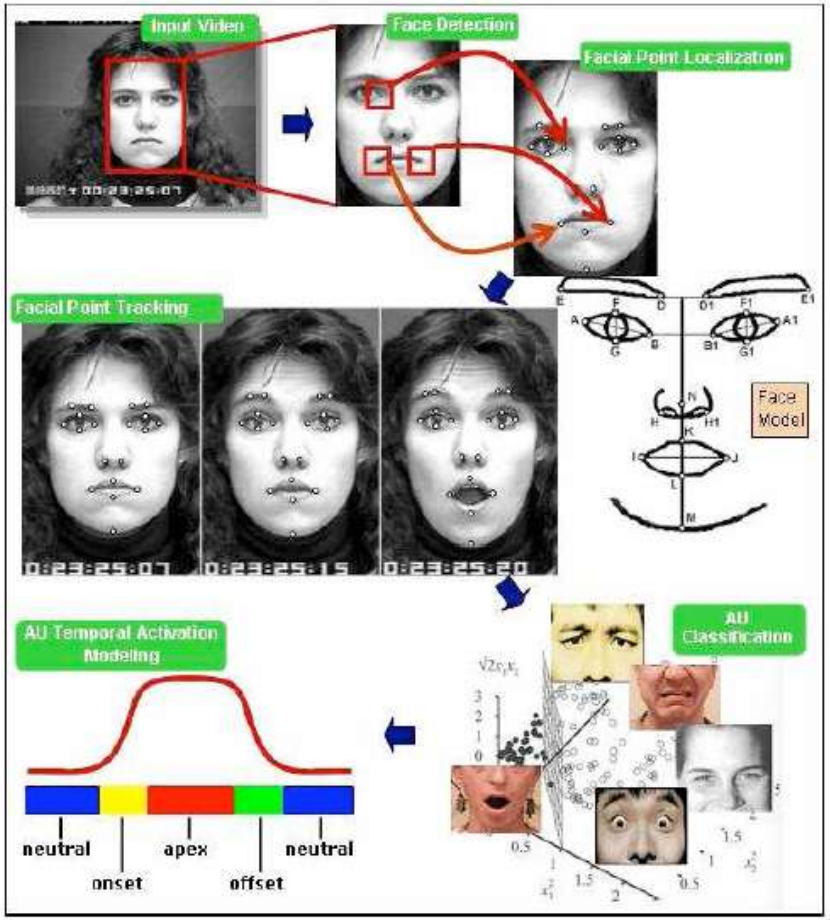
Human behaviour analysis

Activity Recognition

Visual analysis of the human body



Facial expression



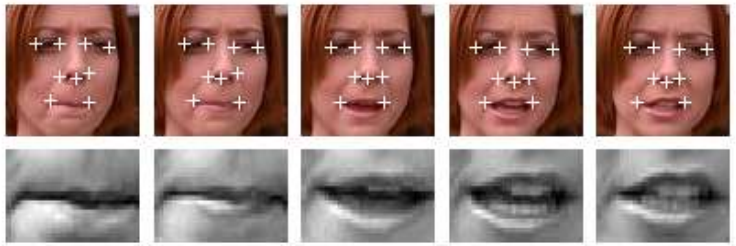
Vinciarelli, Pantic, Bourlard, 2009


Buffy
Joyce
Dawn
Buffy
Buffy
Giles
Other
Tara
Willow
Willow
Buffy
Dawn
Willow
Tara
Willow
Willow
Buffy
Dawn
Tara
Buffy
Willow
Tara
Xander
Anya
Buffy
Riley
Spike

Facial expression

(a) Face detections in original frames (b) Localized facial features

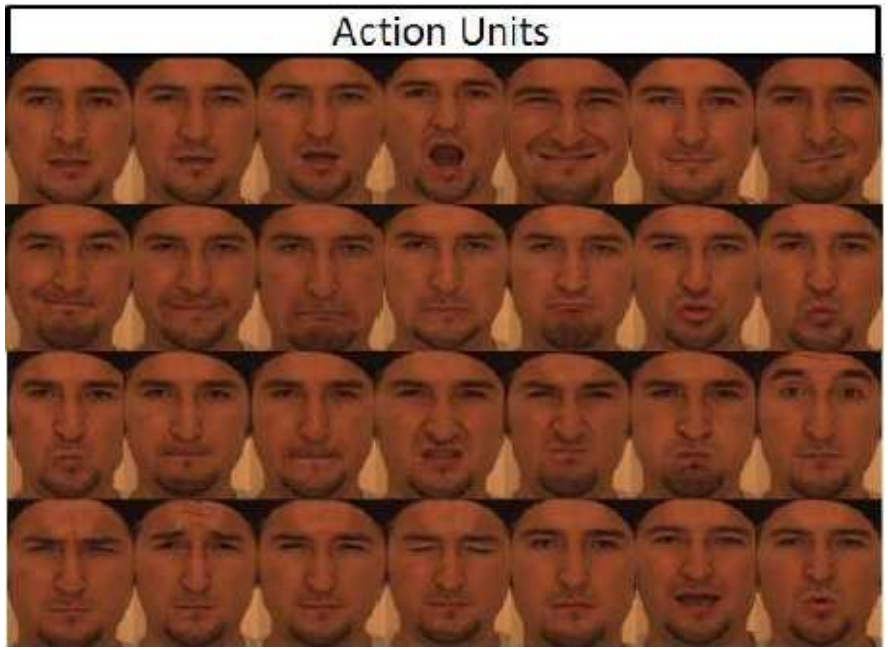
Figure 3: Matching characters across shots using clothing appearance. In the two examples shown the face is difficult to match because of the variation in pose, facial expression and motion blur. The strongly coloured clothing allows correct matches to be established in these cases.



(b)

Figure 5: Speaker identification by detecting lip movement.

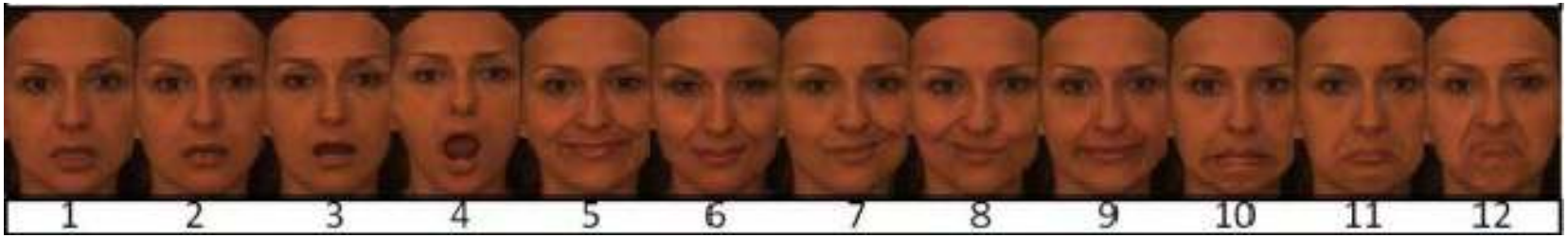
Facial Action Units



Savran et al., “3D Face Recognition Performance under Adversorial Conditions”, *Proc. Enterface Workshop on Multimodal Interfaces*, 2007

Facial Action Units

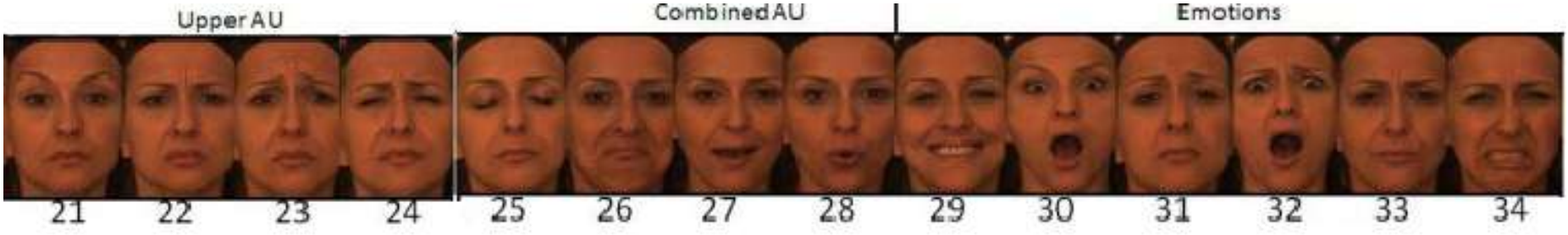
Expressions	Scan No	Explanation	v.2	v.1
Lower AUs	1	Lower Lip Depressor - AU16	•	
	2	Lips Part - AU25	•	
	3	Jaw Drop - AU26	•	
	4	Mouth Stretch - AU27	•	•
	5	Lip Corner Puller - AU12	•	•
	6	Left Lip Corner Puller - AU12L	•	
	7	Right Lip Corner Puller - AU12R	•	
	8	Low Intensity Lip Corner Puller - AU12LW	•	
	9	Dimpler - AU14	•	
	10	Lip Stretcher - AU20	•	
	11	Lip Corner Depressor - AU15	•	
	12	Chin Raiser - AU17	•	
	13	Lip Funneler - AU22	•	
	14	Lip Puckerer - AU18	•	
	15	Lip Tightener - AU23	•	
	16	Lip Presser - AU24	•	
	17	Lip Suck - AU28	•	•
	18	Upper Lip Raiser - AU10	•	
	19	Nose Wrinkler - AU9	•	•
	20	Cheek Puff - AU34	•	•



Facial Action Units

Upper/Combined Action Units + Basic Expressions

Expressions	Scan No	Explanation	v.2	v.1
Upper AUs	21	Outer Brow Raiser - AU2	•	•
	22	Brow Lowerer - AU4	•	•
	23	Inner Brow Raiser - AU1	•	
	24	Squint - AU44	•	
	25	Eyes Closed - AU43	•	•
Combined AUs	26	Jaw Drop (26) + Low Intensity Lip Corner Puller	•	
	27	Lip Funneler (22) + Lips Part (25)	•	•
	28	Lip Corner Puller (12) + Lip Corner Depressor (15)	•	
Emotions	29	Happiness	•	•
	30	Surprise	•	
	31	Fear	•	
	32	Sadness	•	
	33	Anger	•	
	34	Disgust	•	



Video-based expression recognition

Fake and real smiles

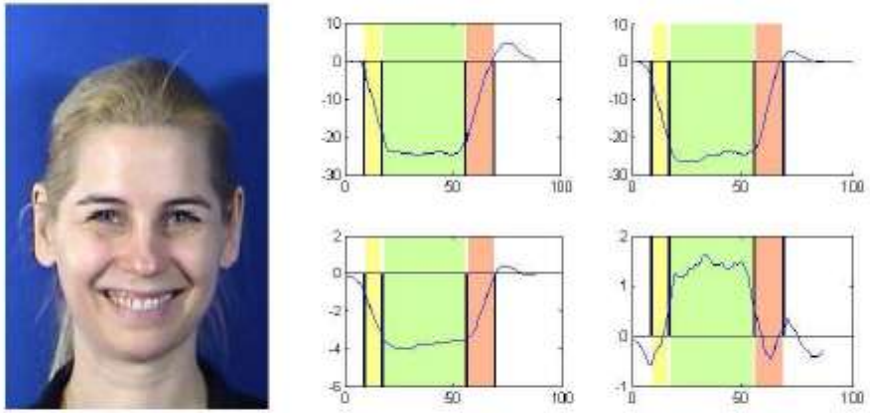


Figure 4: Temporal analysis of Action Unit 12 (smile). Shown are the four most informative features for AU12. The yellow shaded area depicts the period manually labeled as onset, the green area depicts the apex phase and the orange area depicts the offset phase.

Valstar, Pantic, Fully Automatic Facial Action Unit Detection and Temporal Analysis

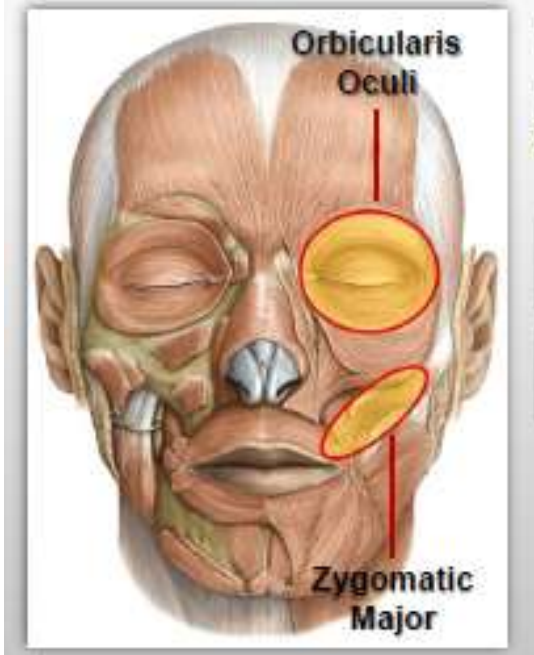
Video-based expression recognition

Real vs. fake smile detection

Dibeklioğlu, Valenti, Salah, Gevers, ACM Multimedia, 2010

Video-based expression recognition

Real vs. fake smile detection



✓ Zygomatic major raises the corners of the mouth.

✓ Orbicularis oculi muscle raises the cheeks and forms crows-feet around the eyes.

Zygomatic Major + Orbicularis Oculi → Spontaneous Smile

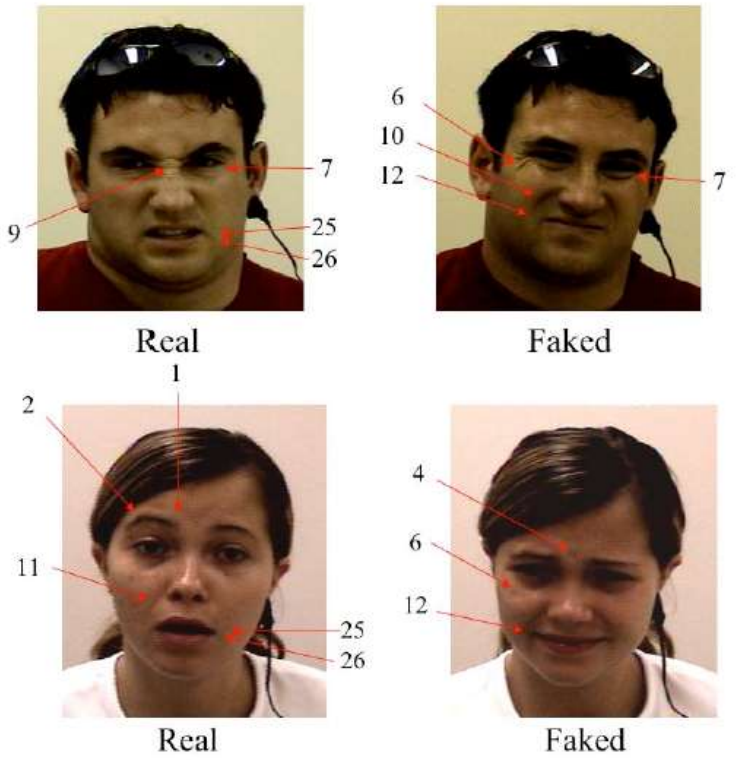
Zygomatic Major → Posed Smile

✓ It is also reported that asymmetry and timing of smiles are discriminative to classify different types of smiles.

P. Ekman and W. V. Friesen. Felt, false, and miserable smiles. Journal of Nonverbal Behavior, 6:238–252, 1982.

Video-based expression recognition

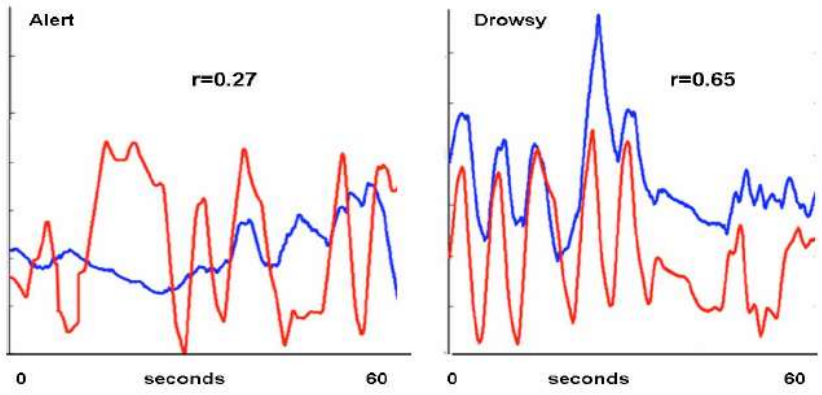
Detecting real vs. faked pain



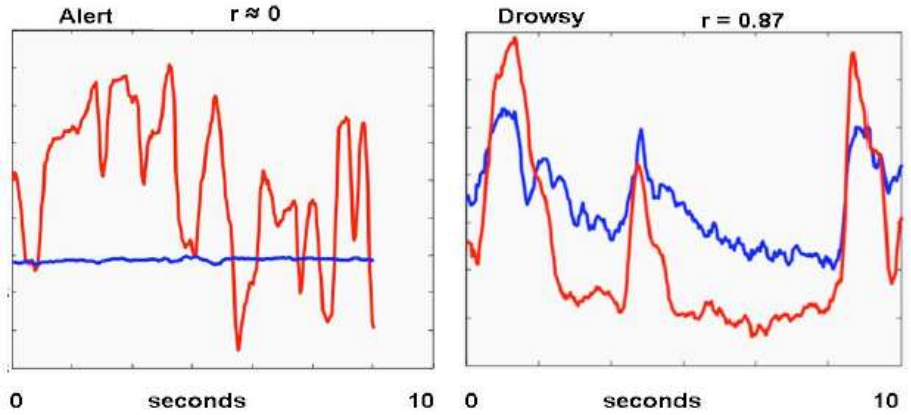
Bartlett et al., 2008

Video-based expression recognition

Drowsiness detection



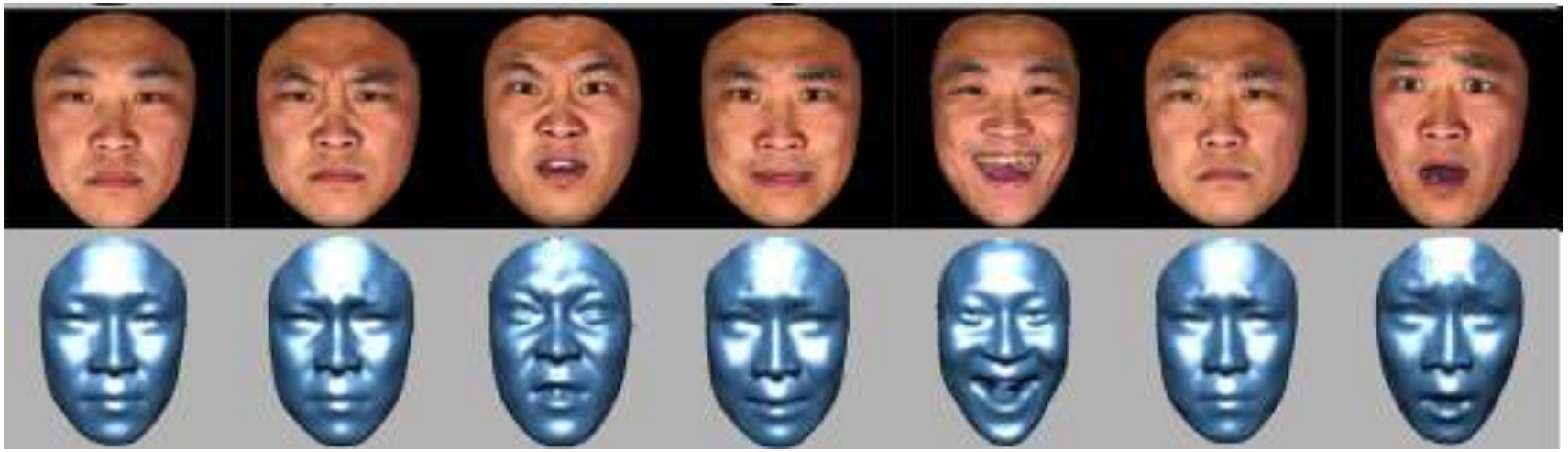
Head motion (blue) and steering position (red)



Eye Brow Raises AU12 (blue) and eye openness (red)

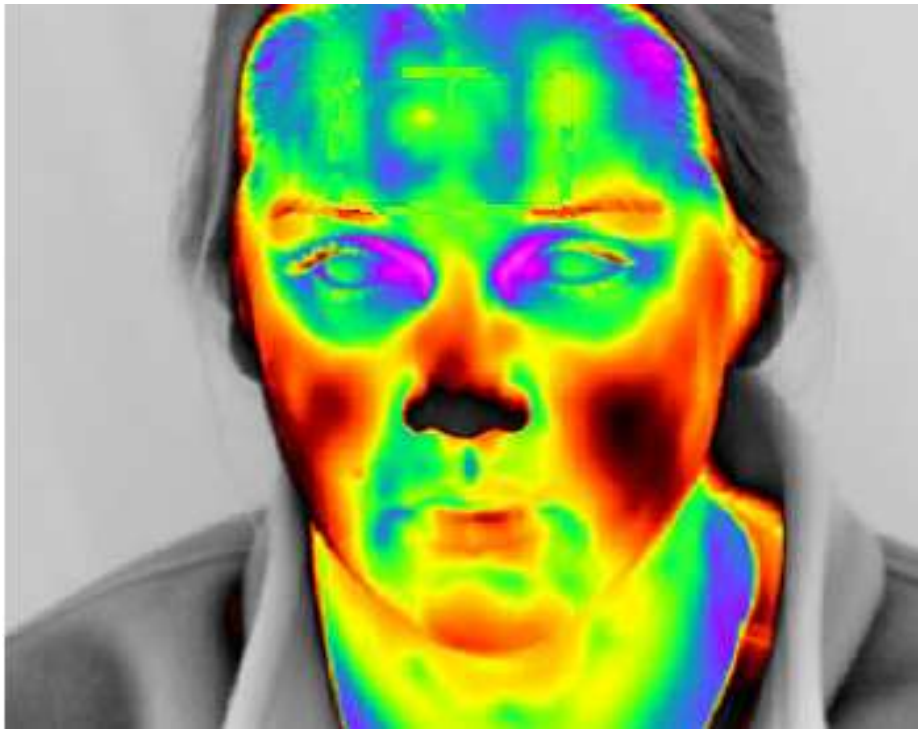
Bartlett et al., 2008

Expression from 3D



Yin et al., AFGR 2006

Other modalities



Thermal imaging of the face can reveal stress (Pavlidis, 2007)

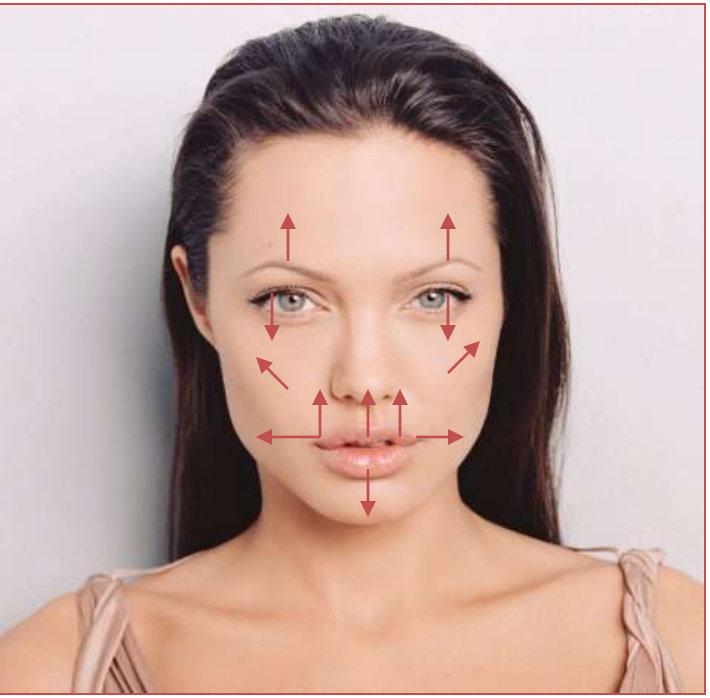
Speech

Emotion and speech parameters

Table 6. Emotions and Speech Parameters (from Murray and Arnott, 1993).					
	Anger	**Happiness**	**Sadness**	**Fear**	**Disgust**
Rate	Slightly faster	Faster or slower	Slightly slower	Much faster	Very much faster
Pitch Average	Very much higher	Much higher	Slightly lower	Very much higher	Very much lower
Pitch Range	Much wider	Much wider	Slightly narrower	Much wider	Slightly wider
Intensity	Higher	Higher	Lower	Normal	Lower
Voice Quality	Breathy, chest	Breathy, blaring tone	Resonant	Irregular voicing	Grumble chest tone
Pitch Changes	Abrupt on stressed	Smooth, upward inflections	Downward inflections	Normal	Wide, downward terminal inflects
Articulation	Tense	Normal	Slurring	Precise	Normal

Affective Computing

Facial Expression Recognition



12 facial motion measurements

vertical movement of the lips

horizontal movement of the mouth corners

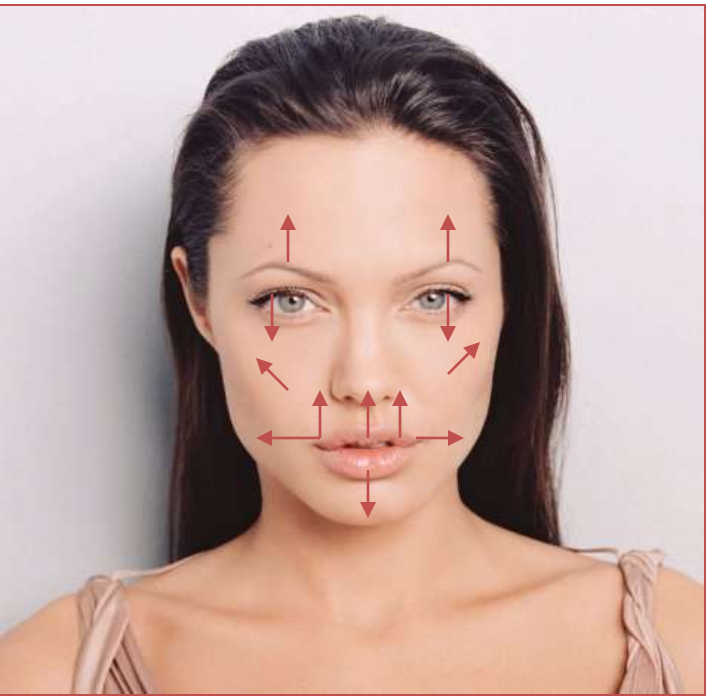
vertical movement of the mouth corners

vertical movement of the eye brows

lifting of the cheeks

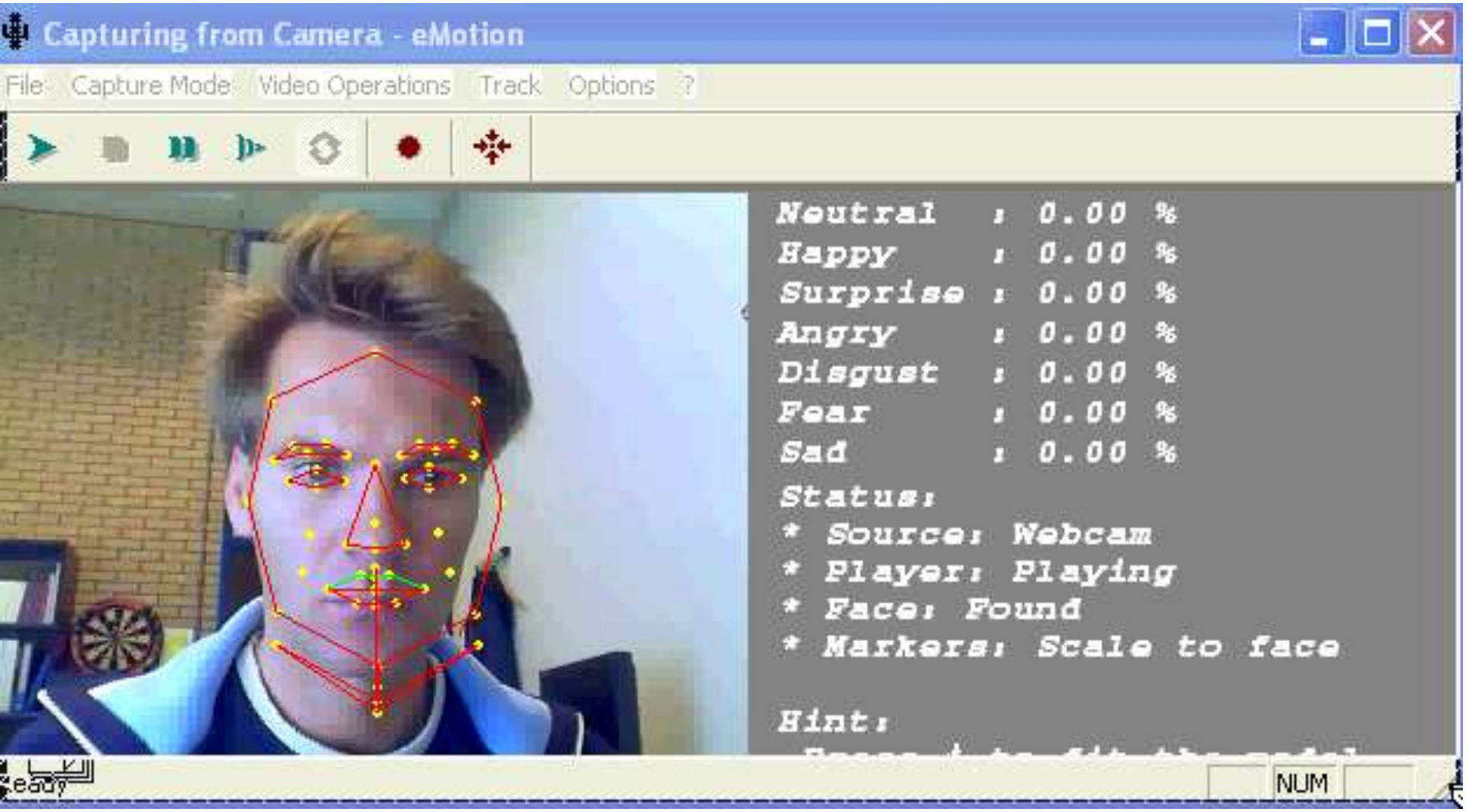
blinking of the eyes

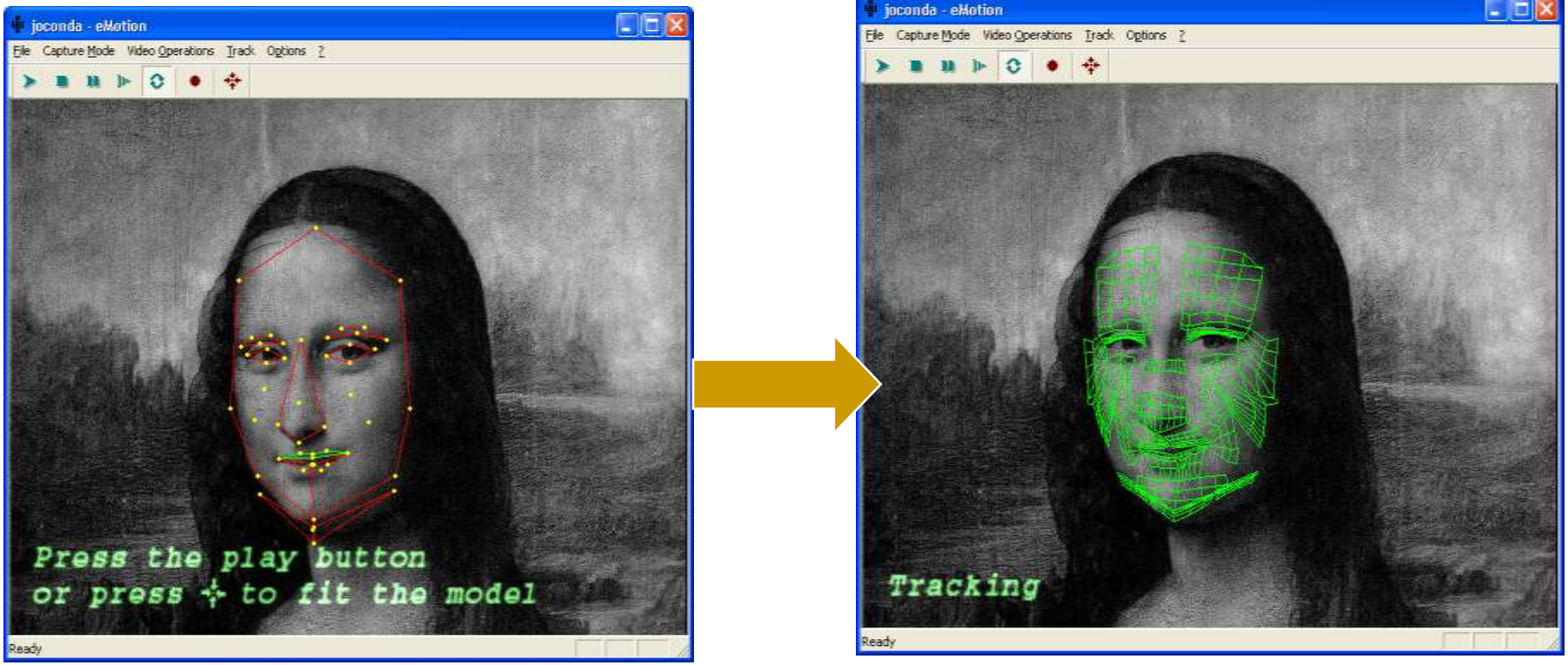
Facial Expression Recognition

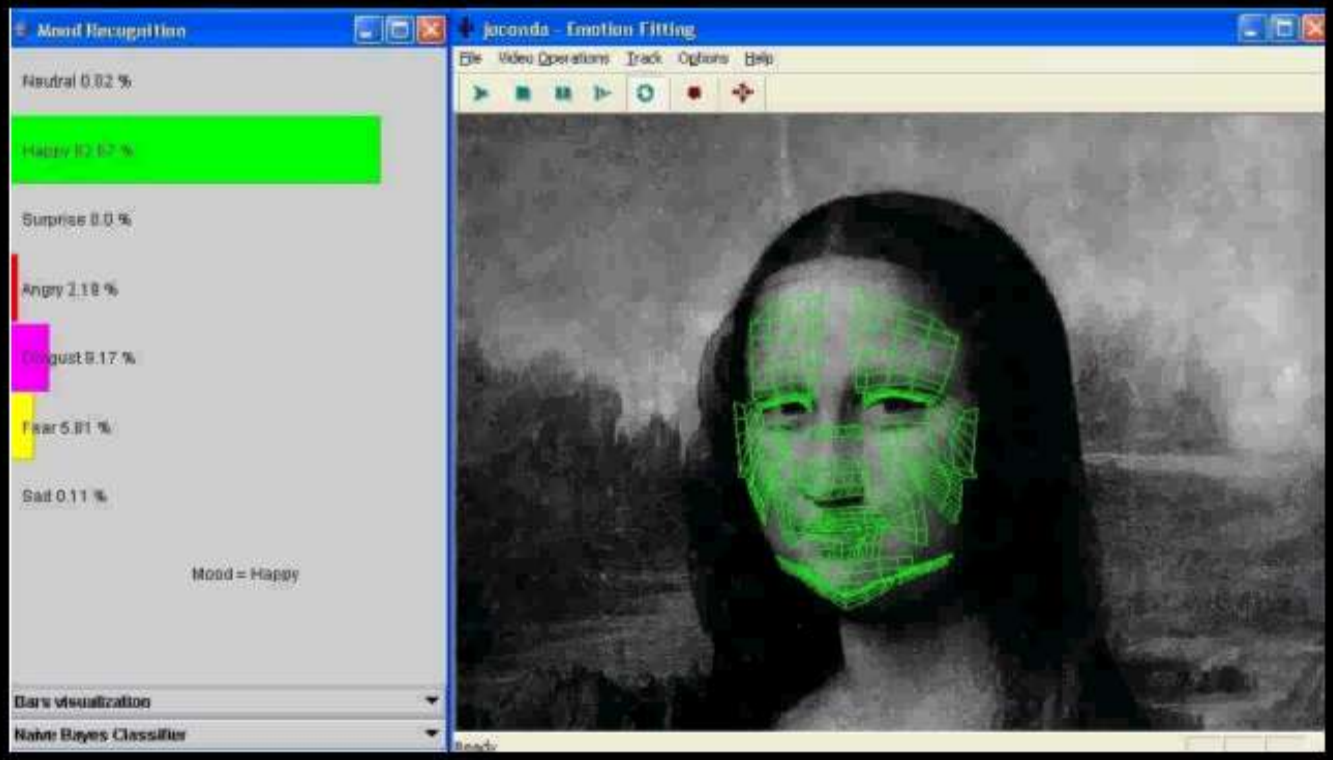


We use 12 facial features = 12 facial motion measurements

The combination of these features define the 7 basic classes of facial expression we want to classify: ***Neutral, Happy, Anger, Disgust, Fear, Sad, Surprise***


Capturing from Camera - eMotion
File
Capture Mode
Video Operations
Track
Options
?
Neutral : 0.00 %
Happy : 0.00 %
Surprise : 0.00 %
Angry : 0.00 %
Disgust : 0.00 %
Fear : 0.00 %
Sad : 0.00 %
Status:
* Source: Webcam
* Player: Playing
* Face: Found
* Markers: Scale to face
Hint:
NUM


joconda - eMotion
File Capture Mode Video Operations Track Options ?
Press the play button
or press to fit the model
Ready
joconda - eMotion
File Capture Mode Video Operations Track Options ?
Tracking
Ready


Mood Recognition
Neutral 0.02 %
Surprise 0.0 %
Angry 2.18 %
Fear 5.81 %
Sad 0.11 %
Mood = Happy
Bars visualization
Naive Bayes Classifier
joconda - Emotion Fitting
File
Video Operations
Track
Options
Help

Glad or Sad

[Emoties aantoonbaar gemaakt]

- Home
- Upload foto
- Stem op foto
- Foto galerij
- Over
- Contact

Glad or Sad is een samenwerking tussen

ilse media

Universiteit van Amsterdam

Stap 1

Upload een foto.

Stap 2

Laat zoveel mogelijk mensen stemmen op jouw foto.

Stem

Aantal stemmen: 15%

Stap 3

Bekijk de resultaten in de foto galerij.

Analyse resultaten

Vrolijk	91%
Verrast	8%
Boos	0%
Walging	0%
Angstig	1%
Droevig	0%

Verfijnen / Sorteren

Selecteer een categorie

- Politici
- Sporters
- Schoonmoeders
- Overige

Sorteer op

- Datum
- Meest vrolijk
- Meest verrast
- Meest boos
- Meeste walging
- Meest angstig
- Meest droevig

Selecteer een tag

Acda Balkenende Bangebroek Duh Gezicht Glimlach Huilen Lachen Man Smile Voetbal Vrouw

Change language to

[English]

Pong

Emotion driven table tennis

TECHNOLOGIE

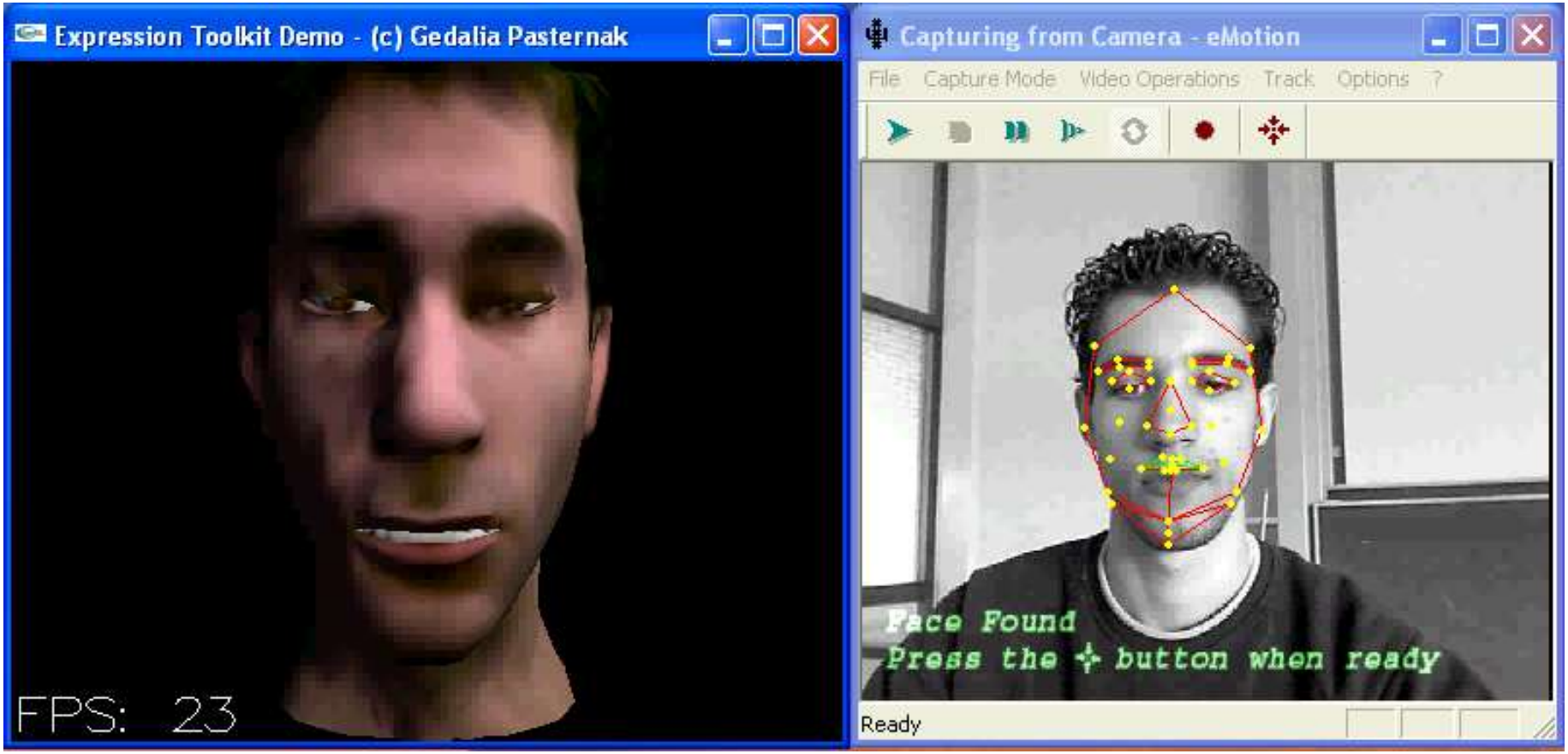
Pong nieuwe stijl

In 2007 was het wereldnieuws: het gezicht van Mona Lisa bestond voor 83 procent uit blijdschap, 9 procent walging, 6 procent angst en 2 procent boosheid. Met de software die werd gebruikt om de gemoedstoestand van Mona Lisa te ontleden, kun je Pong spelen, het oude computerspelletje met de 2 batjes en het balletje. Alleen speel je het spel nu niet met een joystick, maar met je gezicht.

De bedenker ervan is Theo Gevers van het Instituut voor Informatica van de Universiteit van Amsterdam. Gevers: 'De 6 basisemoties (vreugde, verdriet, angst, woede, verbazing, afschuw) van de mens zitten in de hersenen in het limbisch systeem. Overal ter wereld, bij elk mens, zijn deze 6 emoties hetzelfde. Bij een bepaalde emotionele gebeurtenis komen stoffen vrij die de gezichtsspieren aansturen. We hebben gekeken naar welke spieren corresponderen met welke emotie.'
De software ziet welke gezichtspieren worden gebruikt en welke basisemotie daarbij hoort. Bijvoorbeeld: kijk je blij, dan gaat het batje omhoog, kijk je triest dan gaat het batje omlaag. De enige benodigdheden zijn een webcam en de speciale versie van het spel.

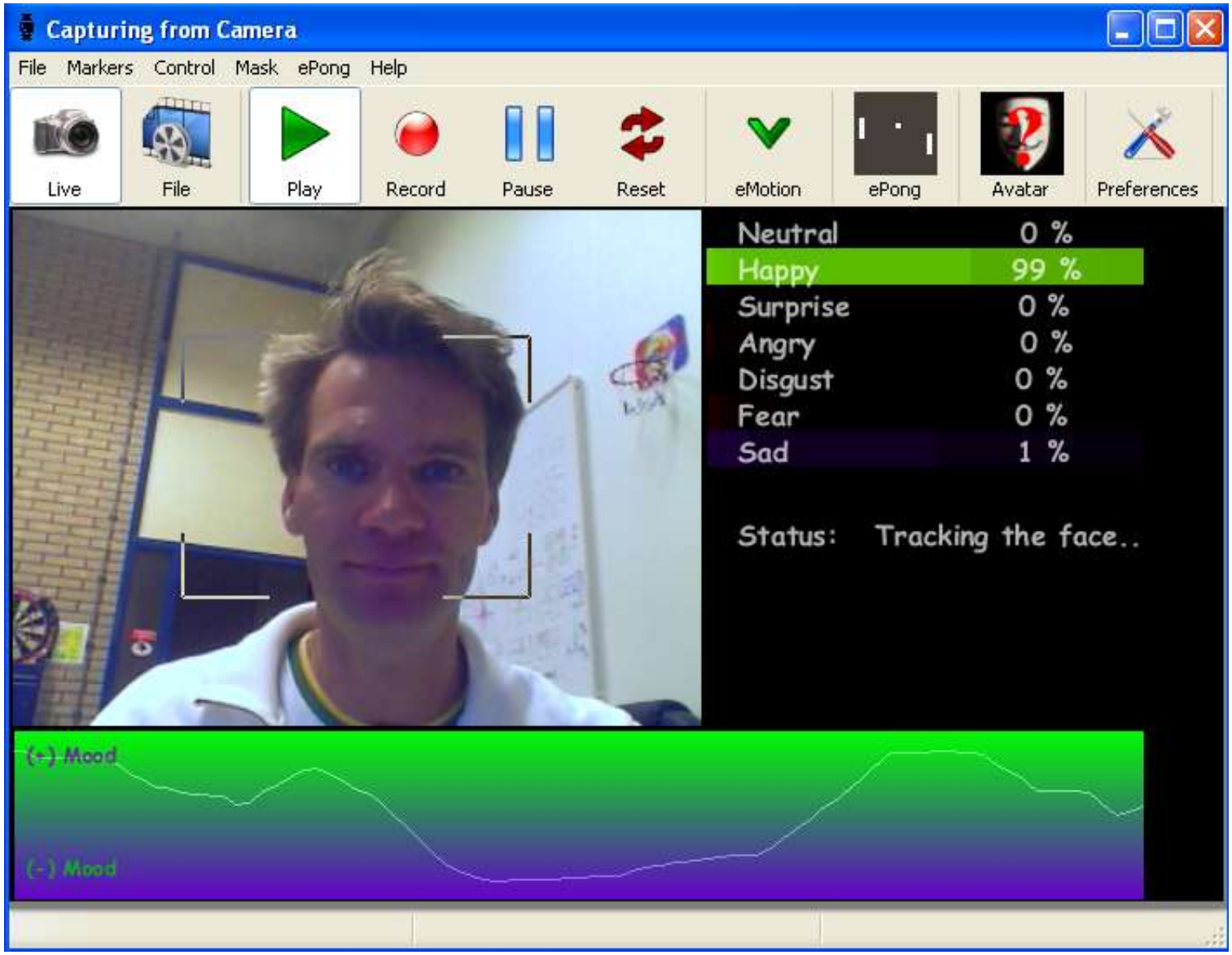
Vision-based Interaction (Games)

Avatar

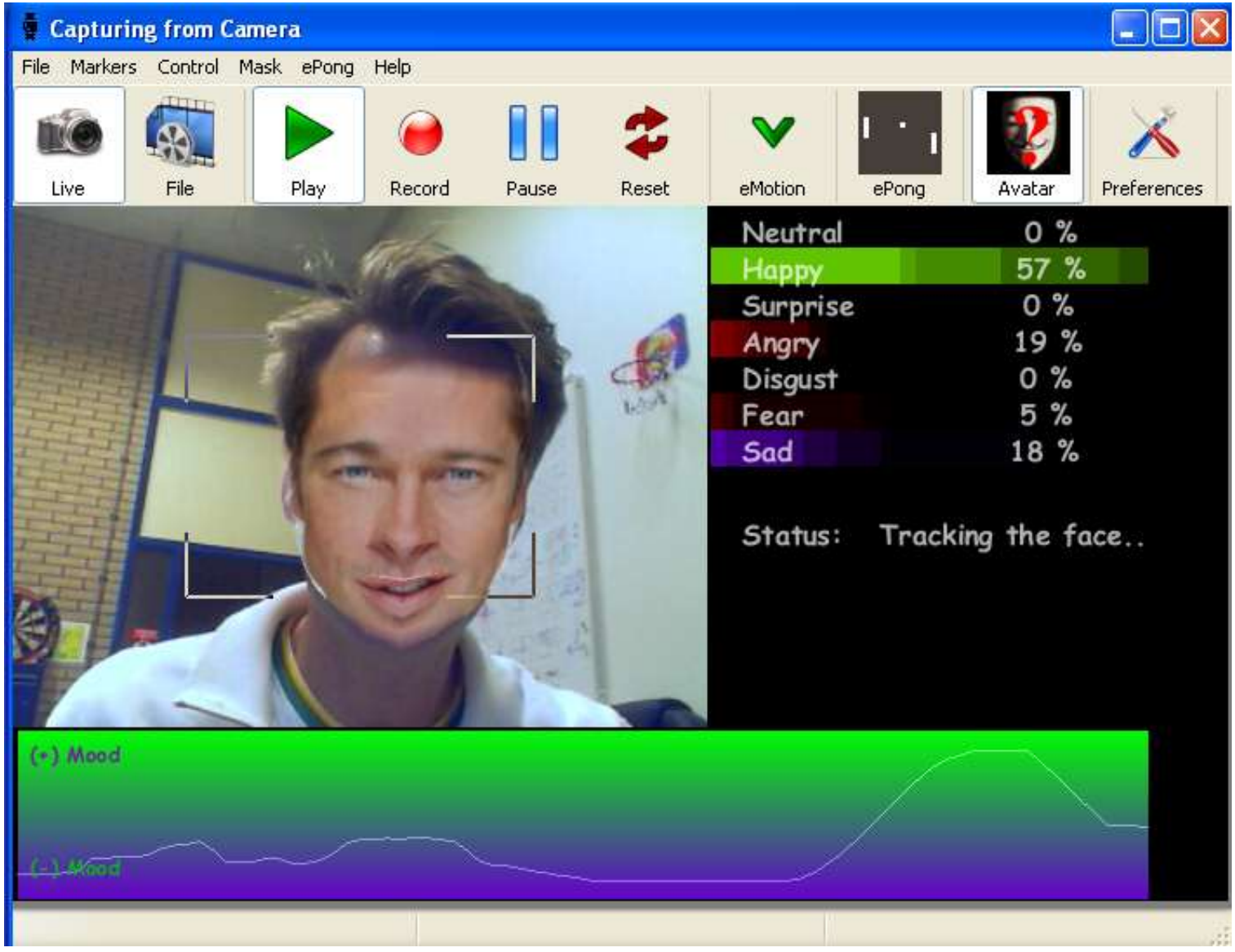


Face off

The mask


Capturing from Camera
File Markers Control Mask ePong Help
Live
File
Play
Record
Pause
Reset
eMotion
ePong
Avatar
Preferences
Neutral 0 %
Happy 99 %
Surprise 0 %
Angry 0 %
Disgust 0 %
Fear 0 %
Sad 1 %
Status: Tracking the face..
(+) Mood
(-) Mood

The mask


Capturing from Camera
File Markers Control Mask ePong Help
Live
File
Play
Record
Pause
Reset
eMotion
ePong
Avatar
Preferences
Neutral 0 %
Happy 57 %
Surprise 0 %
Angry 19 %
Disgust 0 %
Fear 5 %
Sad 18 %
Status: Tracking the face..
(+) Mood
(-) Mood

The mask


Capturing from Camera
File
Markers
Control
Mask
ePong
Help
Live
File
Play
Record
Pause
Reset
eMotion
ePong
Avatar
Preferences
Neutral 0 %
Happy 86 %
Surprise 0 %
Angry 6 %
Disgust 0 %
Fear 1 %
Sad 7 %
Status: Tracking the face..
(+) Mood
(-) Mood

[Valenti, cvpr08]

Human Behaviour Understanding

- Facial expression
- Head pose
- Eye Tracking
- Voice

Smart cars

- Mobileye
  - Vision systems currently in high-end BMW, GM, Volvo models
  - By 2010: 70% of car manufacturers.

Slide content courtesy of Amnon Shashua

Google cars

http://www.nytimes.com/2010/10/10/science/10google.html?ref=artificialintelligence

Mobile robots

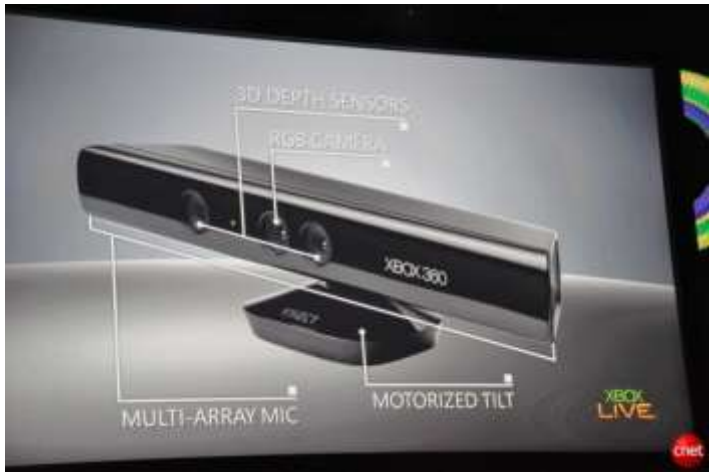
NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover

http://www.robocup2013.org/

Saxena et al. 2008
STAIR at Stanford

Interactive Games: Kinect

- Object Recognition: http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o
- Mario: http://www.youtube.com/watch?v=8CTJL5lUjHg
- 3D: http://www.youtube.com/watch?v=7QrnwoO1-8A
- Robot: http://www.youtube.com/watch?v=w8BmgtMKFbY
- Student of the Master AI (Daniel Karavolos, Sicco van Sas, Maarten van der Velden) Superman: http://www.youtube.com/watch?v=L6C01tyrhf0



Vision in space

NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

Vision systems (JPL) used for several tasks

- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

Overview

- 29-10-2012, Monday, 15:00-17:00, Science Park A1.04 - Introduction
- 05-11-2011, Monday, 15:00-17:00, Science Park A1.04 - Image and Video Formation
- 12-11-2011, Monday, 15:00-17:00, Science Park A1.04 - Color Invariance and Image Processing
- 19-11-2011, Monday, 15:00-17:00, Science Park A1.04 - Feature Extraction and Object Recognition
- 26-11-2011, Monday, 15:00-17:00, Science Park A1.04 - Learning and Tracking
- 03-12-2011, Monday, 15:00-17:00, Science Park A1.04 - Visual Attention and Affective Computing
- 10-12-2011, Monday, 15:00-17:00, Science Park A1.04 - Human Behavior Analysis
- 18-12-2011, Tuesday, 15:00-18:00, Science Park, C1.10 - Examination