Intelligent Multimedia Systems

Master AI, 2012, Lecture 7

Lecturers: Theo Gevers
Lab: Intelligent Systems Lab Amsterdam (ISLA)
Email: th.gevers@uva.nl
http://staff.science.uva.nl/~gevers

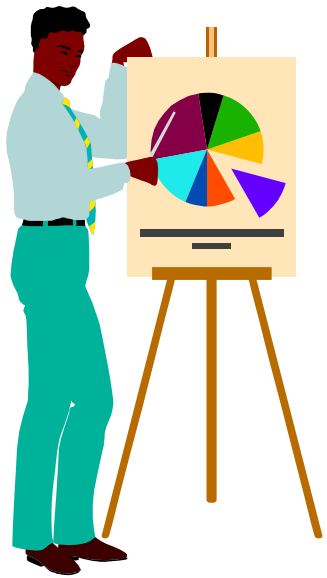
Lectures

- 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 Tracking
- 26-11-2011, Monday, 15:00-17:00, Science Park A1.04 - Learning and Object Recognition
- 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

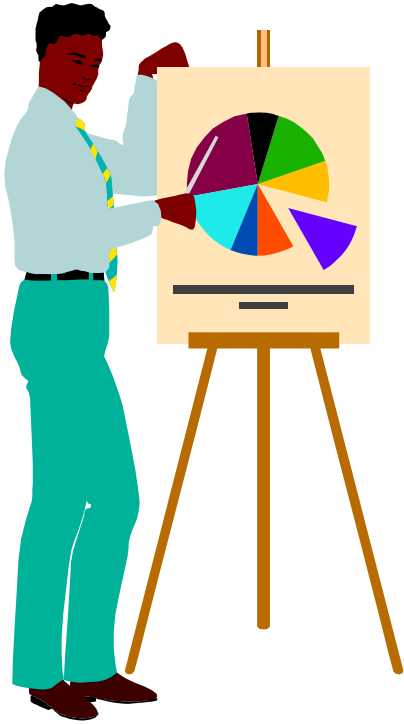
Today's class

Part I: Human Behavior Analysis

Part II: Summary of the Lectures



Today’s class



PART I

Activity Recognition

Social Signal Processing

Face and Facial Expression Recognition

Gaze Estimation

Activity Recognition

With slides by
Derek Hoiem and
Kristen Grauman

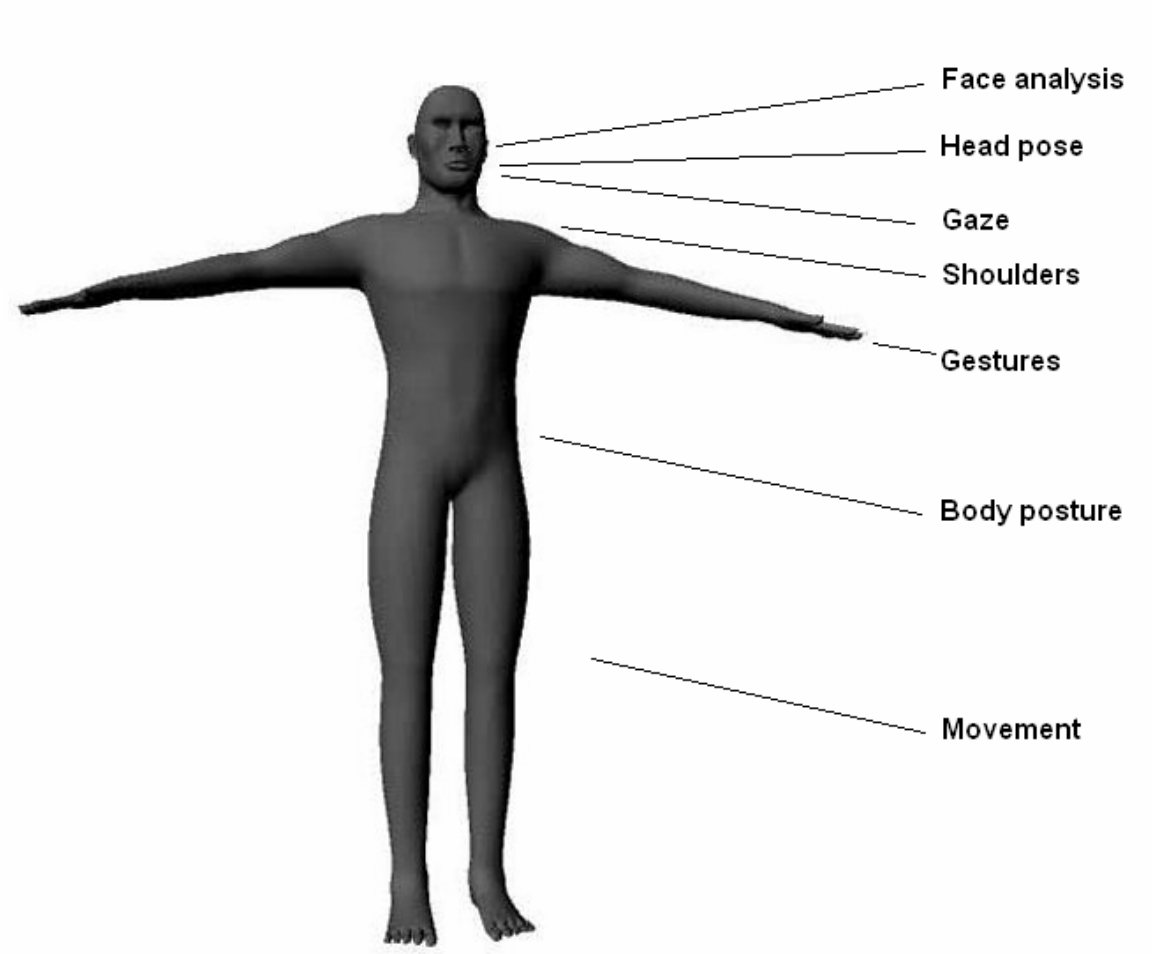
Applications

- Alarm and action detection
  - Fall detection
  - Emergencies
- Activity recognition
  - Activities of daily living
  - Aggression detection
  - Social relations
  - Navigation of people with dementia
- Biometry and therapy
  - Stability and walking
  - Physical therapy

Kinect

Activity Recognition

Visual analysis of the human body



What is an action?

Action: a transition from one state to another

- Who is the actor?
- How is the state of the actor changing?
- What (if anything) is being acted on?
- How is that thing changing?
- What is the purpose of the action (if any)?

Human activity in video

No universal terminology, but approximately:

- "**Actions**": atomic motion patterns -- often gesture-like, single clear-cut trajectory, single nameable behavior (e.g., sit, wave arms)

- "**Activity**": series or composition of actions (e.g., interactions between people)

- "**Event**": combination of activities or actions (e.g., a football game, a traffic accident)

Adapted from Venu Govindaraju

How do we represent actions?

Categories

Walking, hammering, dancing, skiing, sitting down, standing up, jumping

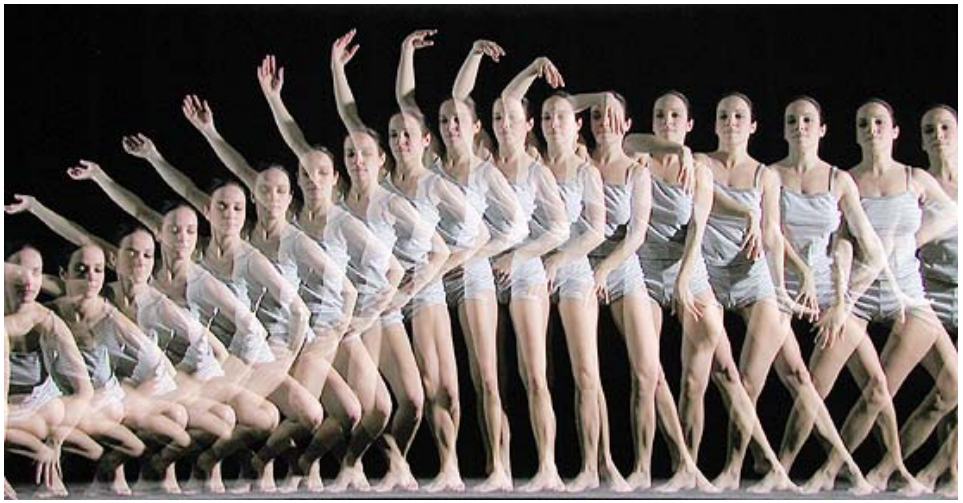
Poses

Nouns and Predicates

<man, swings, hammer>
<man, hits, nail, w/ hammer>

How can we identify actions?

Motion

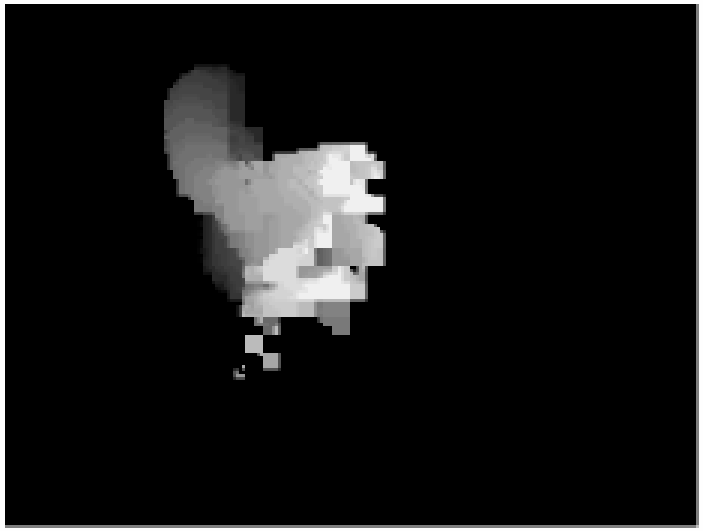


Pose

Representing Motion

Optical Flow with Motion History

sit-down

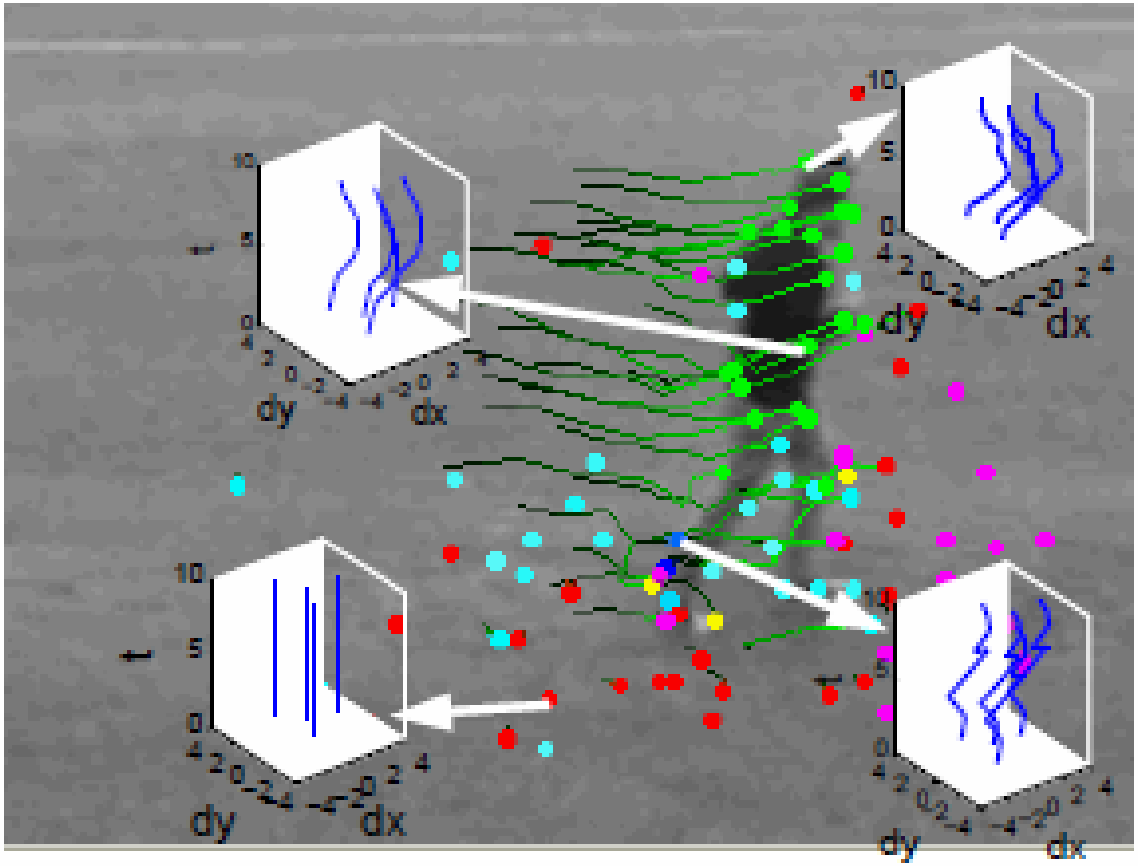


sit-down MHI

Bobick Davis 2001

Representing Motion

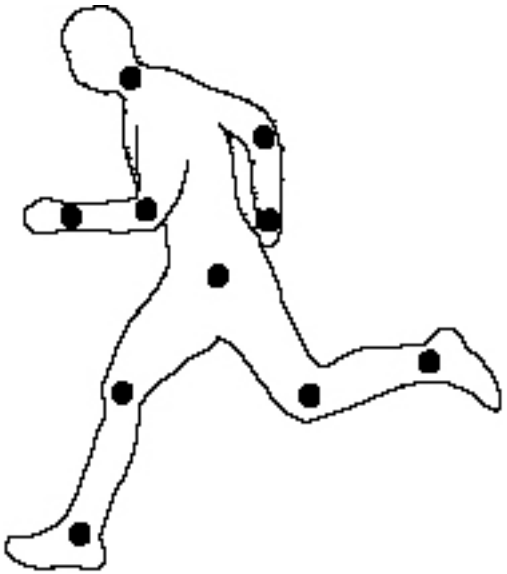
Tracked Points

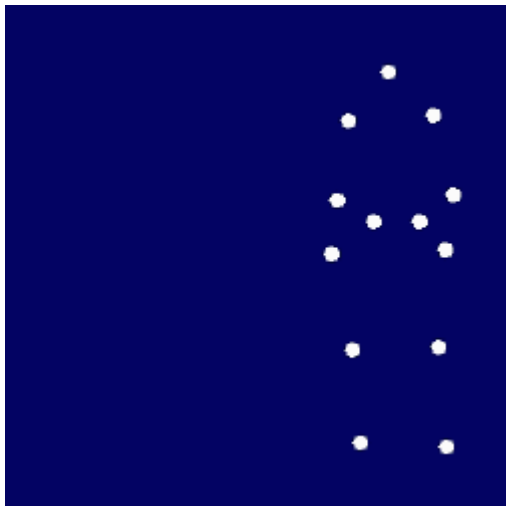


Matikainen et al. 2009

Activity Recognition – traditional

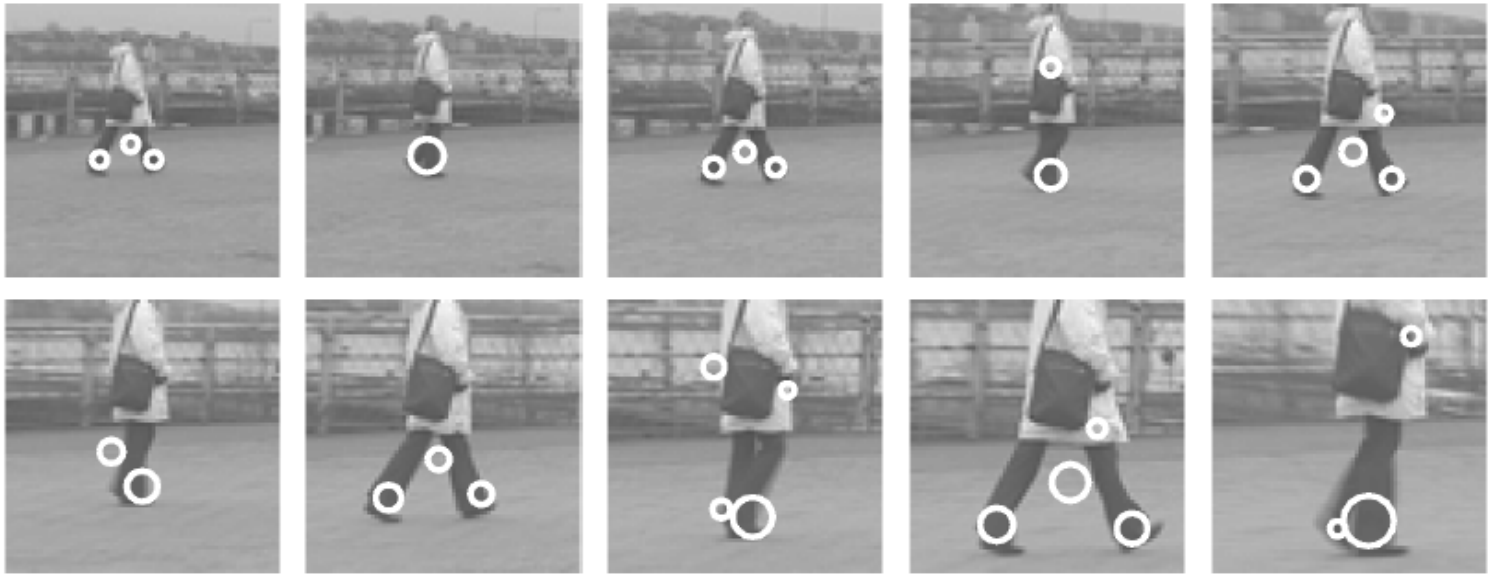
Moving Light Display (MLD) of Gunnar Johansson





Jastorff J., Giese M.A., http://www.compsens.uni-tuebingen.de/index.php?page=project&id=14

Spatio-temporal interest points



Spatial interest points

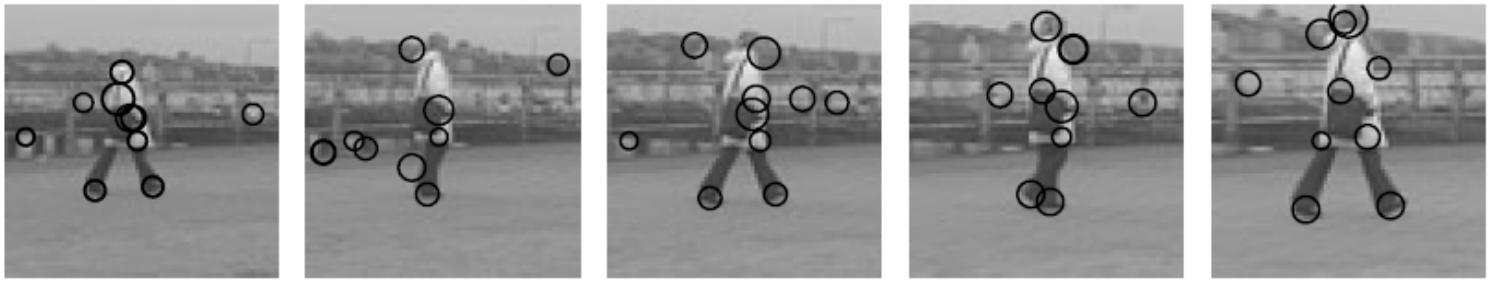


Figure 8: Top: Results of spatio-temporal interest point detection for a zoom-in sequence of a walking person. The spatial scale of the detected points (corresponding to the size of circles) matches the increasing spatial extent of the image structures and verifies the invariance of the interest points with respect to changes in spatial scale. Bottom: Pure spatial interest point detector (here, Harris-Laplace) selects both moving and stationary points and is less restrictive.

Activity Recognition

Point trajectories

Representing Motion

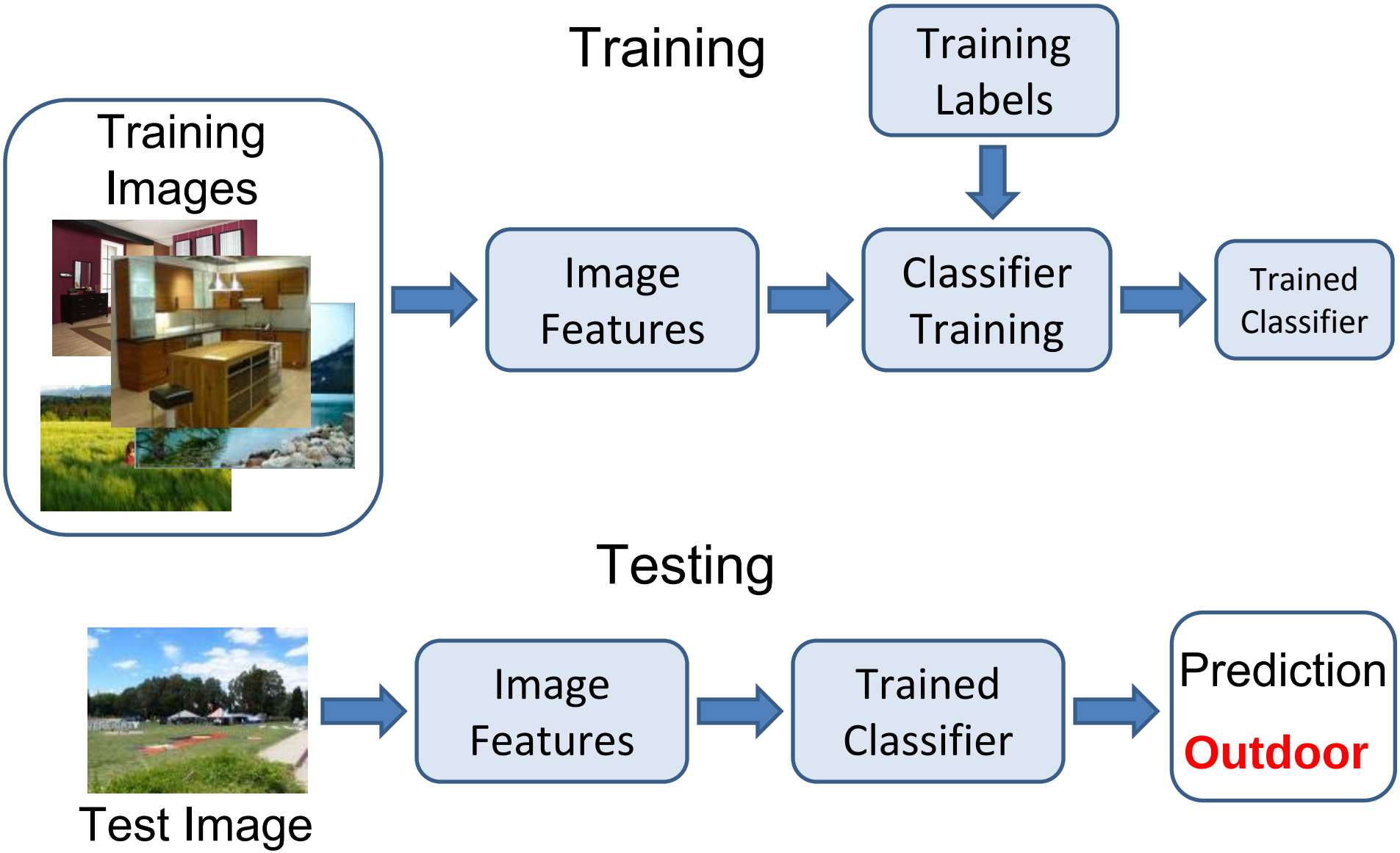
Space-Time Volumes

Blank et al. 2005

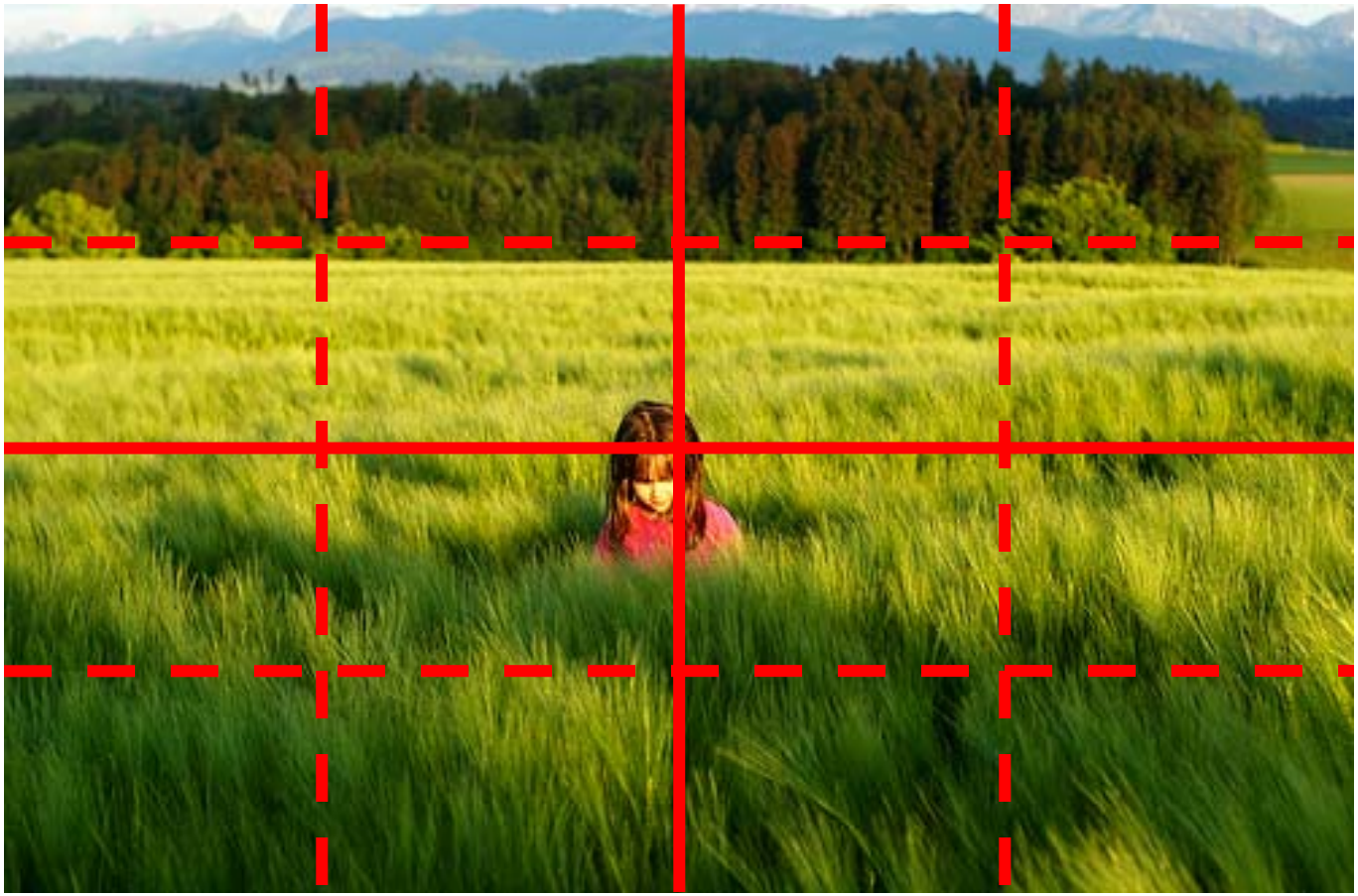
Action recognition as classification

Retrieving actions in movies, Laptev and Perez, 2007

Remember image categorization...



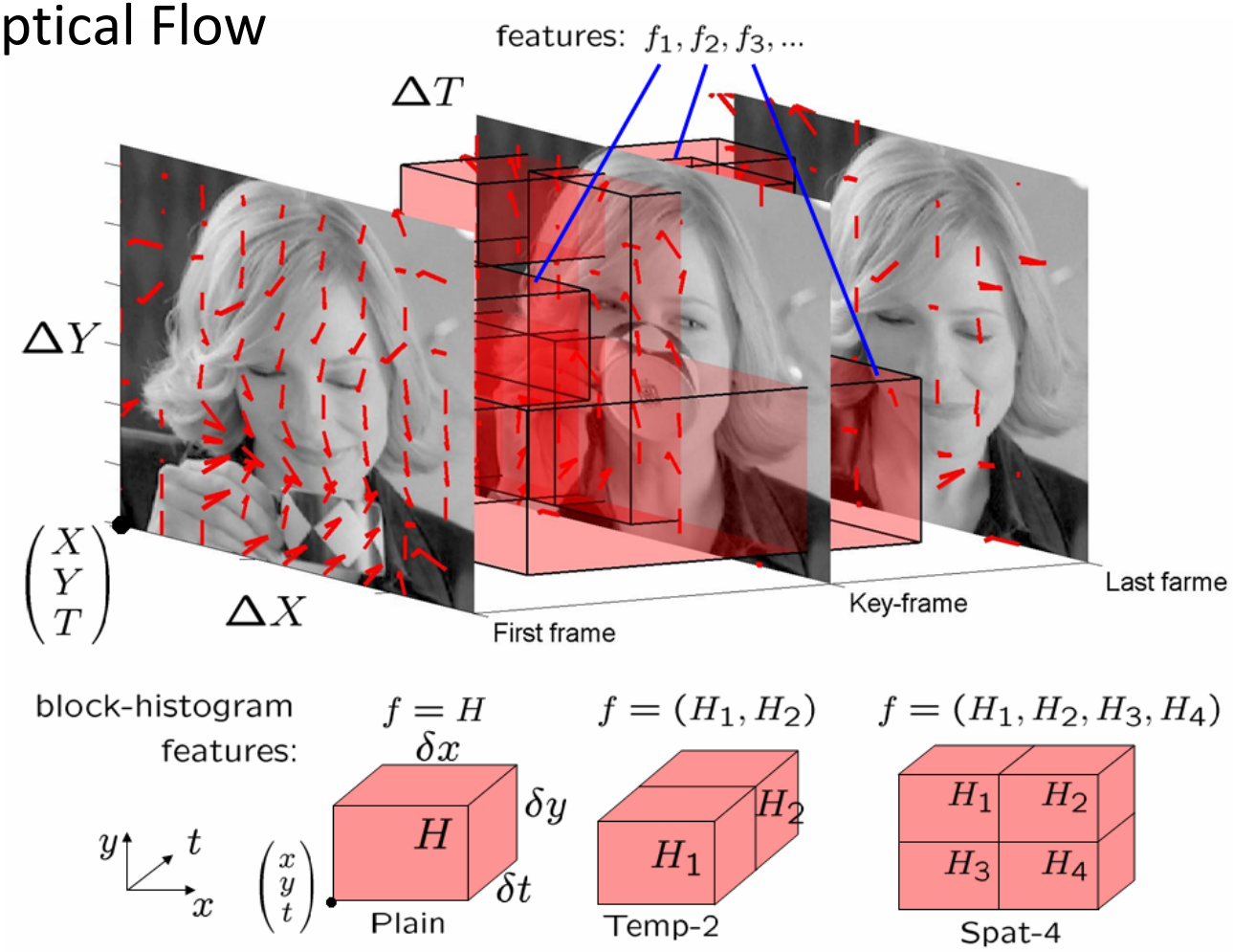
Remember spatial pyramids....



Compute histogram in each spatial bin

Features for Classifying Actions

1. Spatio-temporal pyramids (14x14x8 bins)
   - Image Gradients
   - Optical Flow



Results

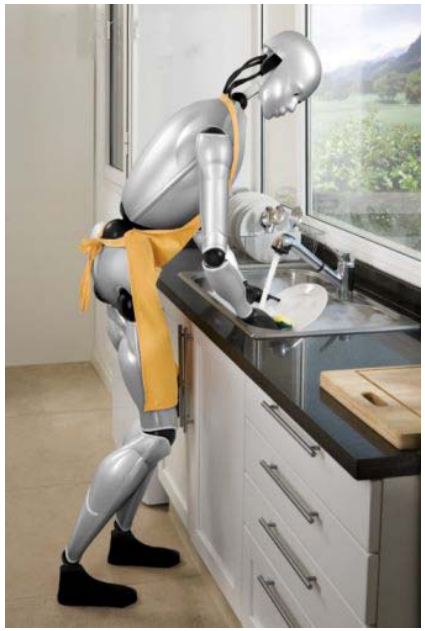
Task	HoG BoF	HoF BoF	Best channel	Best combination
KTH multi-class	81.6%	89.7%	91.1% (hof h3x1 t3)	91.8% (hof 1 t2, hog 1 t3)
Action AnswerPhone	13.4%	24.6%	26.7% (hof h3x1 t3)	32.1% (hof o2x2 t1, hof h3x1 t3)
Action GetOutCar	21.9%	14.9%	22.5% (hof o2x2 1)	41.5% (hof o2x2 t1, hog h3x1 t1)
Action HandShake	18.6%	12.1%	23.7% (hog h3x1 1)	32.3% (hog h3x1 t1, hog o2x2 t3)
Action HugPerson	29.1%	17.4%	34.9% (hog h3x1 t2)	40.6% (hog 1 t2, hog o2x2 t2, hog h3x1 t2)
Action Kiss	52.0%	36.5%	52.0% (hog 1 1)	53.3% (hog 1 t1, hof 1 t1, hof o2x2 t1)
Action SitDown	29.1%	20.7%	37.8% (hog 1 t2)	38.6% (hog 1 t2, hog 1 t3)
Action SitUp	6.5%	5.7%	15.2% (hog h3x1 t2)	18.2% (hog o2x2 t1, hog o2x2 t2, hog h3x1 t2)
Action StandUp	45.4%	40.0%	45.4% (hog 1 1)	50.5% (hog 1 t1, hof 1 t2)


“Talk on phone”


“Get out of car”

Learning realistic human actions from movies, Laptev et al. 2008

Action Recognition using Pose and Objects



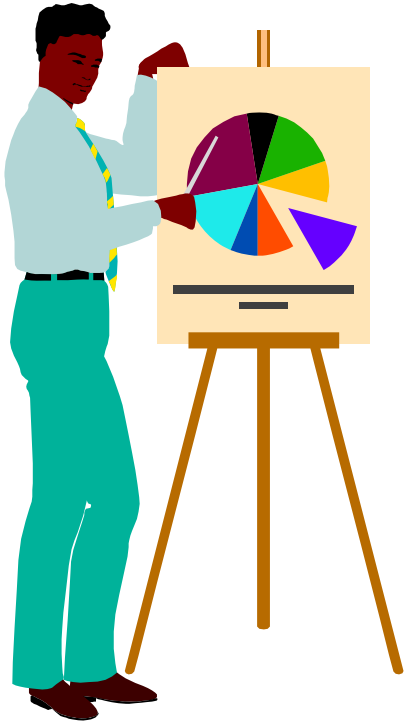
[Modeling Mutual Context of Object and Human Pose in Human-Object Interaction Activities](), B. Yao and Li Fei-Fei, 2010

Slide Credit: Yao/Fei-Fei

Summary

- Some work done, but it is just the beginning of exploring the problem. So far…
  - Actions are mainly simple and categorical
  - Most approaches are classification using simple features (spatial-temporal histograms of gradients or flow, s-t interest points, SIFT in images)
  - Just a couple works on how to incorporate pose and objects
  - Not much idea of how to reason about long-term activities or to describe video sequences

Today’s class



PART I

Activity Recognition

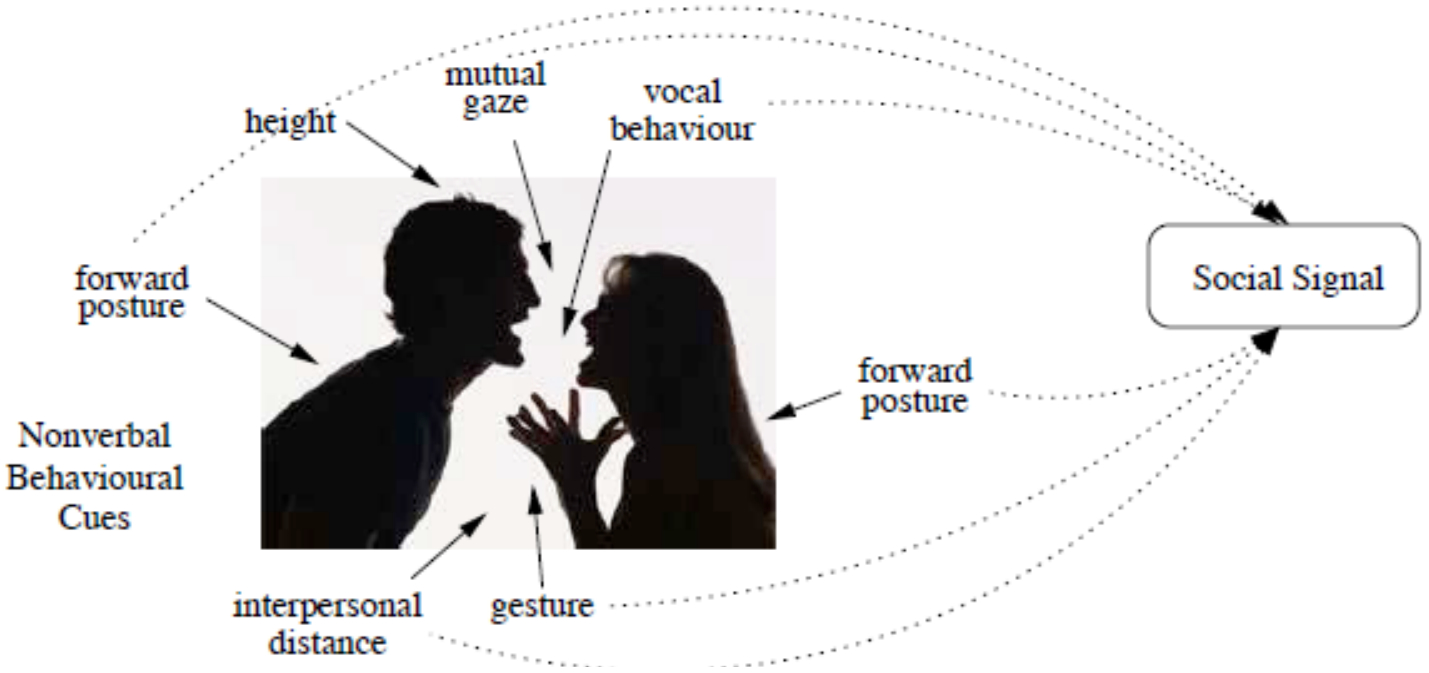
Social Signal Processing

Face and Facial Expression Recognition

Gaze Estimation

Social Signal Processing

Nonverbal cues



Vinciarelli, Pantic, Bourlard, 2009

Social Signal Processing

Example for posture congruence

Congruent postures

Non–congruent postures

Vinciarelli, Pantic, Bourlard, 2009

Social Signal Processing

Taxonomy

	Example Social Behaviours							Tech.		
Social Cues	emotion	personality	status	dominance	persuasion	regulation	rapport	speech anlysis	computer vision	biometry
Physical appearance										
height			√	√					√	√
attractiveness		√	√	√	√		√		√	√
body shape		√		√					√	√

Vinciarelli, Pantic, Bourlard, 2009

Social Signal Processing

Gesture and posture

	Example Social Behaviours							Tech.		
Social Cues	emotion	personality	status	dominance	persuasion	regulation	rapport	speech anlysis	computer vision	biometry
Gesture and posture										
hand gestures	√	√			√	√	√		√	√
posture	√	√	√	√	√	√	√		√	√
walking		√	√	√					√	√

Vinciarelli, Pantic, Bourlard, 2009

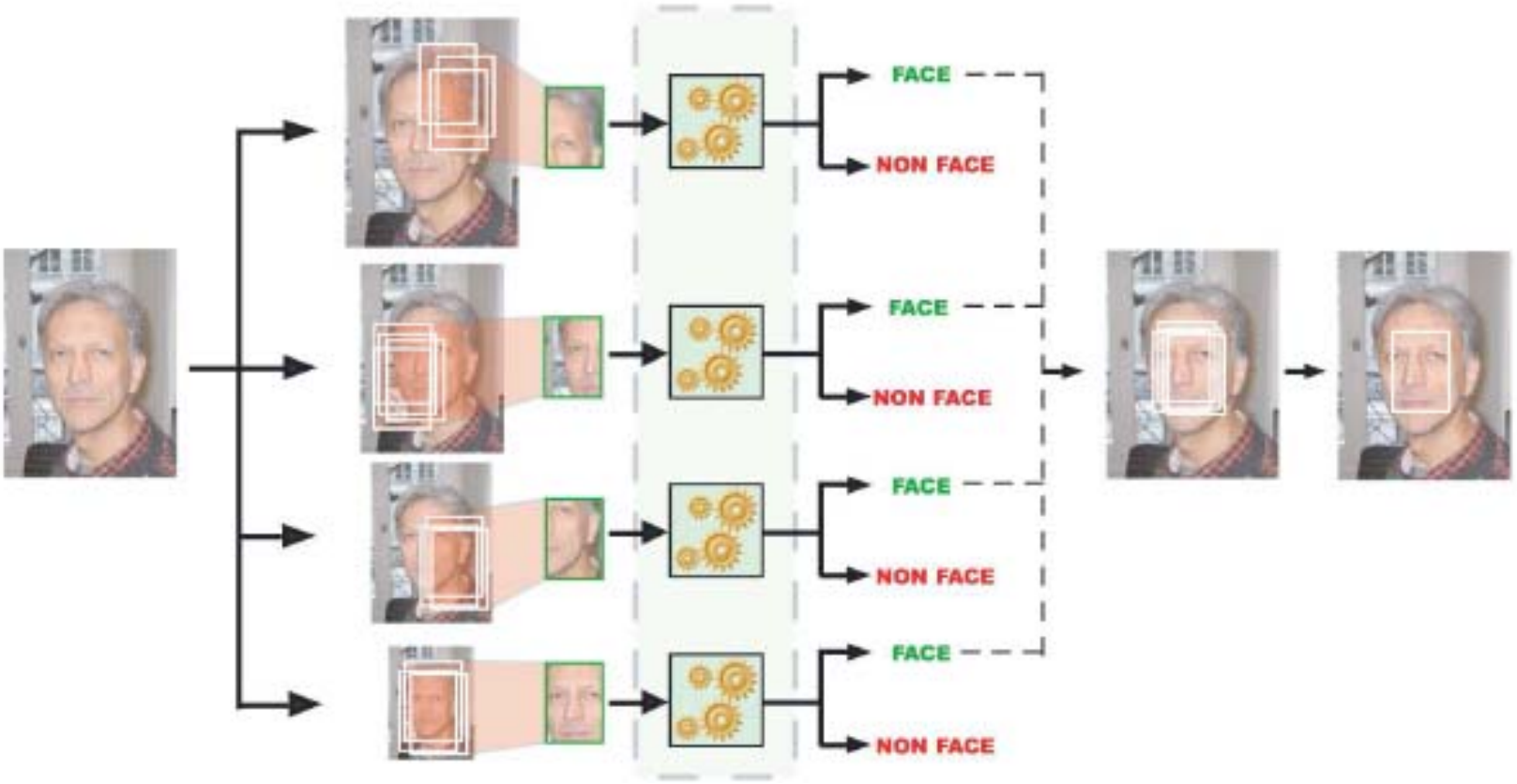
Social Signal Processing

Face and eyes

	Example Social Behaviours							Tech.		
Social Cues	emotion	personality	status	dominance	persuasion	regulation	rapport	speech anlysis	computer vision	biometry
Face and eyes behaviour										
facial expressions	√	√	√	√	√	√	√		√	√
gaze behaviour	√	√	√	√	√	√	√		√	
focus of attention	√	√	√	√	√	√	√		√	

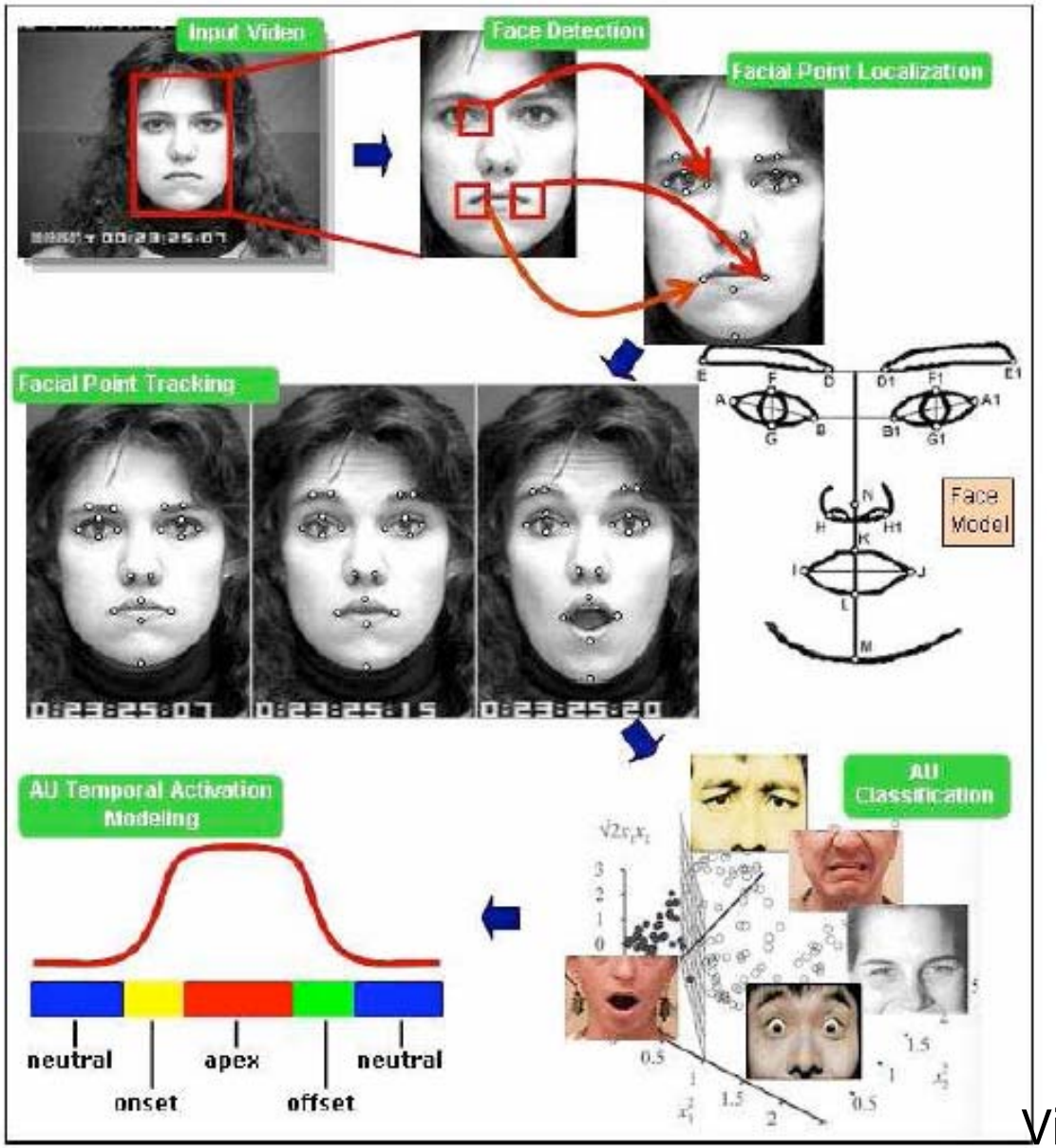
Vinciarelli, Pantic, Bourlard, 2009

Faces



Vinciarelli, Pantic, Bourlard, 2009

Facial expression



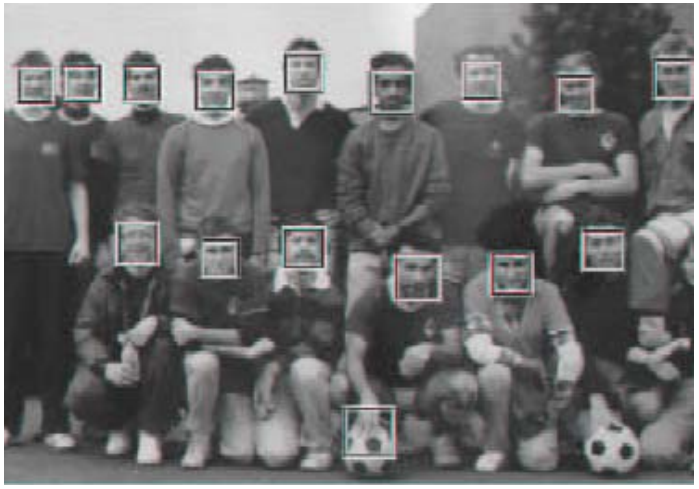
Vinciarelli, Pantic, Bourlard, 2009

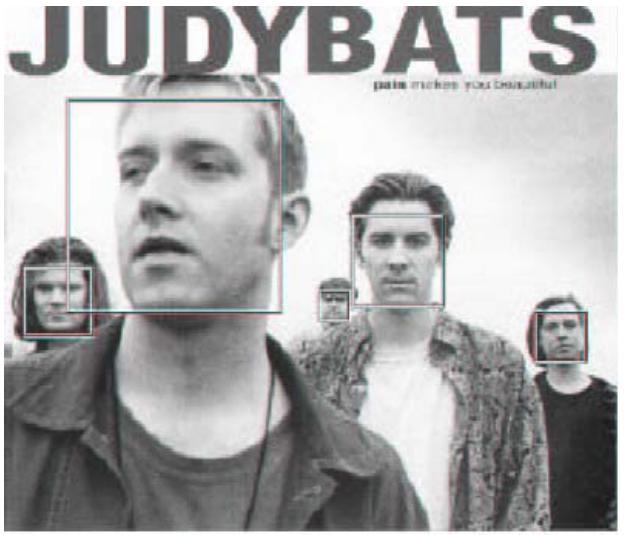
Face Detection: Viola-Jones Algorithm

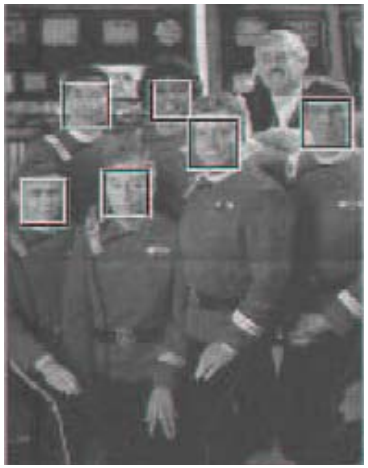
- Cascaded classifiers of simple features

- First level: Two features, 60% of false positives rejected
- Second level: Five features, 80% of false positives rejected
- 32 levels in total
- Trained with 4900 faces, 10000 non-faces

Face Detection: Viola-Jones Algorithm





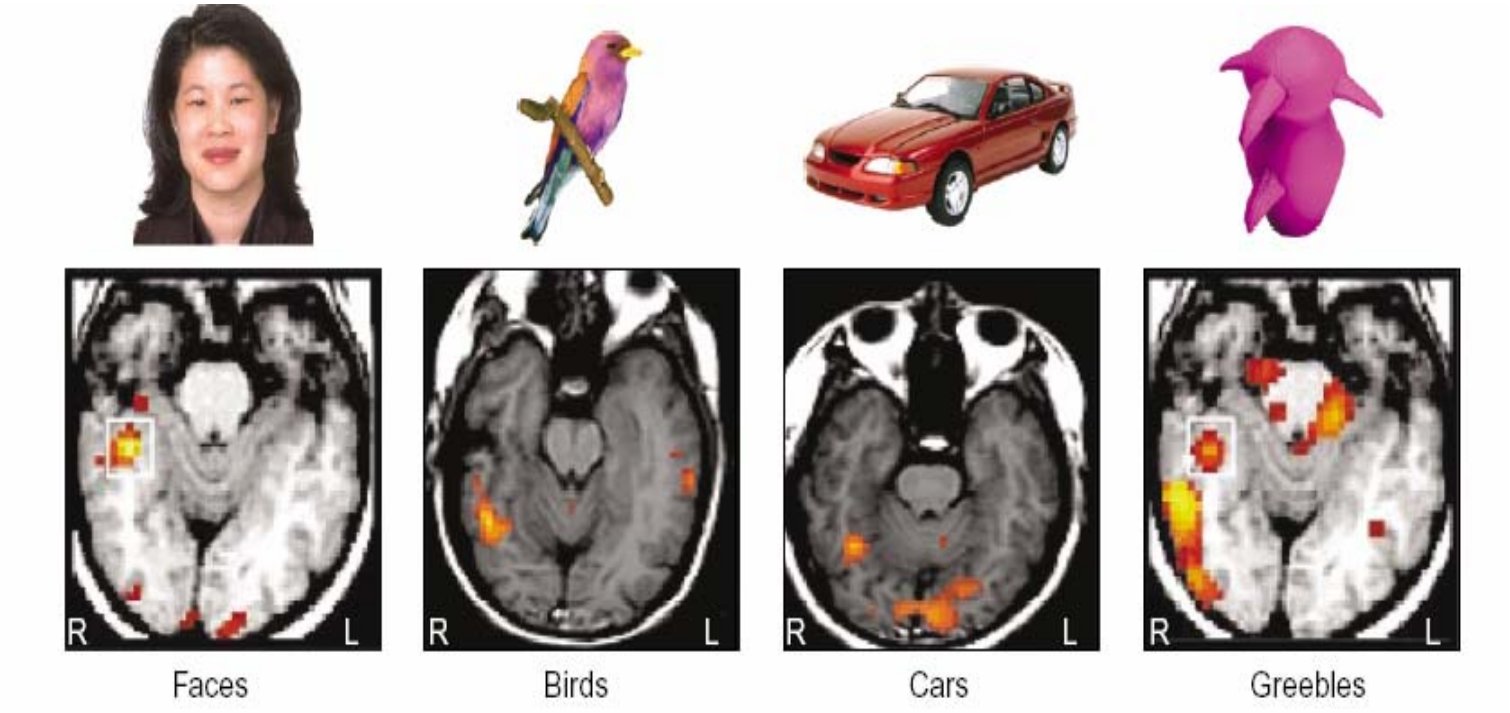



Buffy
Joyce
Dawn
Buffy
Buffy
Giles
Other
Tara
Willow
Willow
Buffy
Dawn
Willow
Tara
Willow
Willow
Buffy
Dawn
Tara
Buffy
Willow
Tara
Xander
Anya
Buffy
Riley
Spike

Face vs. Object Recognition

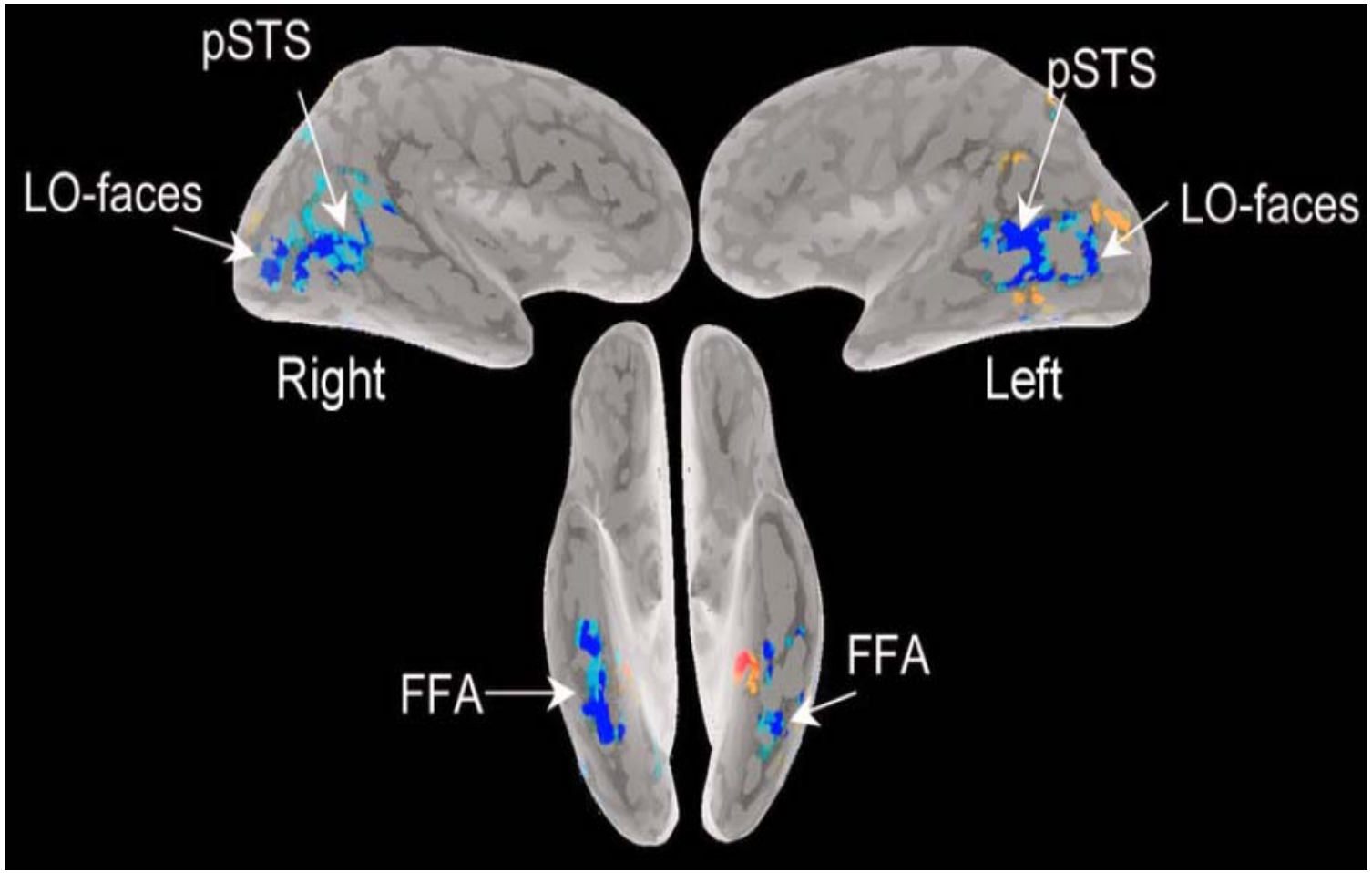
- Is face recognition different from general object recognition?
  - fMRI measurements
  - Prosopagnosia and agnosia
  - Prosopamnesia
  - Capgras syndrome
- Is there a module in the brain for face recognition?

fMRI Experiments



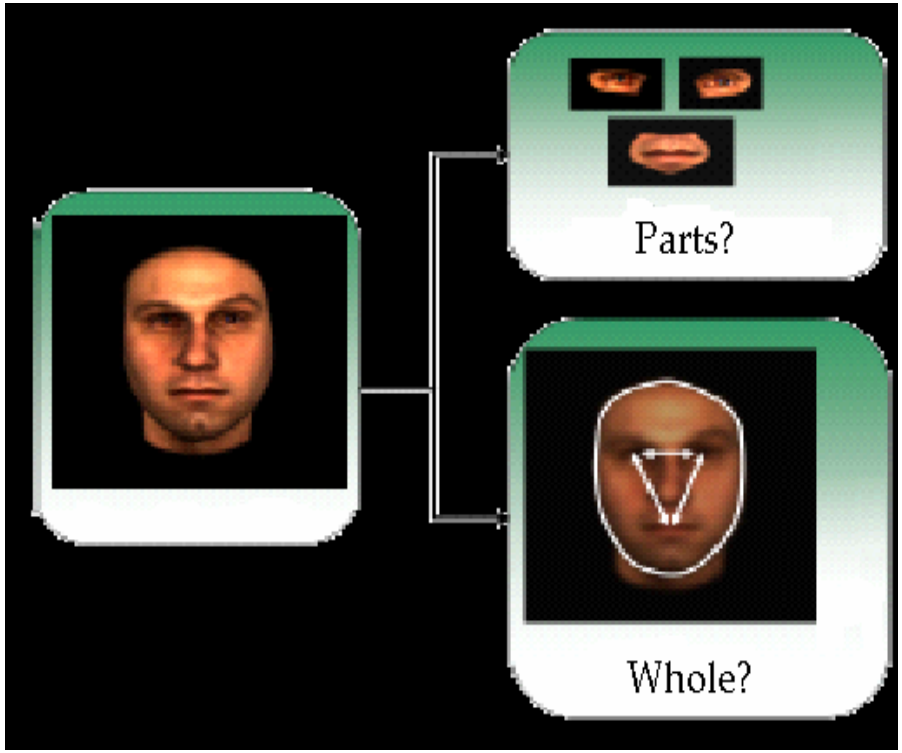
Gauthier, I., M.J. Tarr, *Vision Research* vol.37, pp.1673-1682, 1997

Activation for Faces



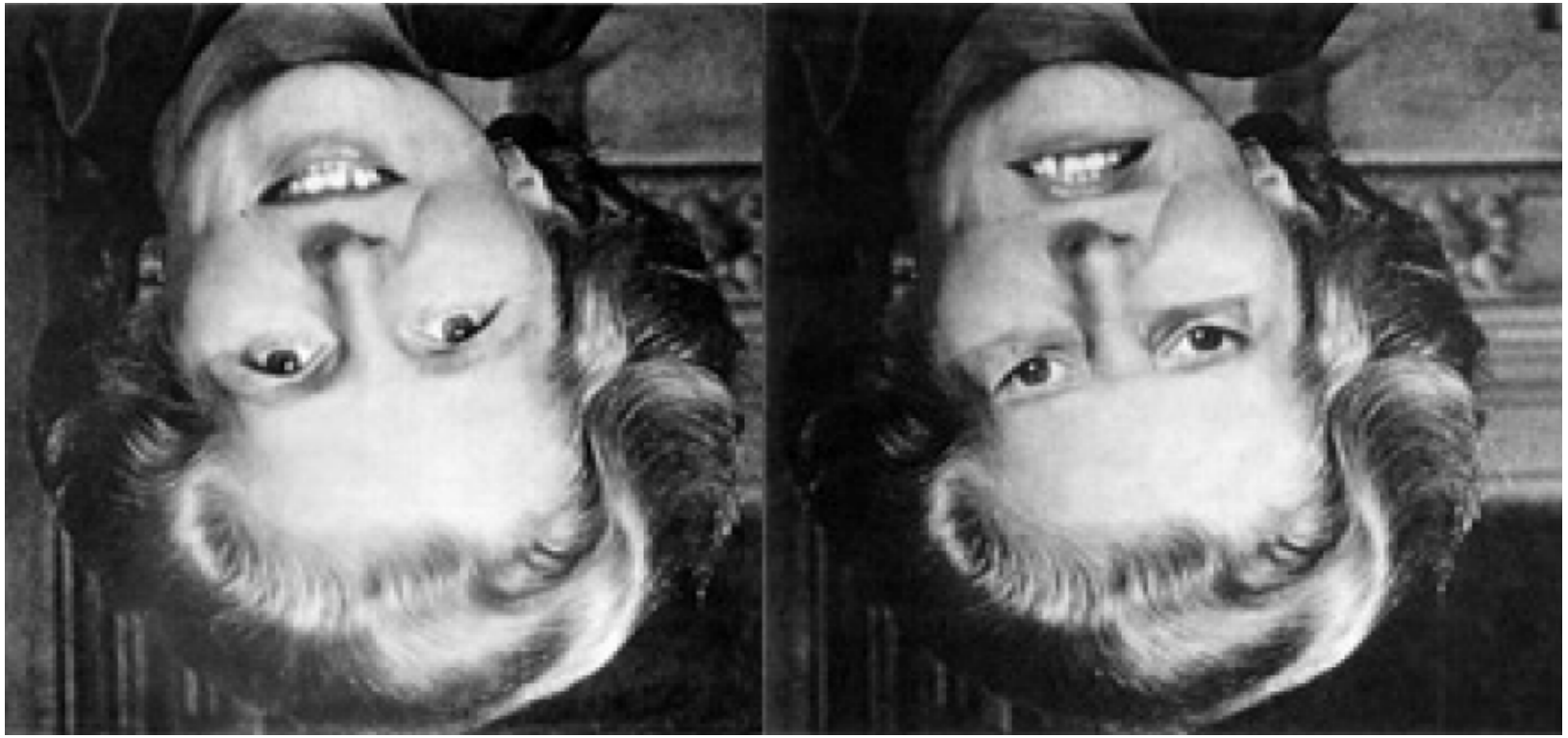
Grill-Spector, Knouf, Kanwisher, Nature Neuroscience 2004

Parts or Wholes?



Wallraven, Schwaninger, Bulthoff

Thatcher Illusion

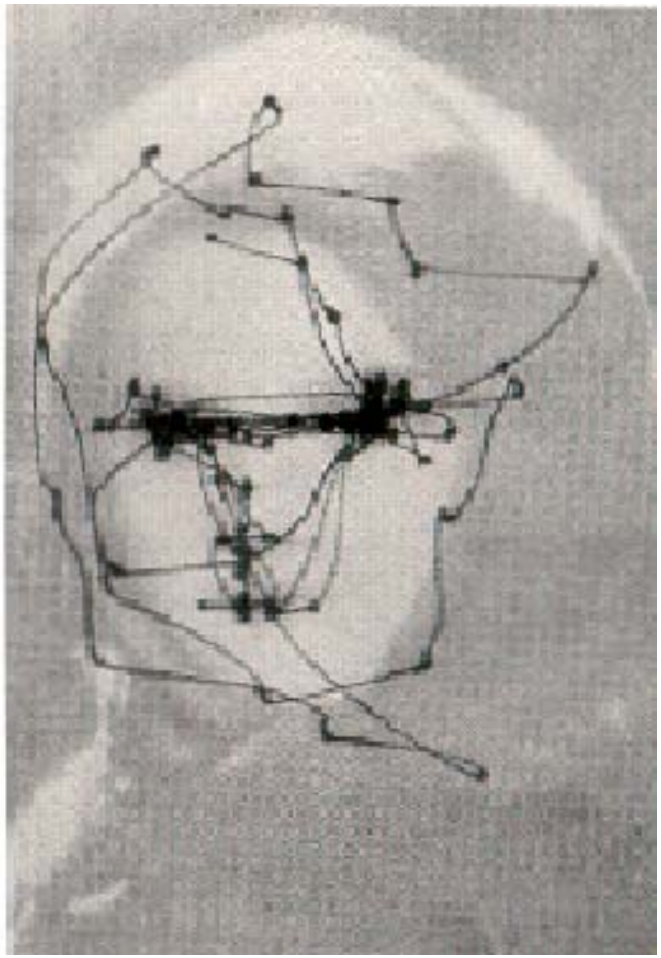


Thompson, *Perception*, vol.9, pp.483-484, 1980

Thatcher Illusion

Thompson, *Perception*, vol.9, pp.483-484, 1980

Selective Attention



Yarbus, A.L.,Eye Movement and Vision, 1976

Outline

- ***Facial emotion recognition***

Face

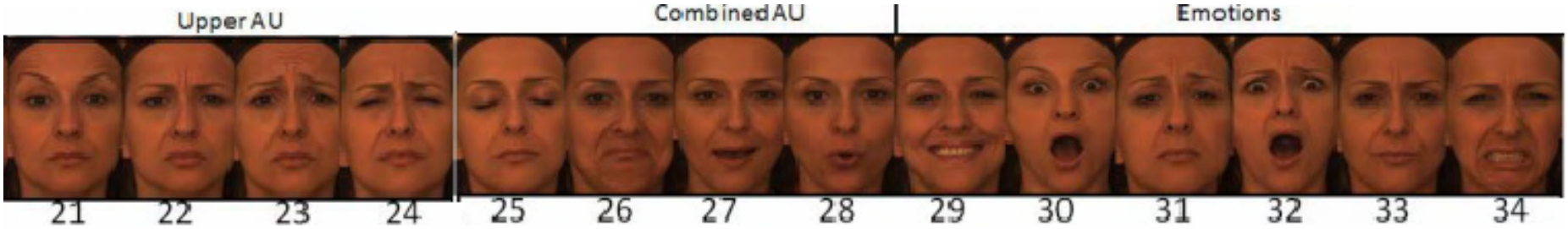
Lower Action Units

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	•	•

Face

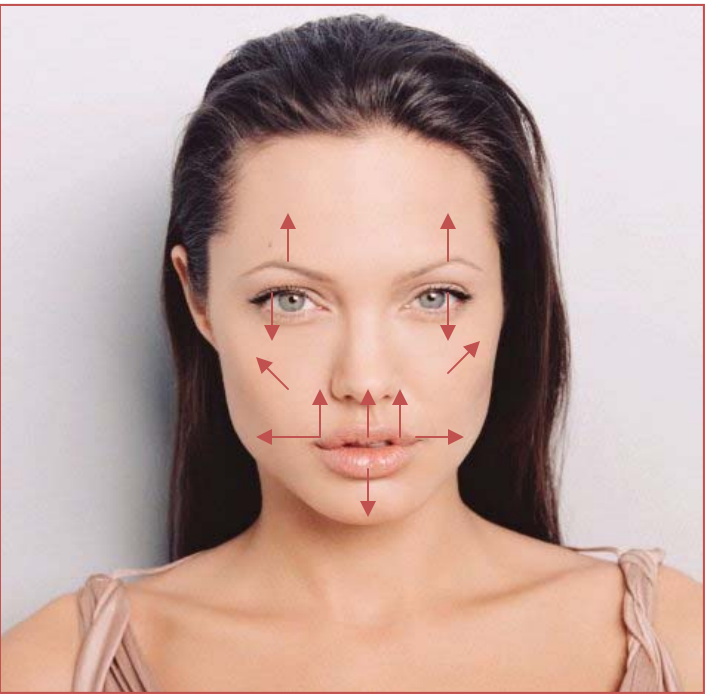
Upper/Combined Action Units + Basic Expressions

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	•	



Software

Facial Expression Recognition



12 facial motion measurements

vertical movement of the lips

horizontal movement of the mouth corners

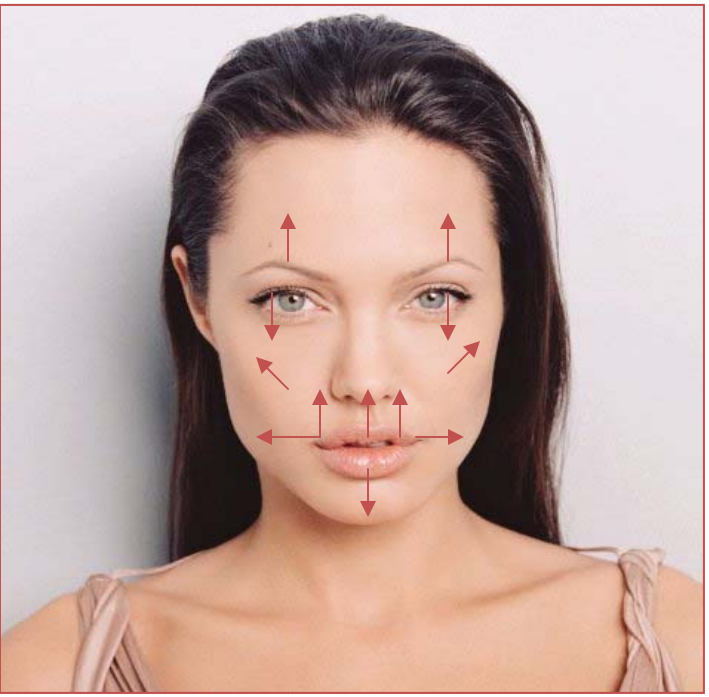
vertical movement of the mouth corners

vertical movement of the eye brows

lifting of the cheeks

blinking of the eyes

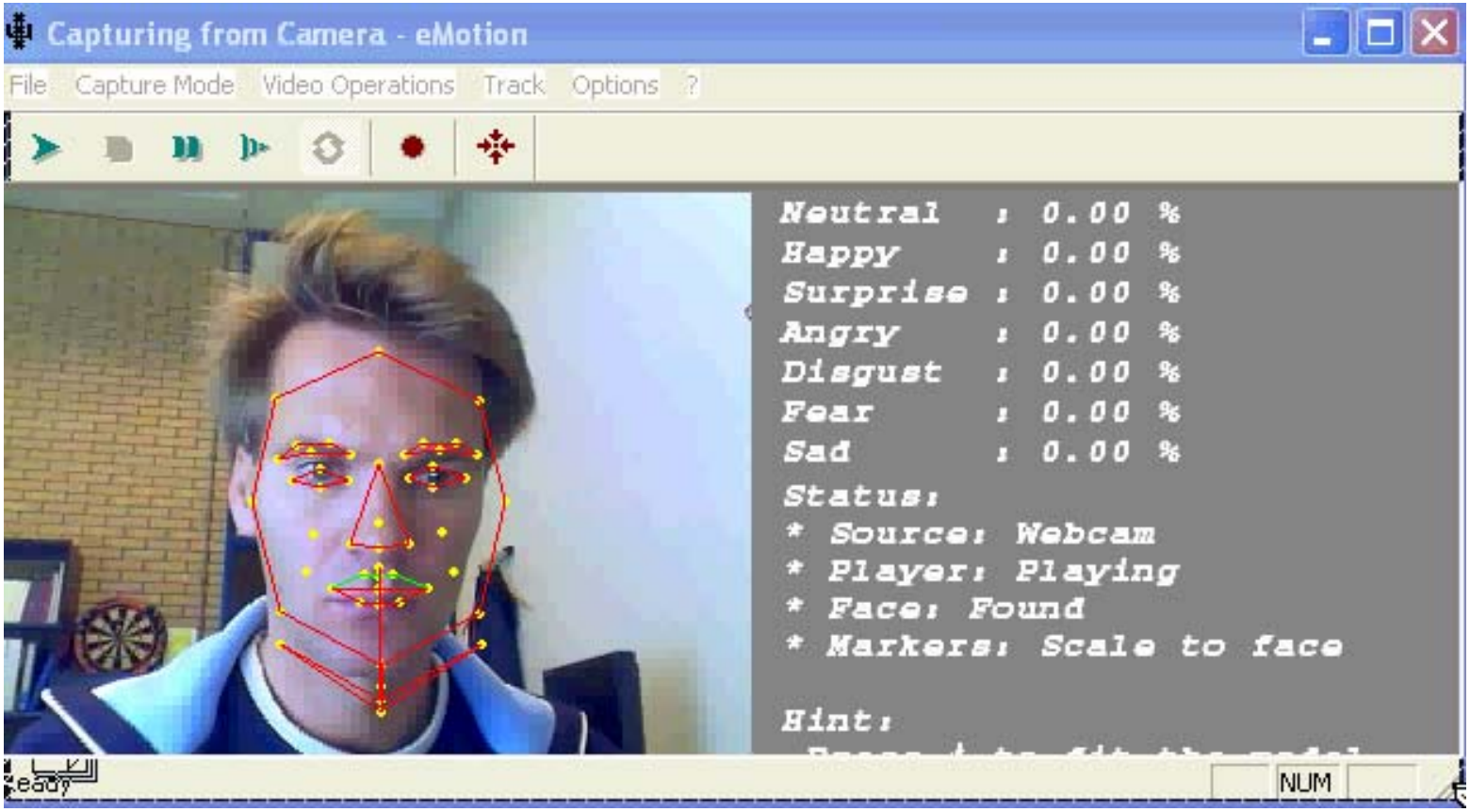
Facial Expression Recognition



We use 12 facial features = 12 facial motion measurements

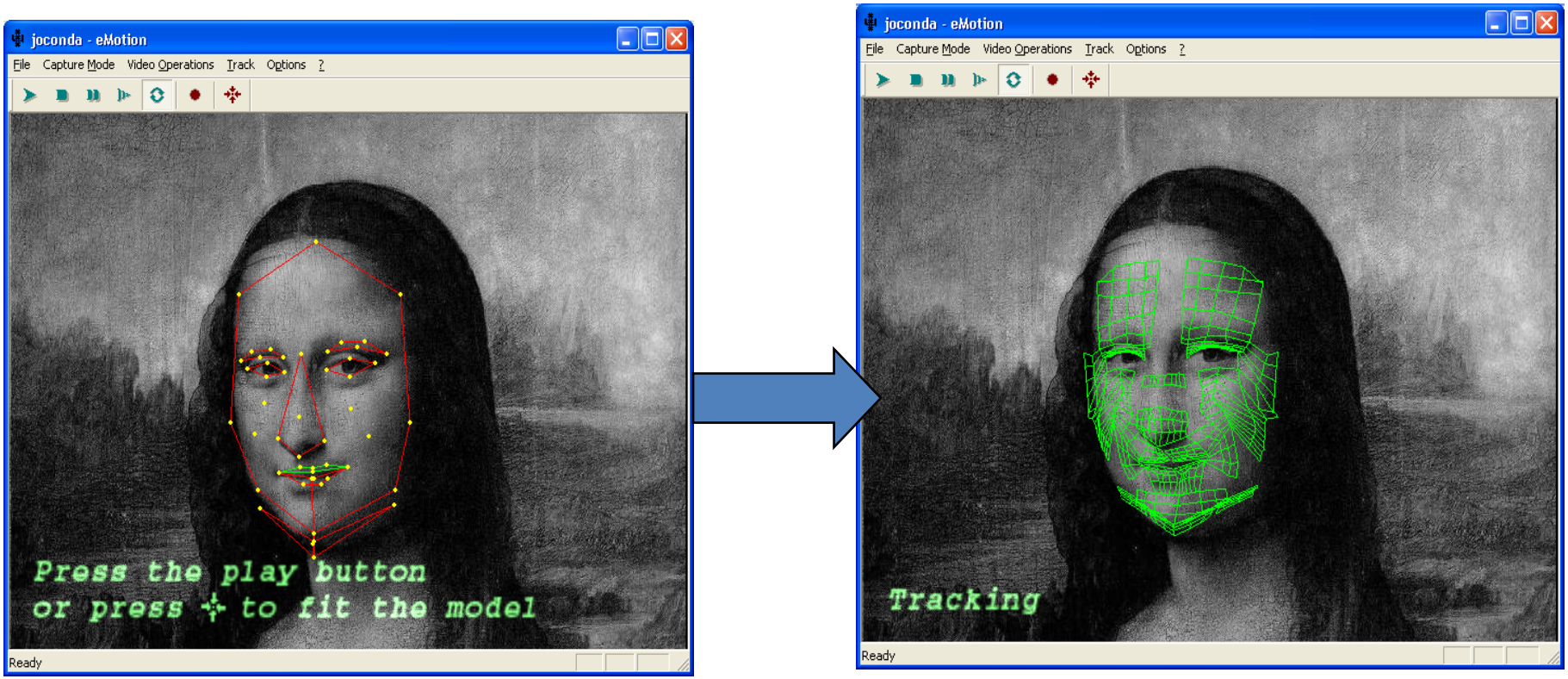
The combination of these features define the 7 basic classes of facial expression we want to classify: ***Neutral, Happy, Anger, Disgust, Fear, Sad, Surprise***

Facial Expression Recognition



Facial Expression Recognition

Nicu Sebe



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

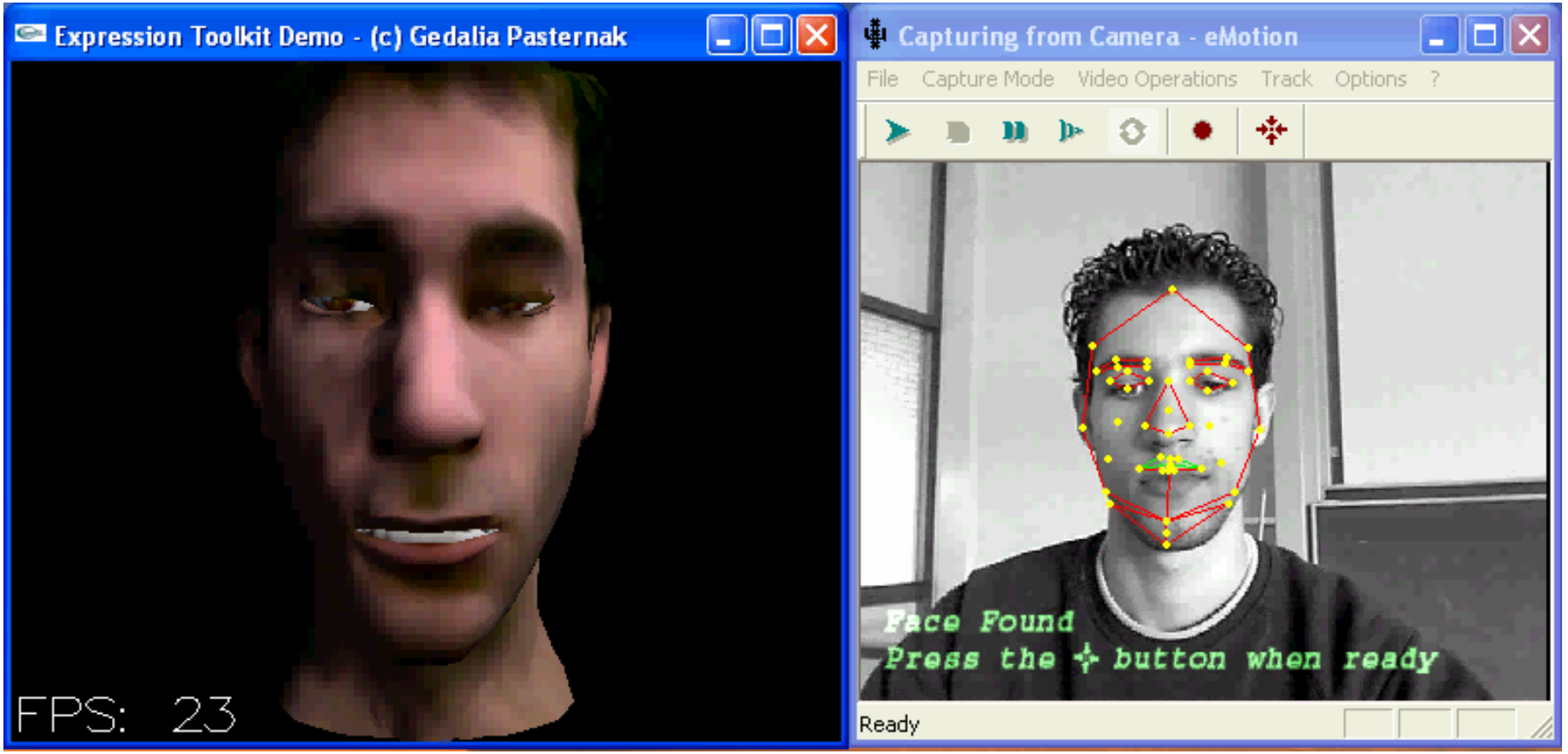
- Datum
- Meest vrolijk
- Meest verrast
- Meest boos
- Meeste walging
- Meest angstig
- Meest droevig

Selecteer een tag

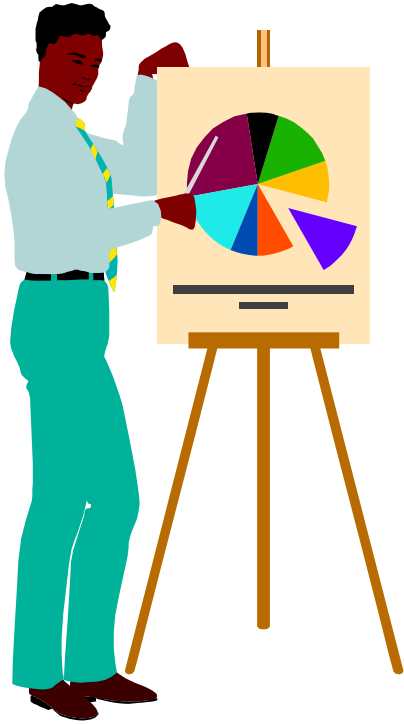
Acda Balkenende Bangebroek Duh Gezicht Glimlach Huilen Lachen Man Smile Voetbal Vrouw

Change language to [English]

Avatar


Expression Toolkit Demo - (c) Gedalia Pasternak
FPS: 23
Capturing from Camera - eMotion
File
Capture Mode
Video Operations
Track
Options
?
Face Found
Press the button when ready
Ready

Today’s class



PART I

Activity Recognition

Social Signal Processing

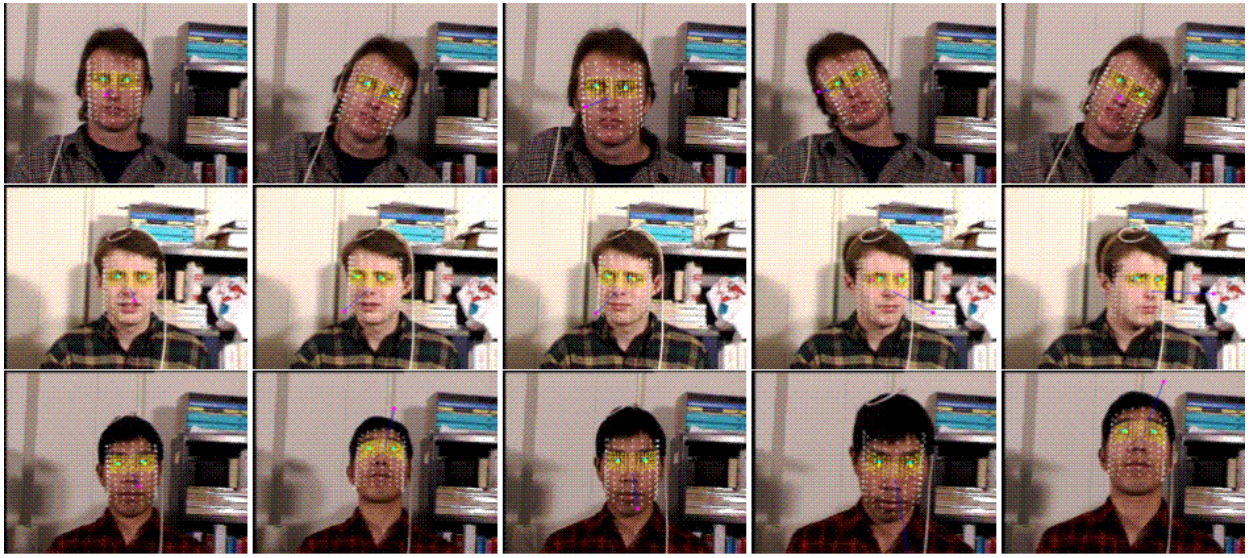
Face and Facial Expression Recognition

Gaze Estimation

[Valenti, cvpr08]

Human behaviour understanding

- Facial expression
- Head pose
- Eye Tracking
- Voice



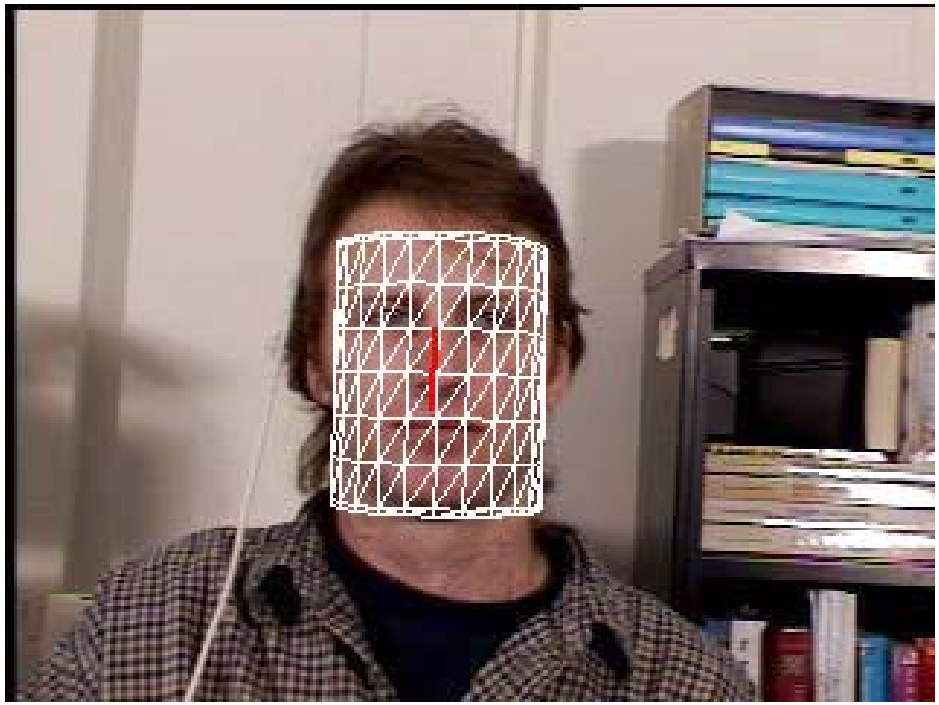
Visual Gaze Estimation

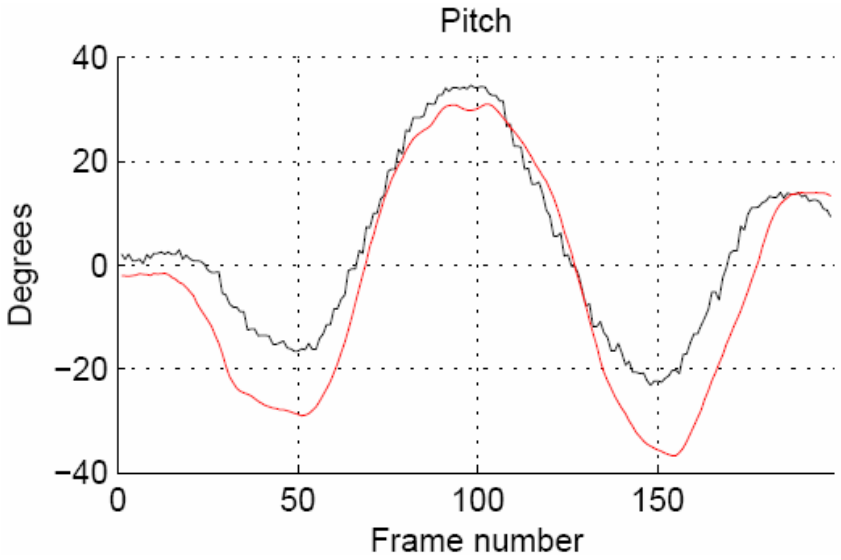
The big picture

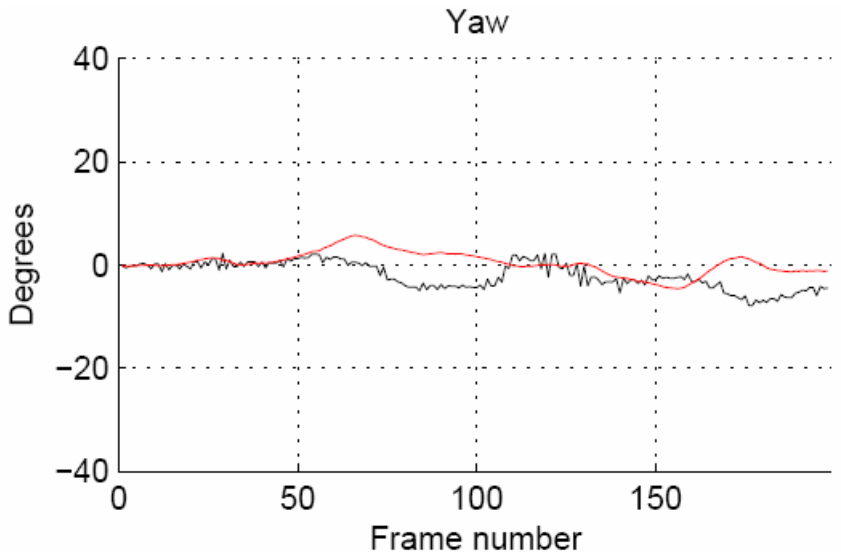
Visual Gaze Estimation

Experiments

- Boston University



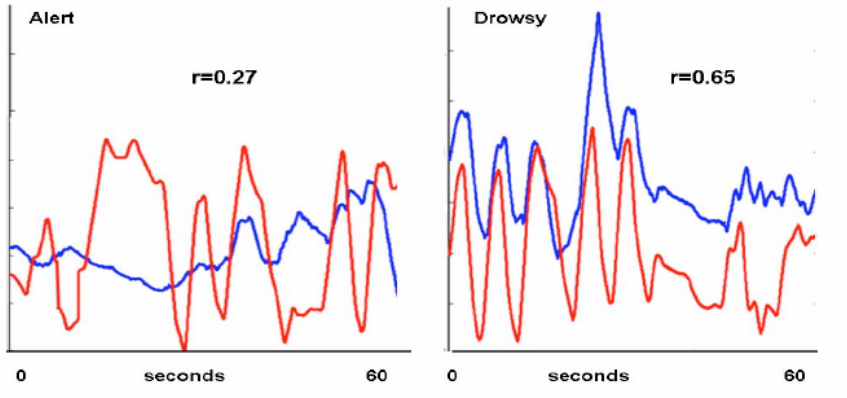




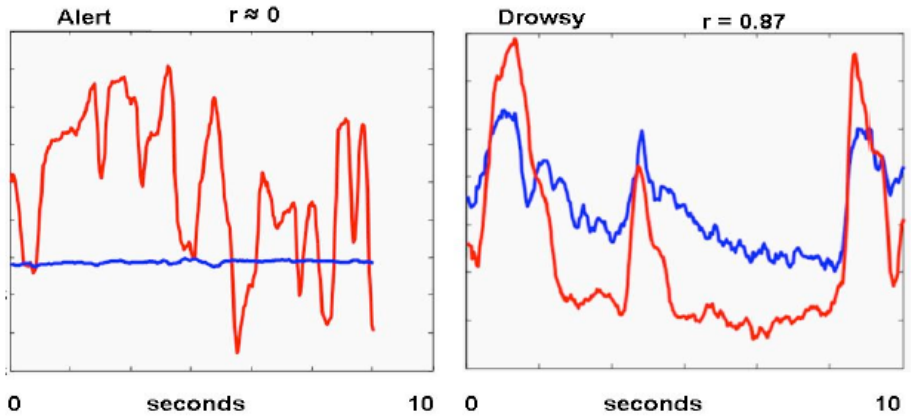
Dataset

Dataset

Detecting driver drowsiness



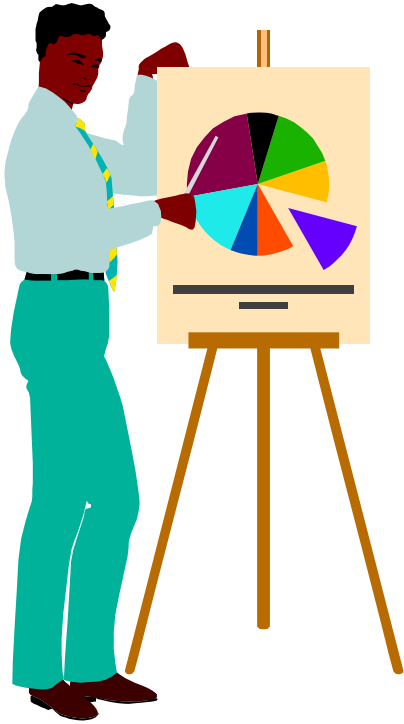
Head motion (blue) and steering position (red)



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

Bartlett et al., 2008

Summary



Activity Recognition

Social Signal Processing

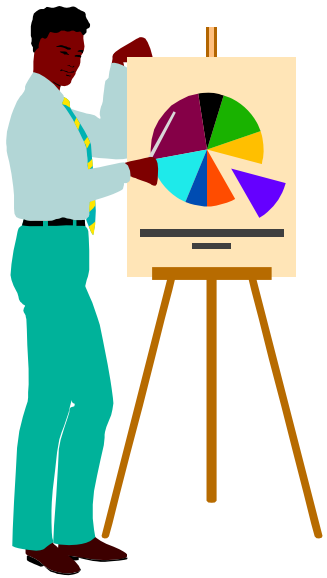
Face and Facial Expression Recognition

Gaze Estimation

Today's class

Part I: Human Behavior Analysis

Part II: Summary of the Lectures



Lecture 2
Image Formation

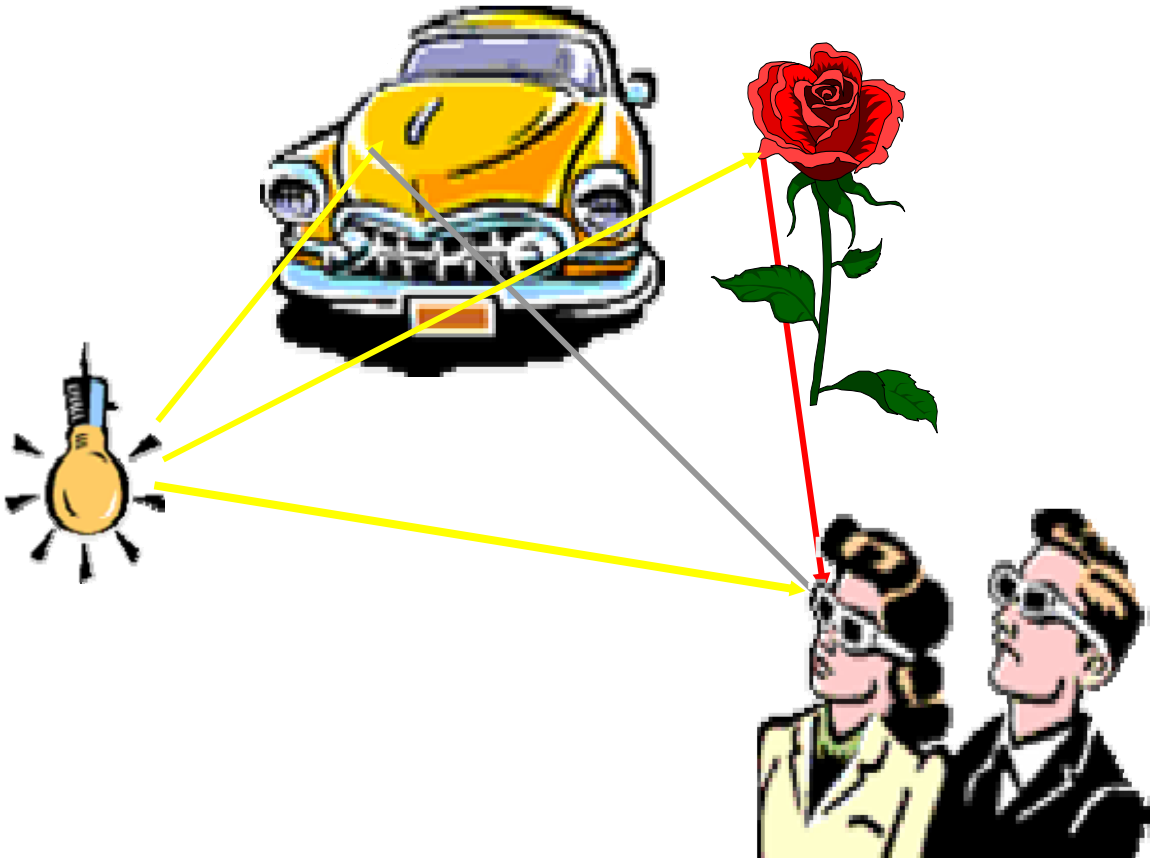
What makes an image?

the triplet light-objects-observer

Light source

Object(s)

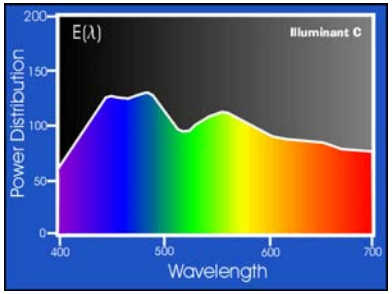
Sensor



What makes an image?

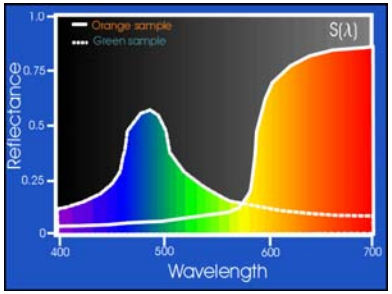
the triplet light-objects-observer

Light source



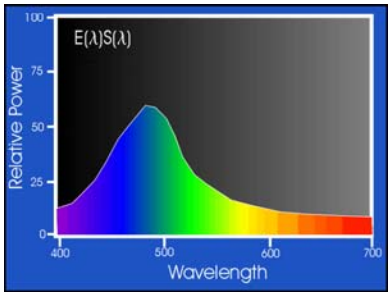
$e(\lambda)$

Object



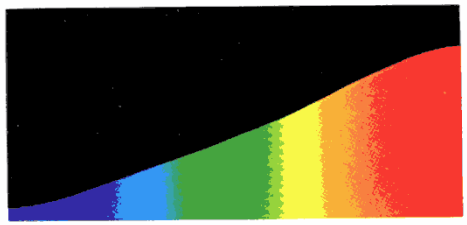
$\rho(\lambda)$

Sensor

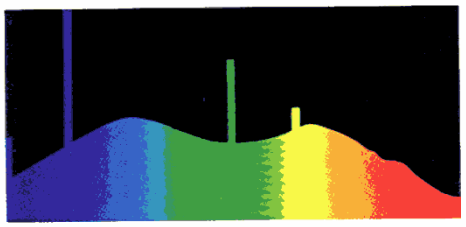


$e(\lambda)\rho(\lambda)$

Light sources and illuminants



Incandescent lamp

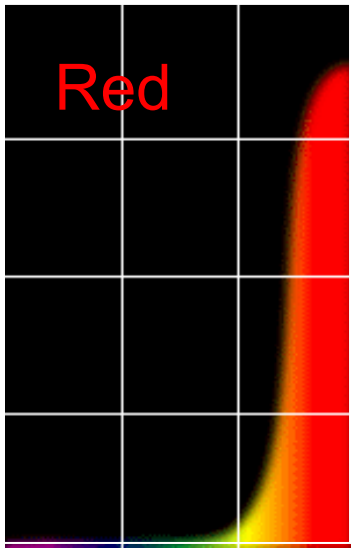


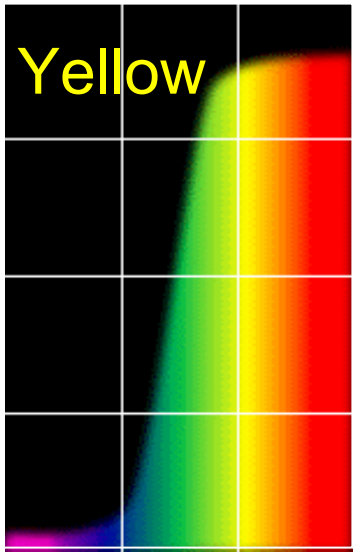
Fluorescent lamp

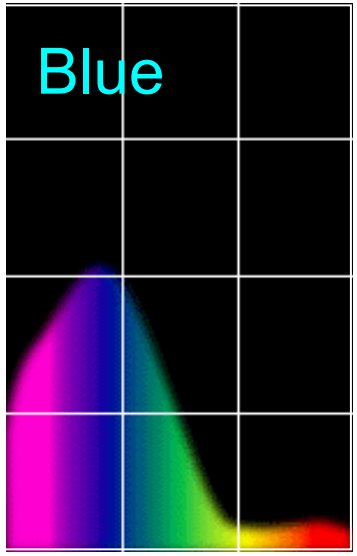
Object Colours

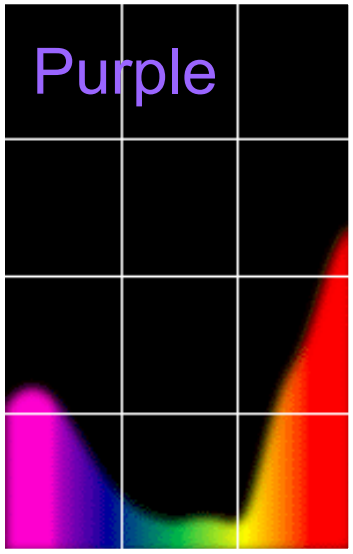
Some examples of the reflectance spectra of surfaces









Wavelength (nm)

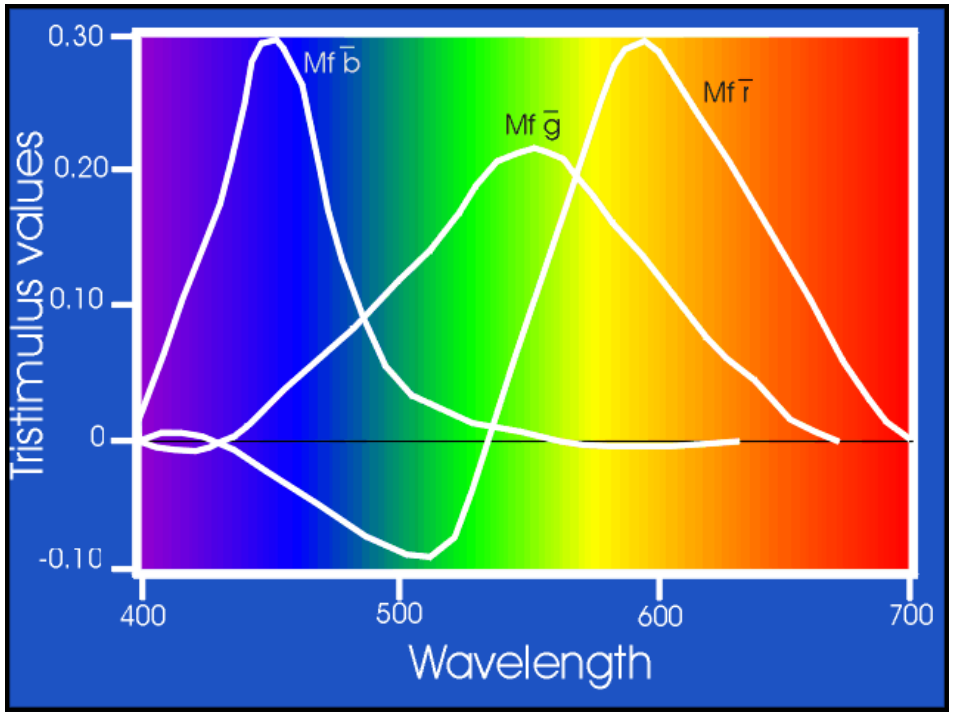
Observer: Trichromacy

Young-Helmholtz approach

Tristimulus values R, G, and B

Wright (7) Guild (10)

Stiles and Burch (50)

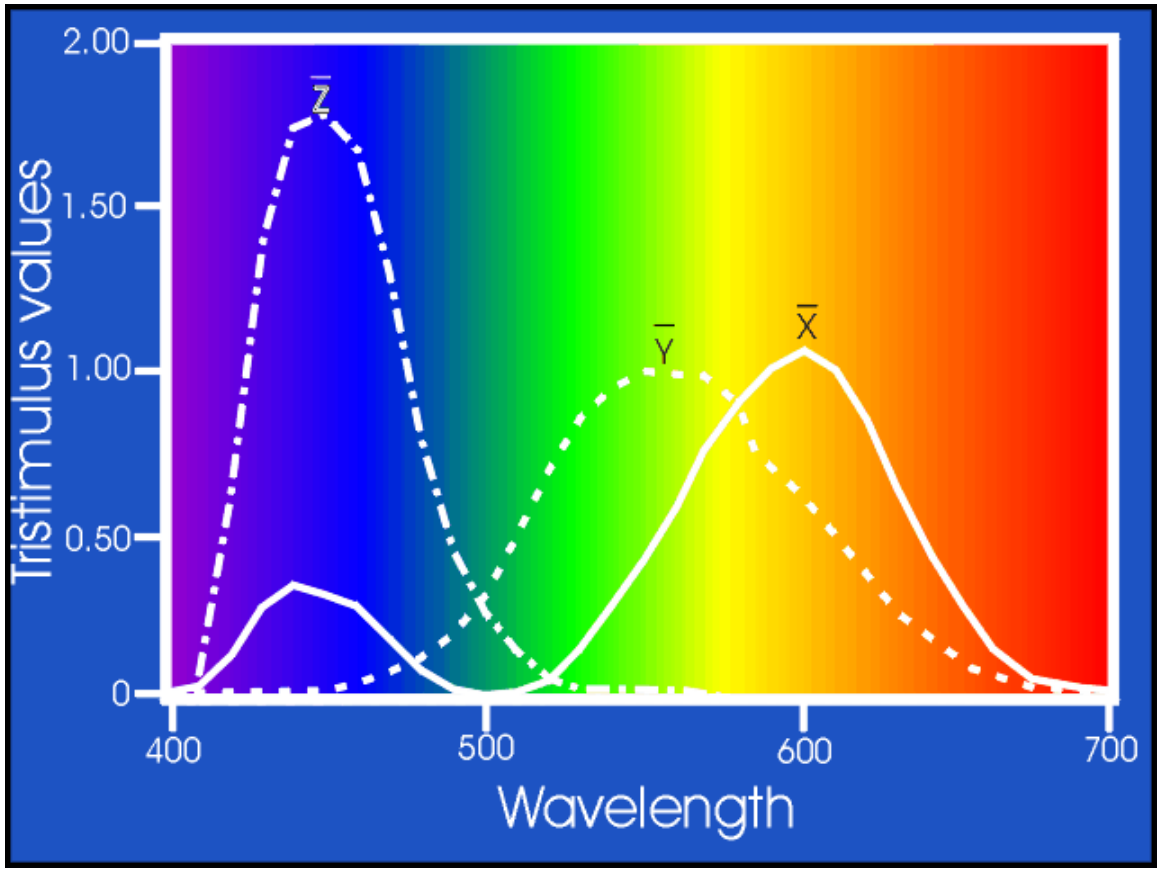


Colorimetry: CIE XYZ-system

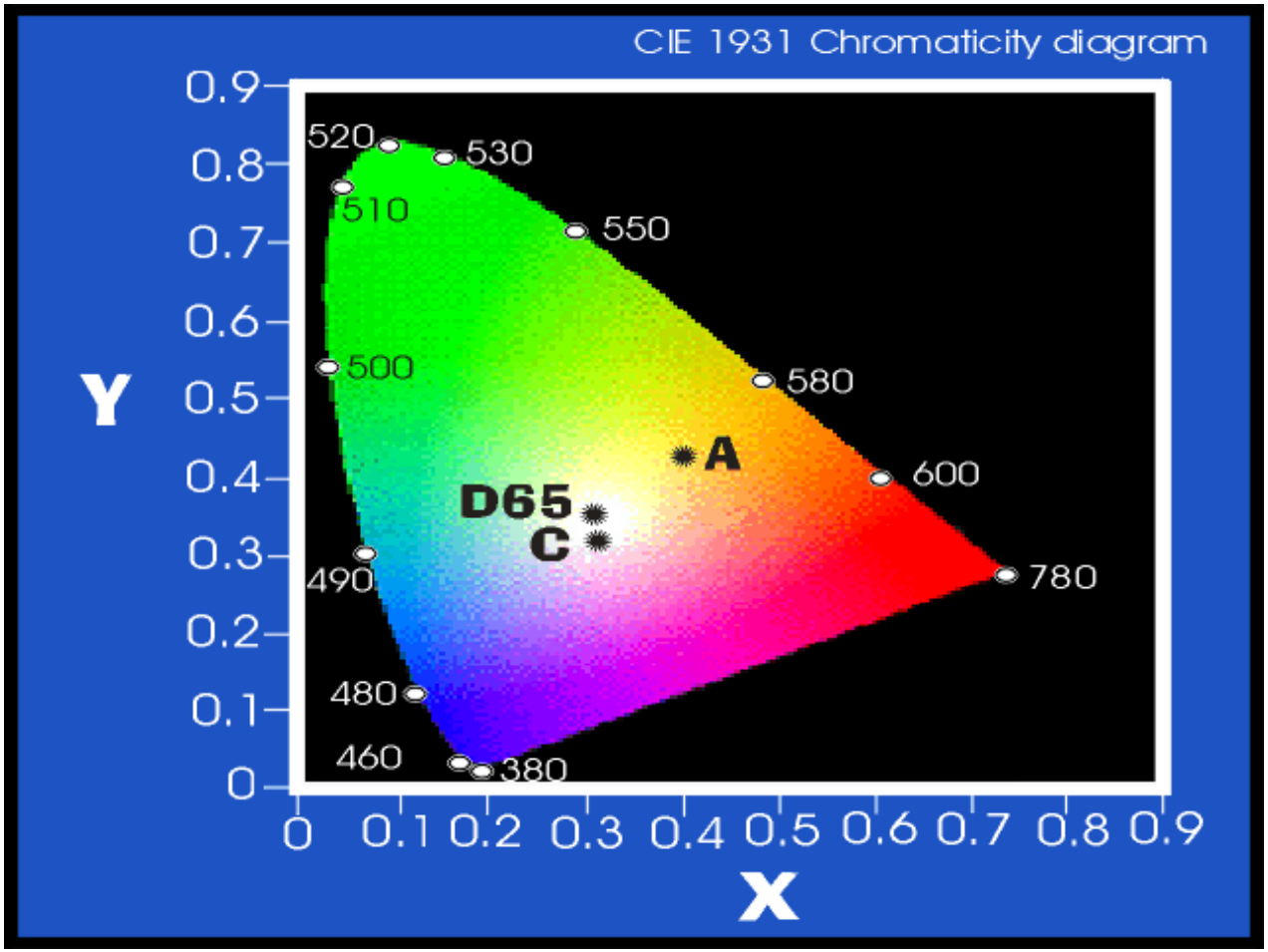
$$X = \int_{\lambda} e(\lambda)\rho(\lambda)\bar{x}(\lambda)d\lambda$$

$$Y = \int_{\lambda} e(\lambda)\rho(\lambda)\bar{y}(\lambda)d\lambda$$

$$Z = \int_{\lambda} e(\lambda)\rho(\lambda)\bar{z}(\lambda)d\lambda$$



Colorimetry: Illuminants in the xy-plane



Lecture 3

Color Invariance and Image Processing

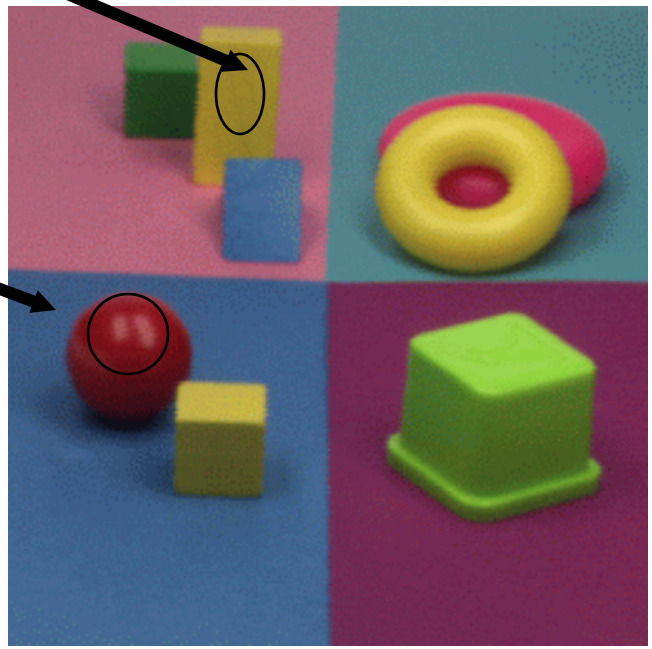
Reflection Model

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

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

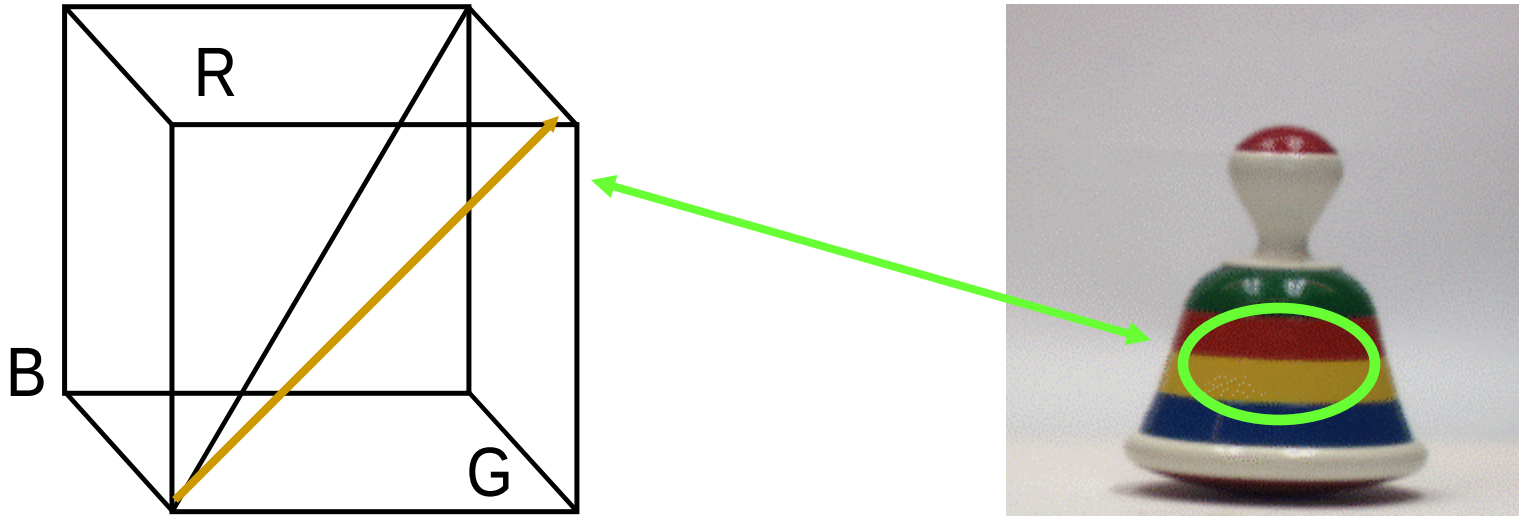
$c_b(\lambda)$	surface albedo	scene & viewpoint invariant
$e(\lambda)$	illumination	scene dependent
$\vec{n}$	object surface normal	object shape variant
$\vec{s}$	illumination direction	scene dependent
$\vec{v}$	viewer's direction	viewpoint variant
$f_C(\lambda)$	sensor sensitivity	scene dependent

body reflectance in RGB - space

Consider the body reflection term :

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

for example the geometric term is Lambertian i.e. $m_b(\vec{n}, \vec{s}) = \vec{n} \cdot \vec{s}$



rgb – Photometric Invariance: Proof

Consider the body reflection term:

$$C_b = em_b(\vec{n}, \vec{s})k_C$$

For $C_b = \{R, G, B\}$ giving the red, green and blue sensor response under white light. Further, with $k_C = \int f_C(\lambda)c_b(\lambda)d\lambda$

$$r(R_b, G_b, B_b) = \frac{em_b(\vec{n}, \vec{s})k_R}{em_b(\vec{n}, \vec{s})(k_R + k_G + k_B)} = \frac{k_R}{(k_R + k_G + k_B)}$$

$$g(R_b, G_b, B_b) = \frac{em_b(\vec{n}, \vec{s})k_G}{em_b(\vec{n}, \vec{s})(k_R + k_G + k_B)} = \frac{k_G}{(k_R + k_G + k_B)}$$

$$b(R_b, G_b, B_b) = \frac{em_b(\vec{n}, \vec{s})k_B}{em_b(\vec{n}, \vec{s})(k_R + k_G + k_B)} = \frac{k_B}{(k_R + k_G + k_B)}$$

Colour invariance - Summary

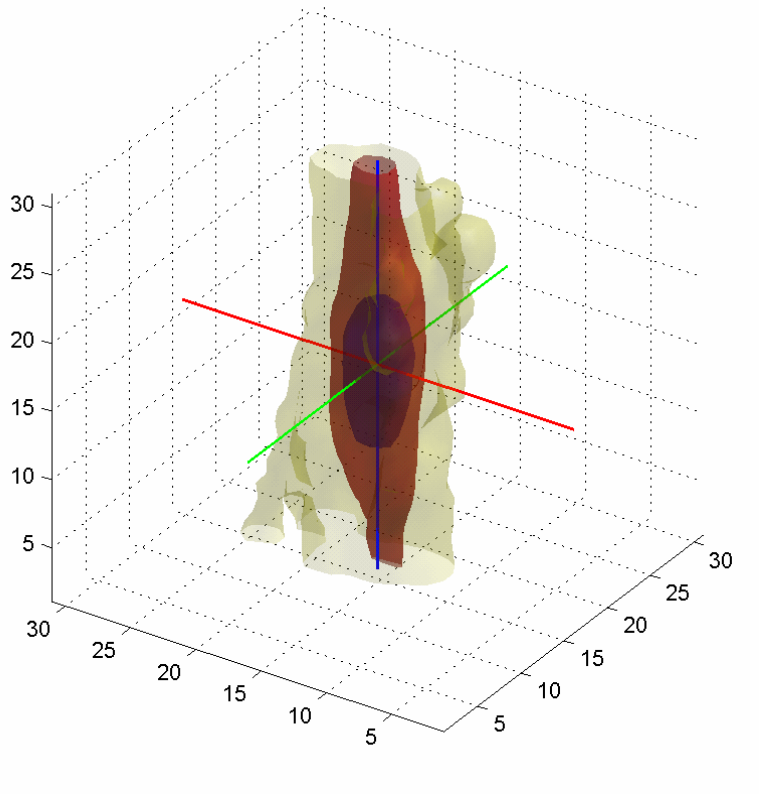
	shadows	shading	highlights	ill. intensity	ill. Colour
I	-	-	-	-	-
R,G,B	-	-	-	-	-
r,g,b	+	+	-	+	-
c1,c2,c3	+	+	-	+	-
Hue	+	+	+	+	-
l1,l2,l3	+	+	+	+	-
m1m2m3	+	+	-	+	+

- no invariance

+ invariance

Color constancy

Color constancy



Color Constancy

- Model images assuming Lambertian reflectance:

$$\mathbf{f}(\mathbf{x}) = \int_{\omega} e(\lambda)\rho_k(\lambda)s(\mathbf{x},\lambda)d\lambda$$

- Image transformation using von Kries model[1]:

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix}$$

(1) - J. von Kries. “Influence of adaptation on the effects produced by luminous stimuli.” In D. MacAdam, editor, *Sources of Color Vision*, pages 109–119. MIT Press, 1970.

Color constancy

color constancy : the ability to recognize colors of objects invariant of the color of the light source.

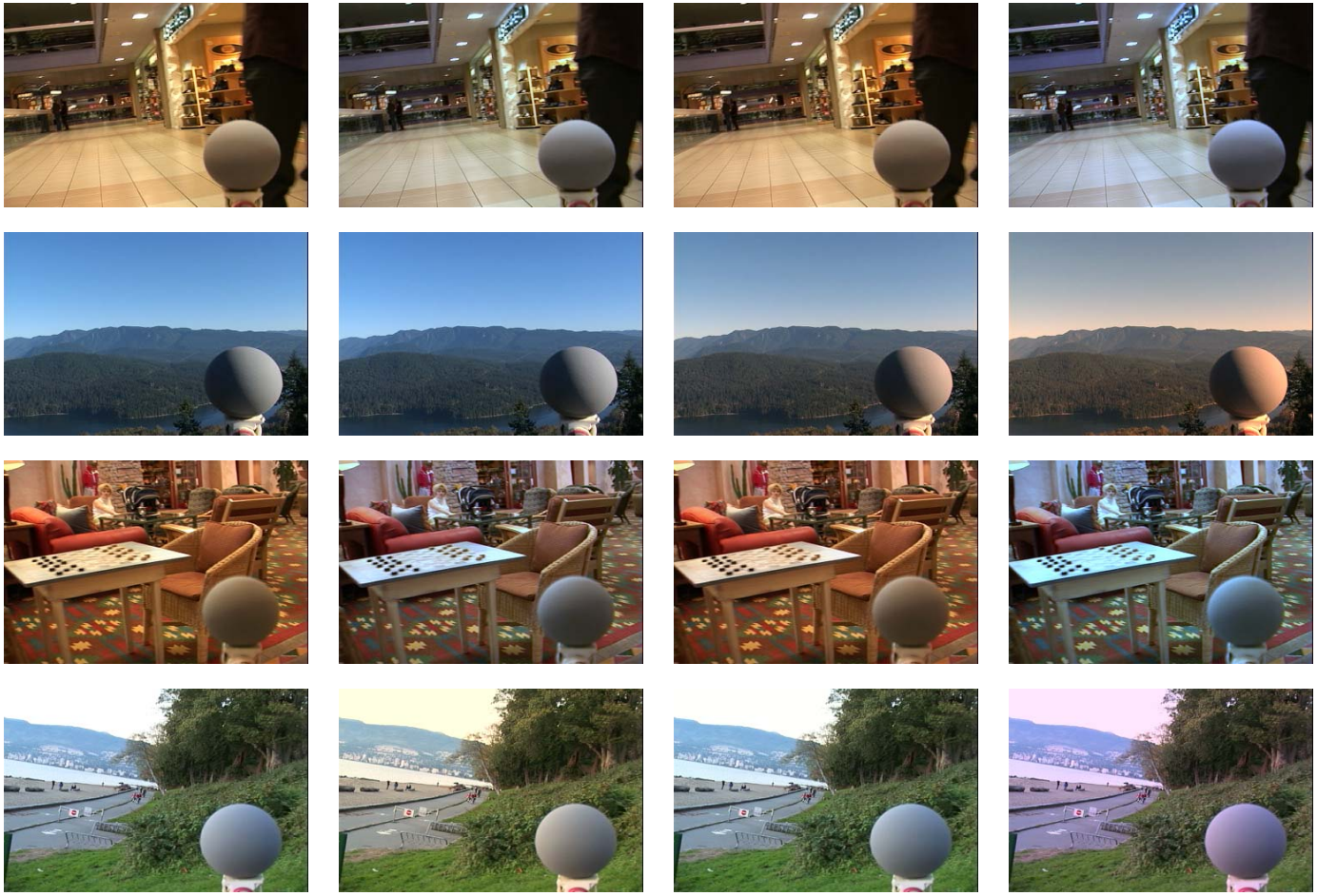
Grey world hypothesis : the average reflectance in a scene is grey.

White patch hypothesis : the highest value in the image is white.

Grey edge hypothesis : the average edge difference in a scene is grey

White Patch and Grey-world Examples

Original image **Ideal correction** **White-Patch** **Grey-world**



Images from “A large image data set for color constancy research”, by F. Ciurea and B. Funt in CIC 2003.

Natural Image Statistics

2nd order Grey-edge

White-Patch

Grey-world

Grey-edge

2nd order Grey-edge

Gamma app. resolution

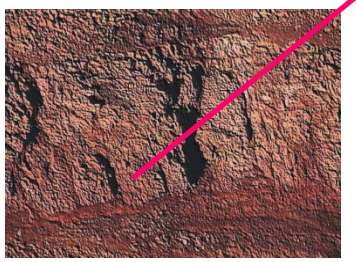
1.5

1

0.5

b=0.012192, g=1.81543

b=0.095042, g=1.50293



b=0.251007, g=1.354492

b=0.01023, g=0.989258

b=0.100916, g=1.02832

b=0.178397, g=1.043945

b=0.010076, g=0.510742

b=0.101547, g=0.641602

b=0.182459, g=0.647461

0.01

0.1

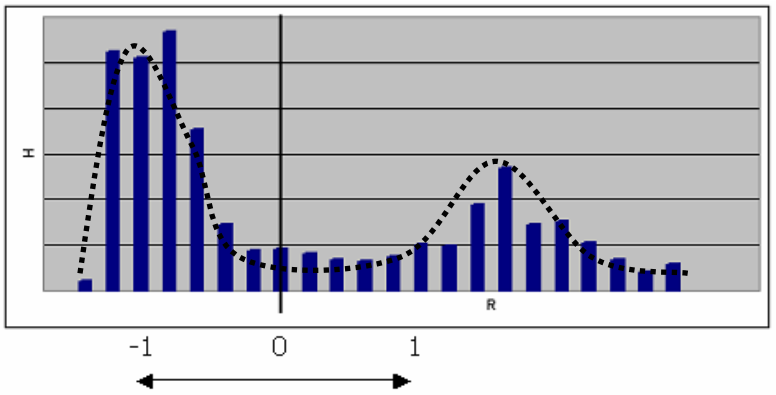
0.2

Beta app. Contrast

Color Constancy Examples

Original image	Ideal correction	White-Patch	Grey-world	NIS-combined

Kernel Density Estimation for Object Recognition



Color Invariants: Noise Propagation

Suppose that $u,...,w$ are measured with corresponding uncertainties $\sigma_u,...,\sigma_w$ to compute function $q(u,...,w)$.

The predicted uncertainty is defined by :

$$\sigma_q = \sqrt{\left(\frac{\partial q}{\partial u}\sigma_u\right)^2 + ... + \left(\frac{\partial q}{\partial w}\sigma_w\right)^2}$$

The uncertainty in q is never larger than the ordinary sum

$$\sigma_q \leq \left|\frac{\partial q}{\partial u}\right|\sigma_u + ... \left|\frac{\partial q}{\partial w}\right|\sigma_w$$

if and only if the uncertainties $\sigma_u,...,\sigma_w$ are relatively small.

Color Invariants: Noise Propagation

Example: rgb

Function $q(x,\ldots,z)$ then $\delta_q = \sqrt{\left(\frac{\partial q}{\partial x}\delta x\right)^2 + \ldots + \left(\frac{\partial q}{\partial z}\delta z\right)^2}$

$$r(R,G,B) = \frac{R}{R+G+B}$$

Normalized color value

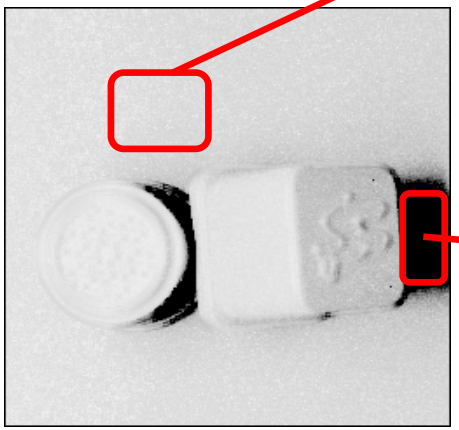
$$\delta_r = \frac{\sqrt{R^2(\delta_B^2+\delta_G^2)+(B+G)\delta_R^2}}{(B+G+R)^4}$$

Predicted uncertainty of normalized color value

Measured values and predicted uncertainty

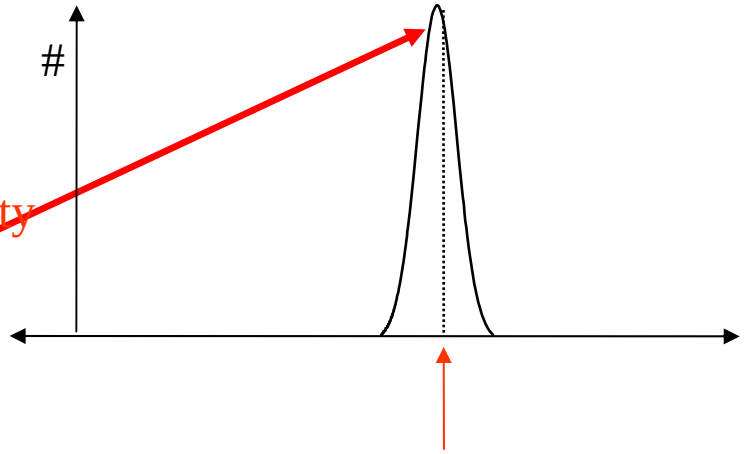
Variable Kernel Density Estimation Example

image



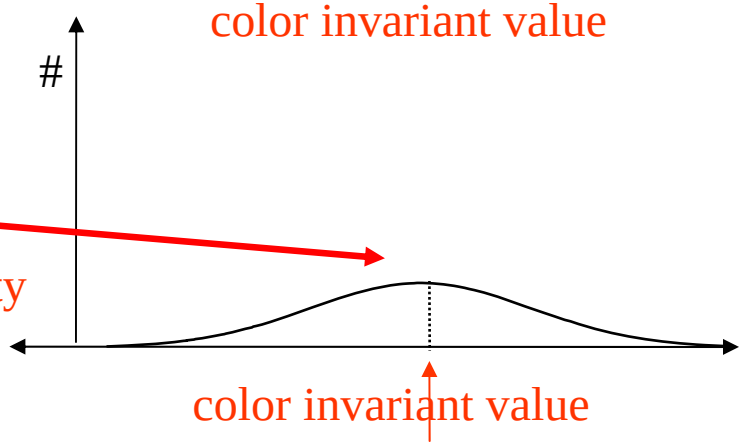
uncertainty

High certainty



color invariant value

Low certainty

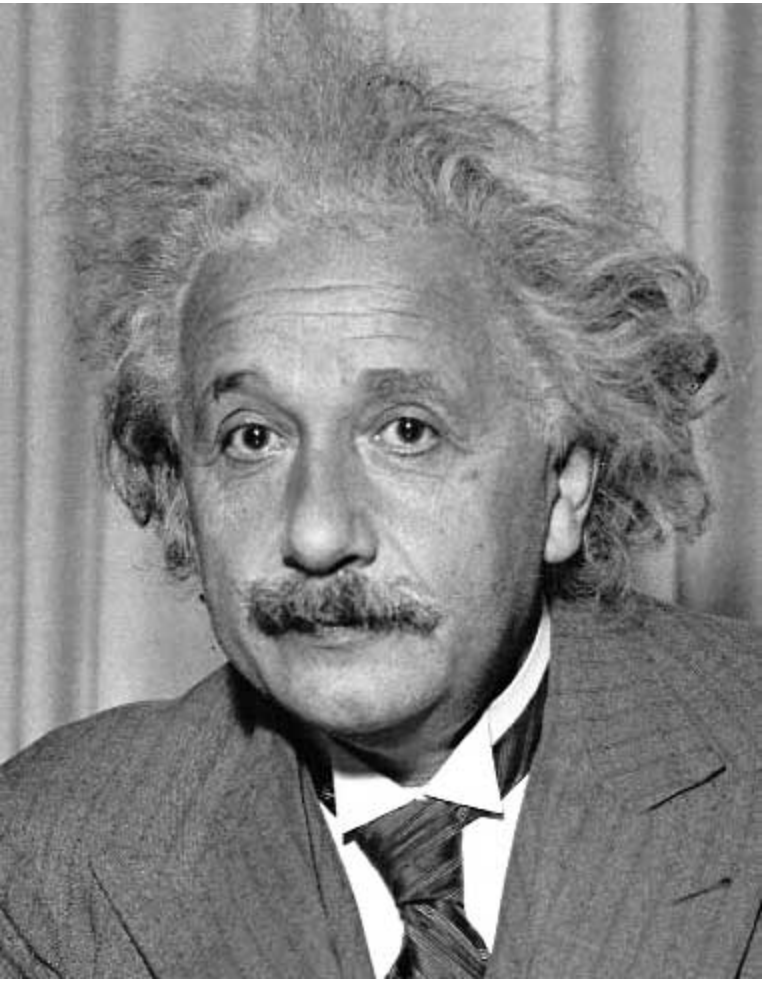


color invariant value

Lecture 3

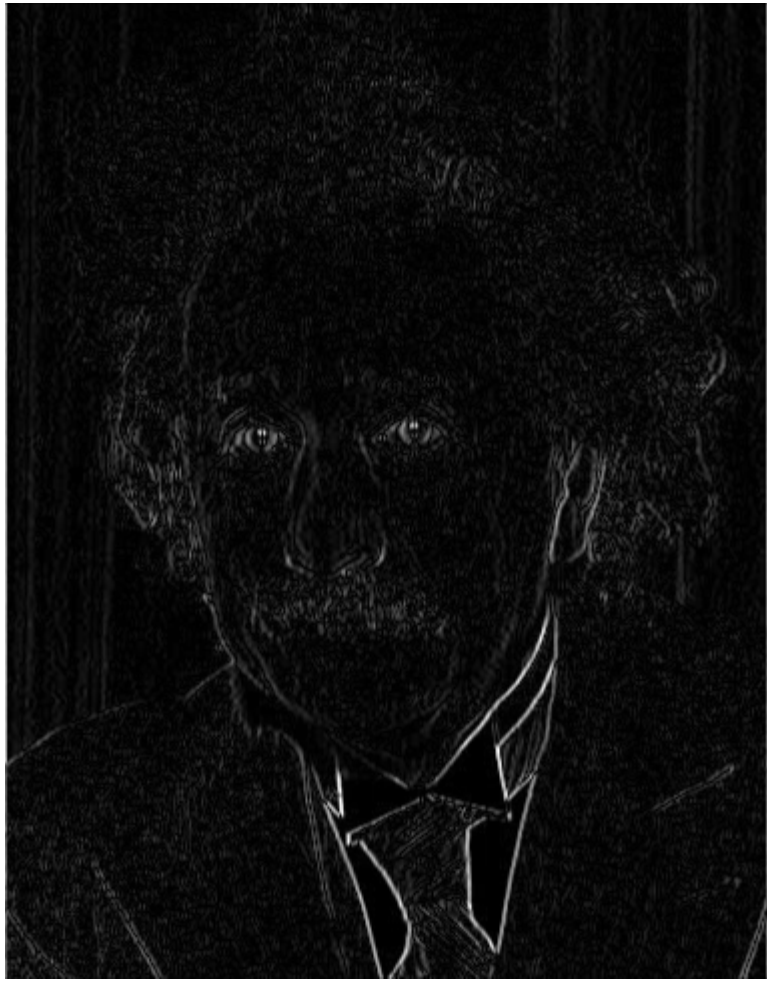
Image Processing

Edge Filters



1	0	-1
2	0	-2
1	0	-1

Sobel



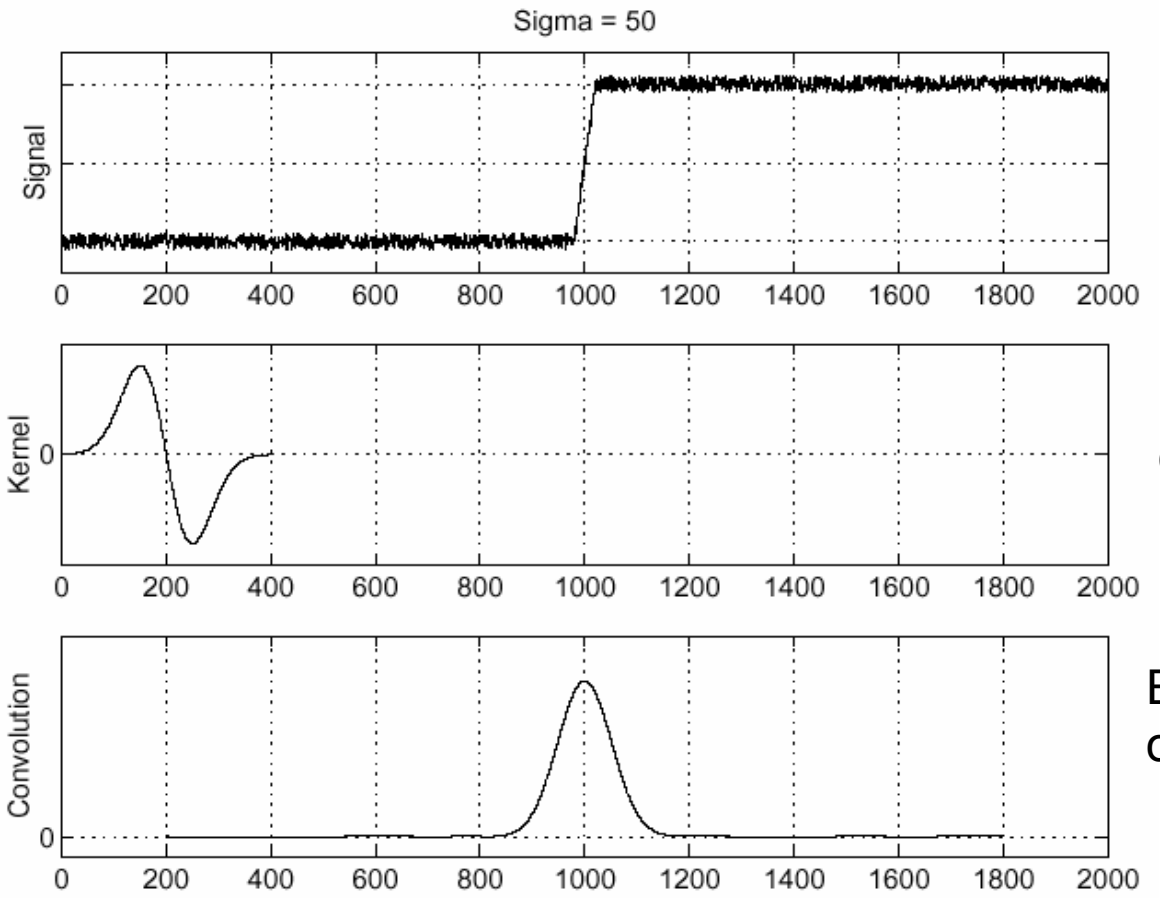
Vertical Edge
(absolute value)

Review: Edge Detection

f

$\frac{d}{dx} g$

$f * \frac{d}{dx} g$



Edge

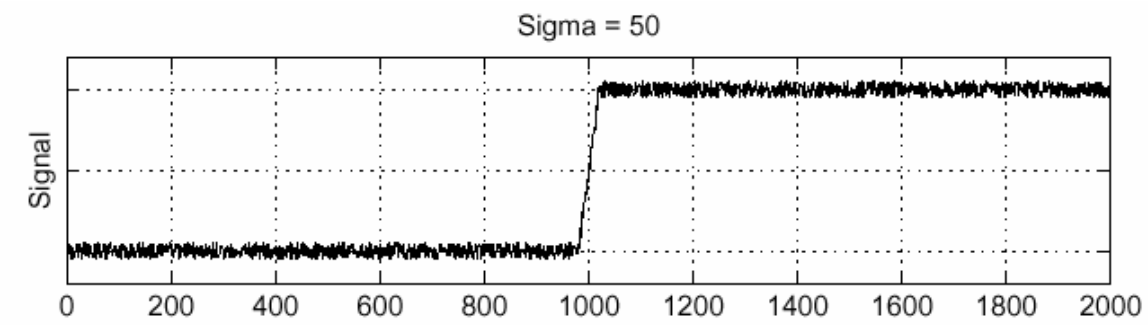
Derivative of Gaussian

Edge = maximum of derivative

Source: S. Seitz

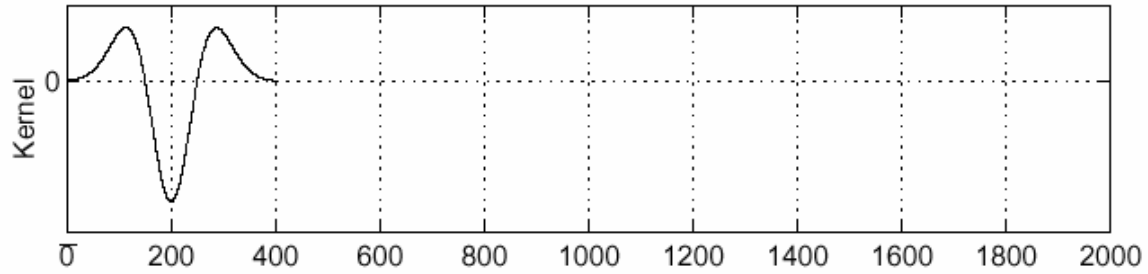
Review: Edge Detection

f



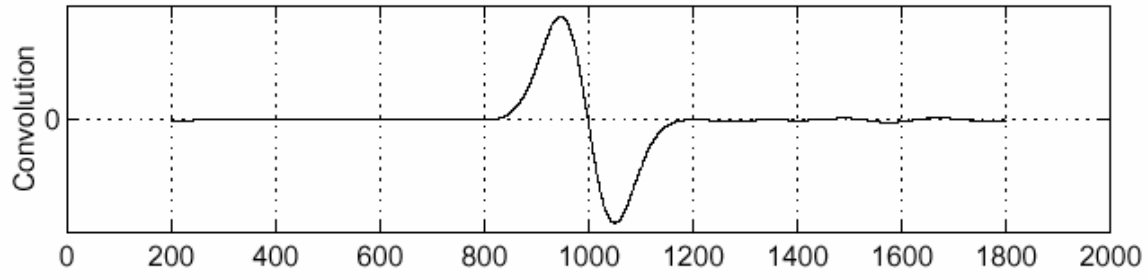
Edge

$\frac{d^2}{dx^2} g$



Second derivative of Gaussian (Laplacian)

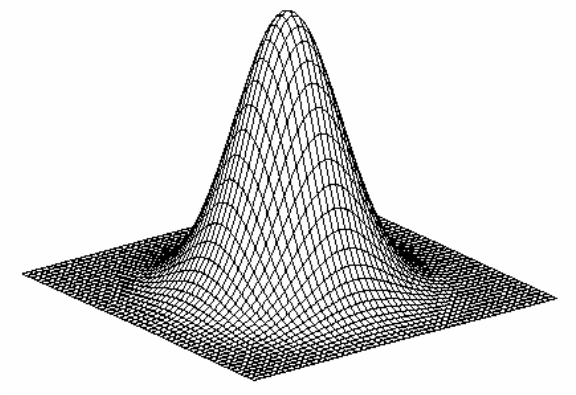
$f * \frac{d^2}{dx^2} g$



Edge = zero crossing of second derivative

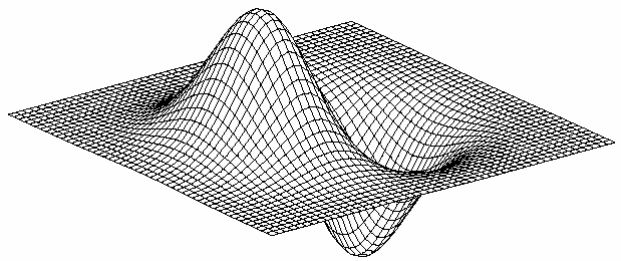
Source: S. Seitz

2D Edge Detection Filters



Gaussian

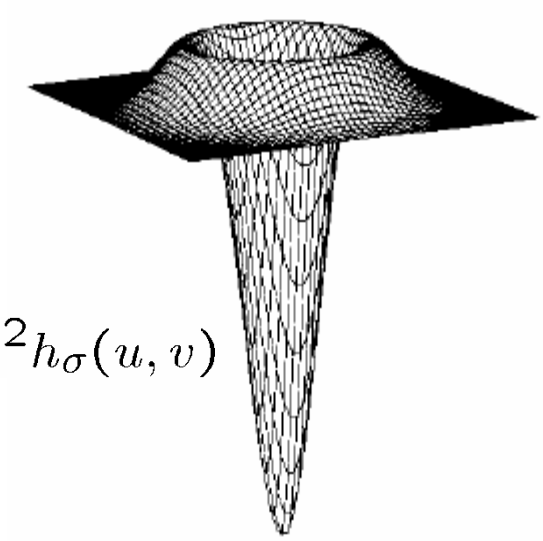
$$h_\sigma(u,v) = \frac{1}{2\pi\sigma^2} e^{-\frac{u^2+v^2}{2\sigma^2}}$$



derivative of Gaussian

$$\frac{\partial}{\partial x} h_\sigma(u,v)$$

Laplacian of Gaussian



$$\nabla^2 h_\sigma(u,v)$$

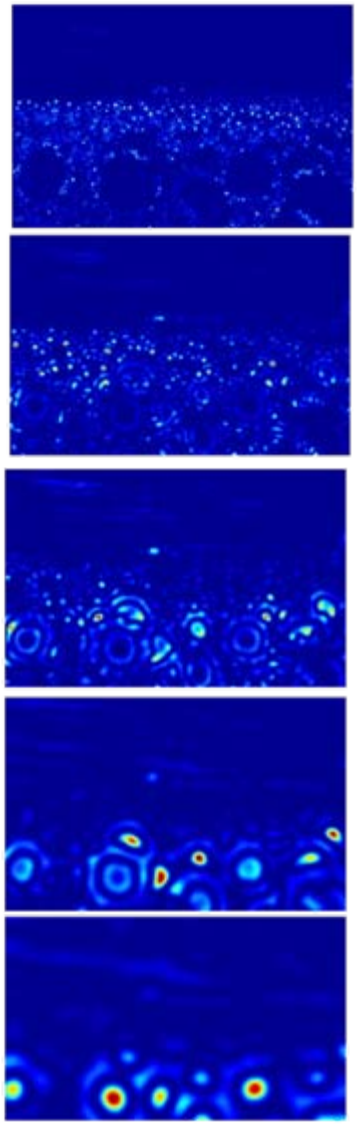
∇^2 is the **Laplacian** operator:

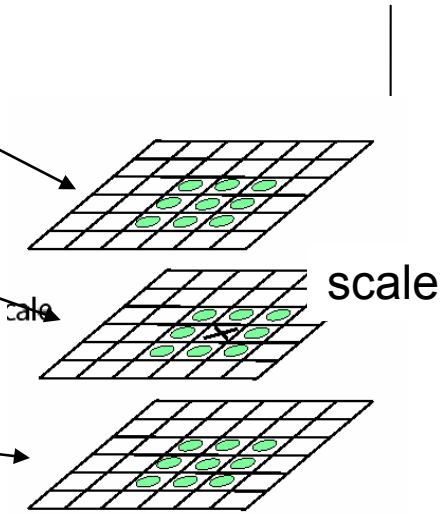
$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

Scale invariant interest points

Interest points are local maxima in both position and scale.

$$L_{xx}(\sigma) + L_{yy}(\sigma)$$





⇒ **List of** ***(x, y, σ)***

Squared filter response maps

Lecture 4

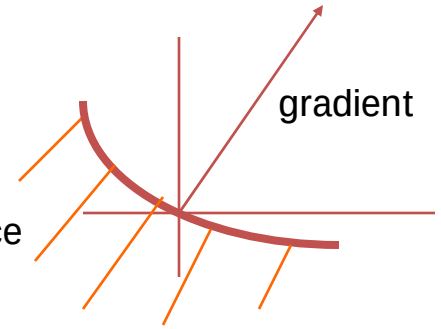
Feature Extraction and Tracking

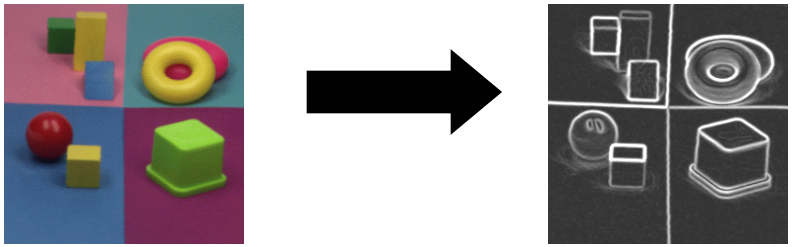
Edge Detection

An intensity edge is defined as a point where...

the gradient is large:

$$|\nabla f(x,y)| = \sqrt{f_x^2 + f_y^2} >> 0$$

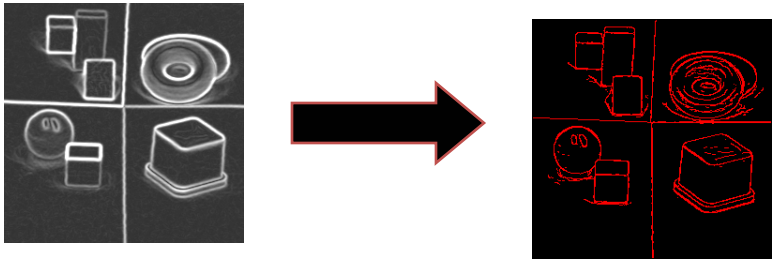




gradient computation

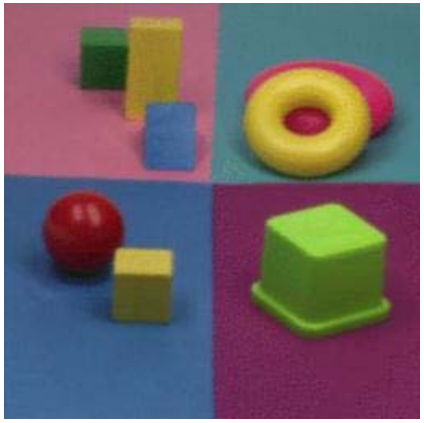
Localization - Laplacian - zero-crossing:

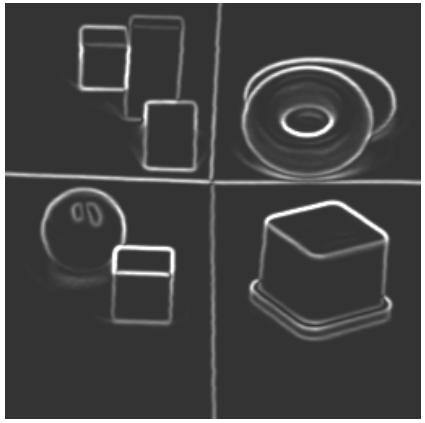
$$\nabla^2 f(x,y) = f_{xx} + f_{yy}$$



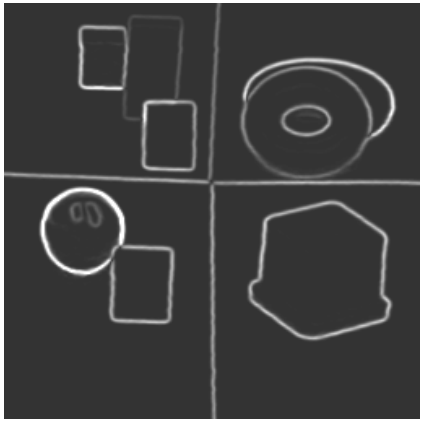
zero-crossings at large gradients

Edge Classification

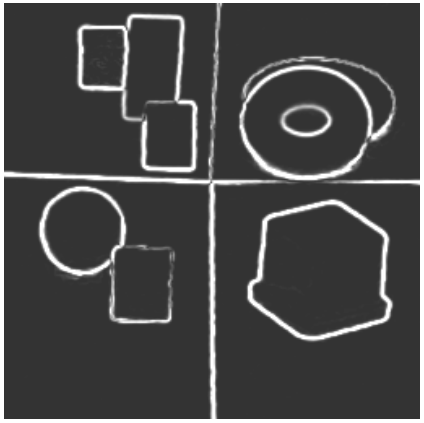




RGB



c_1, c_2, c_3



l_1, l_2, l_3

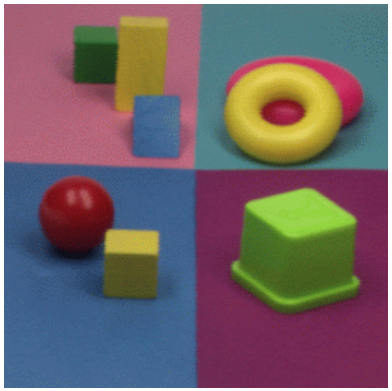
if $(|\nabla C_{c_1c_2c_3}| \geq t_{c_1c_2c_3} \,\&\, |\nabla C_{l_1l_2l_3}| < t_{l_1l_2l_3})$ then classify as highlight edge

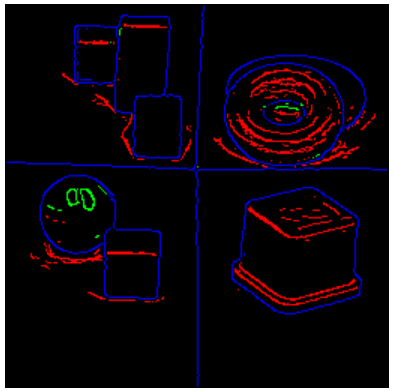
else

if $(|\nabla C_{l_1l_2l_3}| \geq t_{l_1l_2l_3})$ then classify as color edge

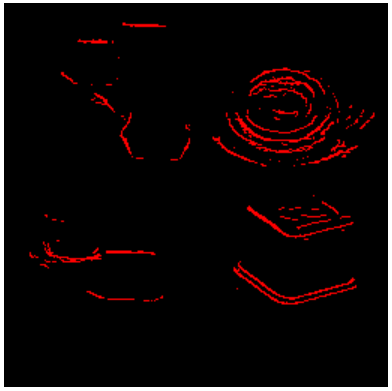
else classify as shadow/geometry edge

Edge Classification

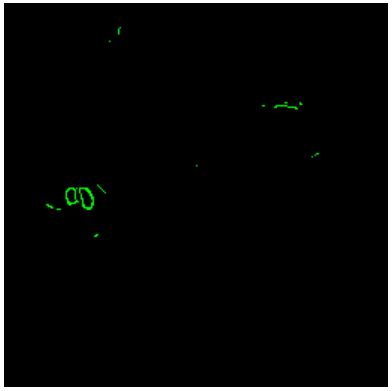




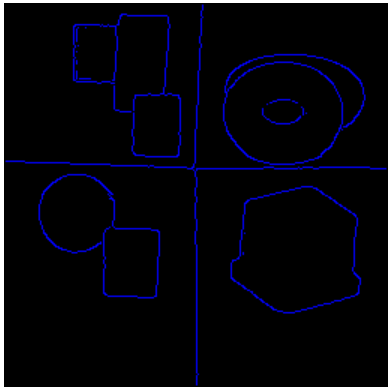
colour edge maxima by type



shadows and geometry

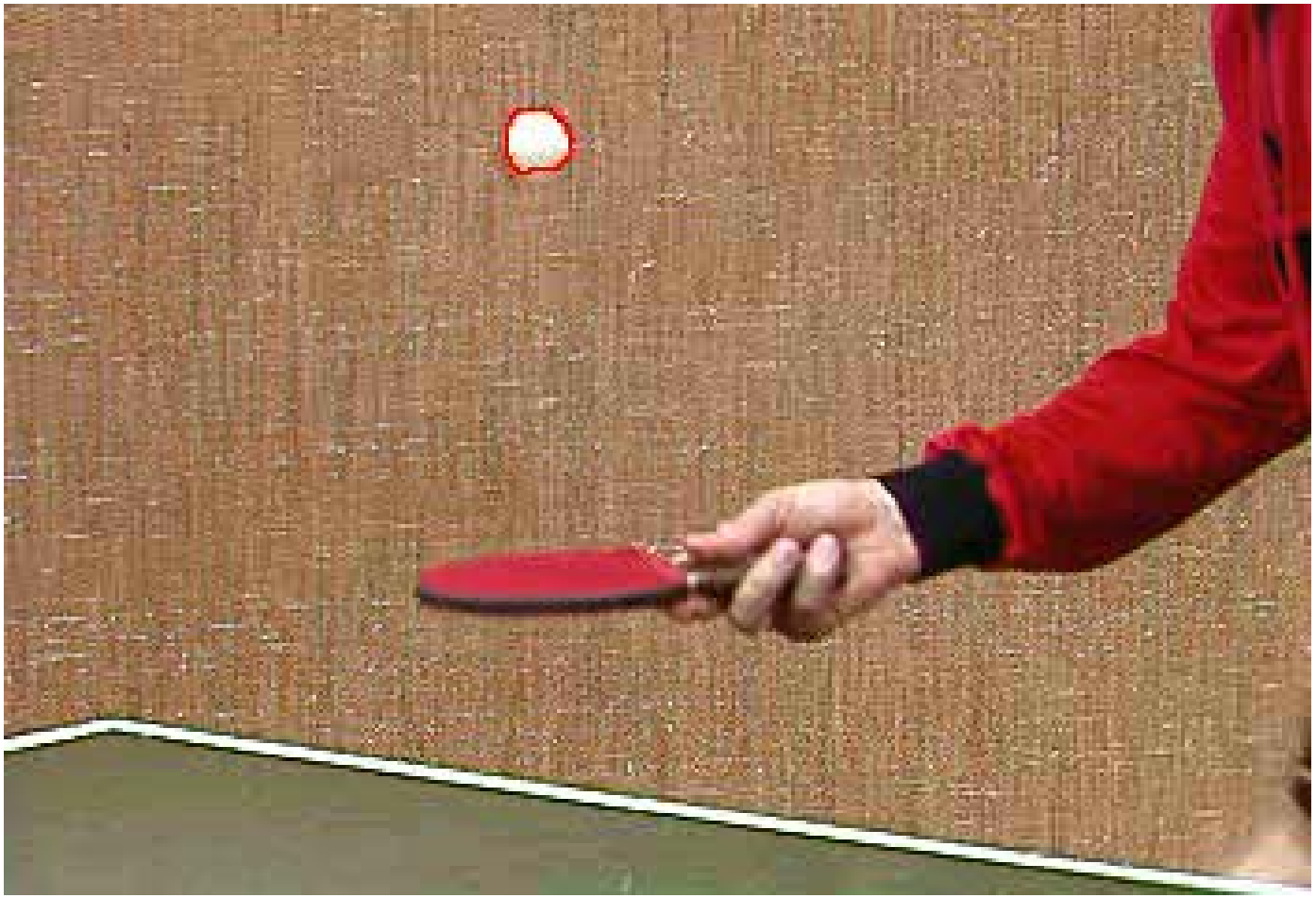


highlights



colour edges

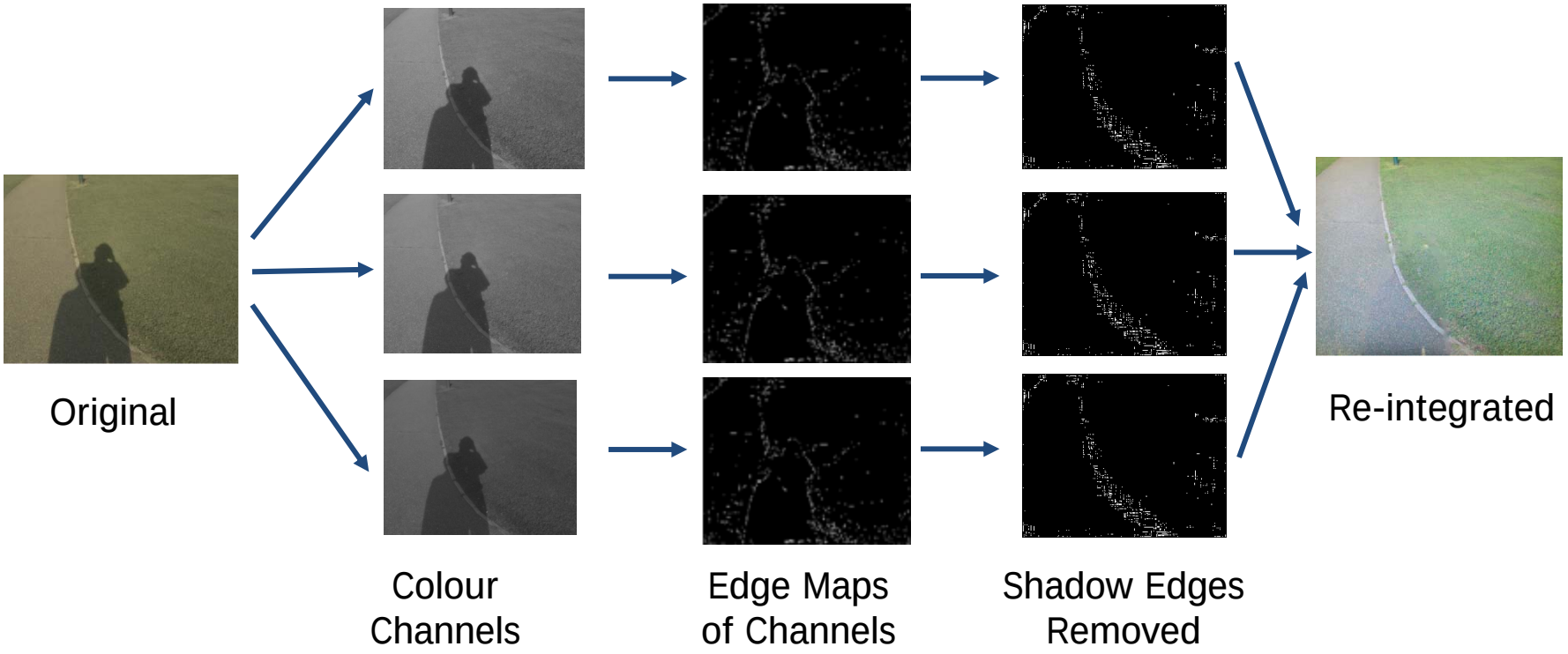
Demo: Tracking by Deformable Contours



Shadow Removal

We would like to go from a colour image with shadows, to the same colour image, but without the shadows.

Shadow Removal



An Example

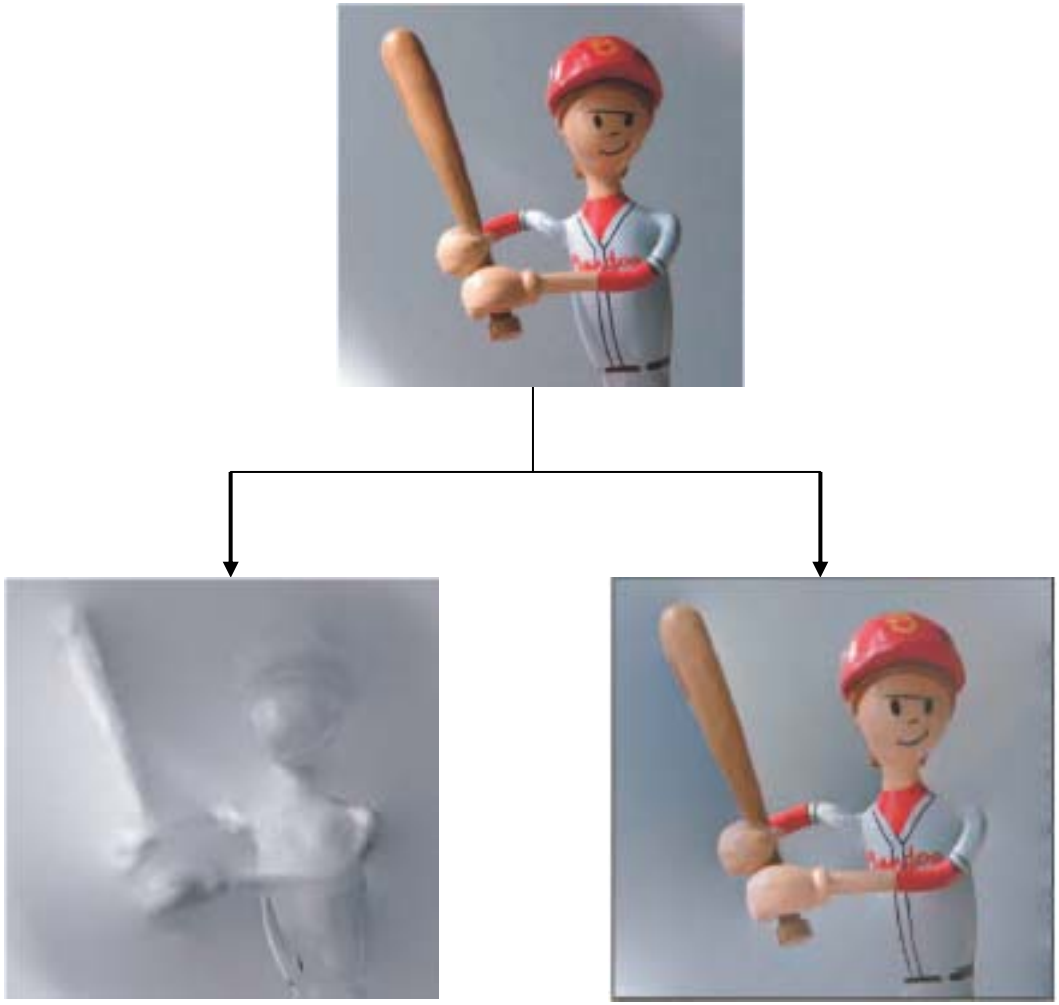
Original Image

Invariant Image

Detected Shadow Edges

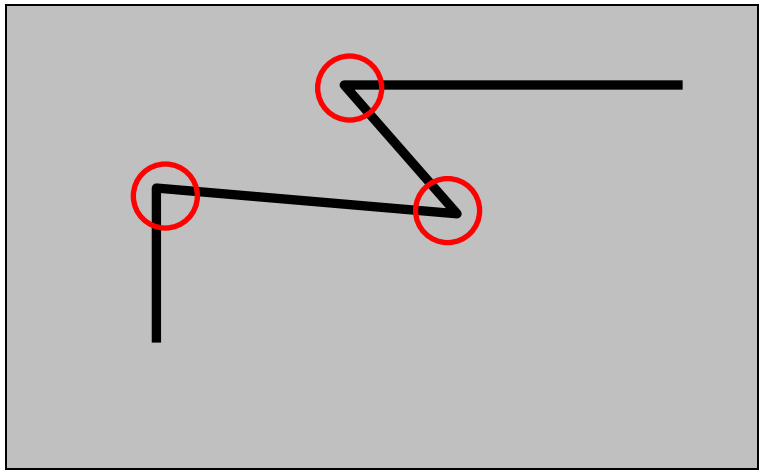
Shadow Removed

Intrinsic Images



An introductory example:

Harris corner detector

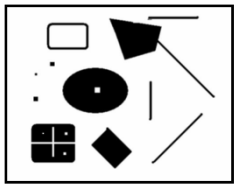


C.Harris, M.Stephens. “A Combined Corner and Edge Detector”. 1988

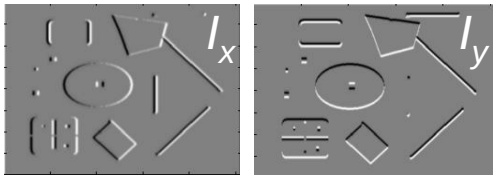
Harris Detector [Harris88]

- Second moment matrix (autocorrelation matrix)

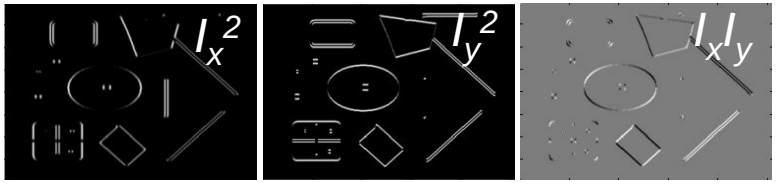
$$\mu(\sigma_I, \sigma_D) = g(\sigma_I) * \begin{bmatrix} I_x^2(\sigma_D) & I_x I_y(\sigma_D) \\ I_x I_y(\sigma_D) & I_y^2(\sigma_D) \end{bmatrix}$$



1. Image derivatives



2. Square of derivatives



3. Gaussian filter $g(\sigma_I)$

$$\det M = \lambda_1 \lambda_2$$

$$\text{trace}\, M = \lambda_1 + \lambda_2$$

4. Cornerness function – both eigenvalues are strong

$$har = \det[\mu(\sigma_I, \sigma_D)] - \alpha[\text{trace}(\mu(\sigma_I, \sigma_D))^2] =$$

$$g(I_x^2)g(I_y^2) - [g(I_x I_y)]^2 - \alpha[g(I_x^2) + g(I_y^2)]^2$$

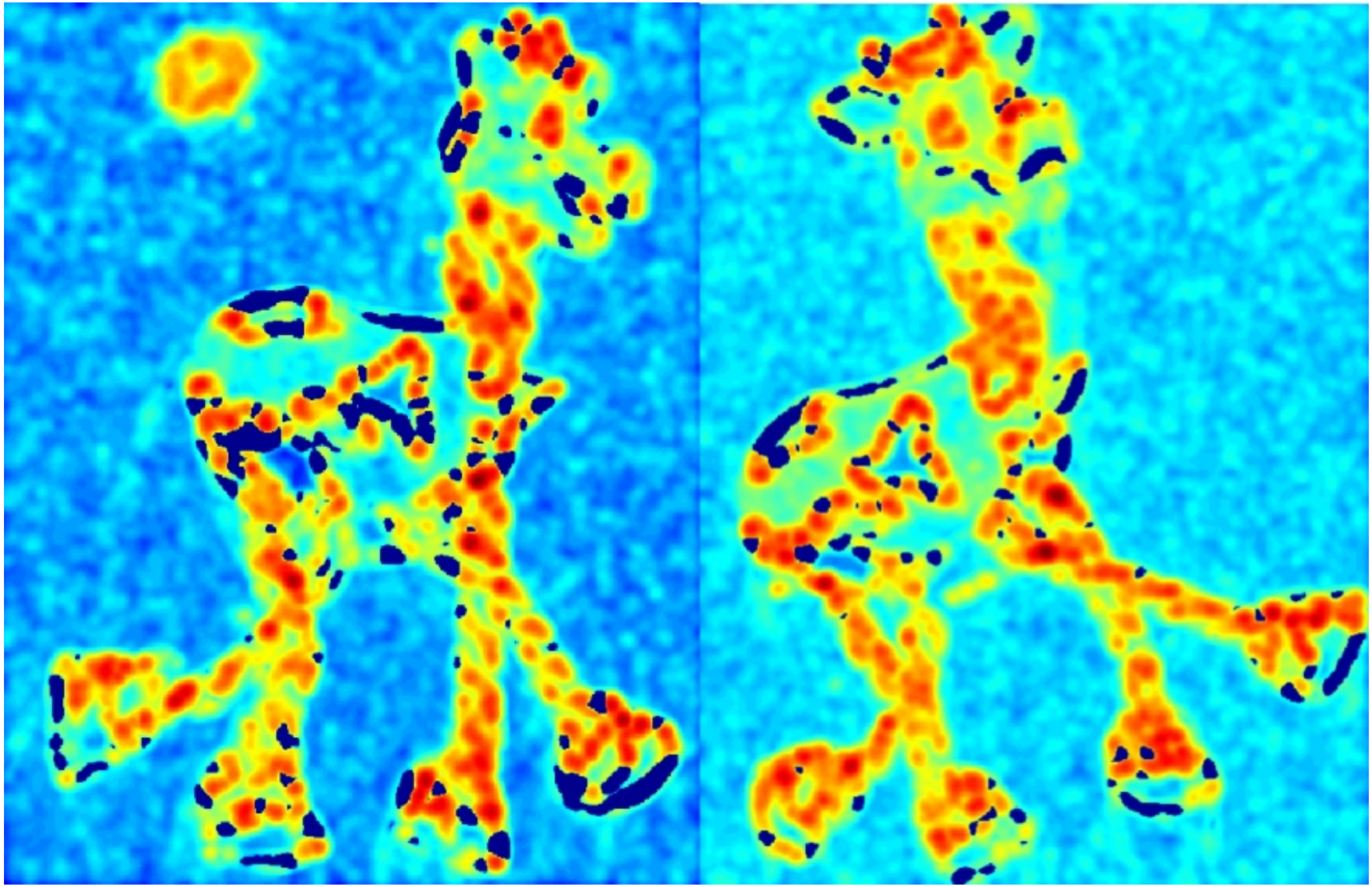
5. Non-maxima suppression

Slide: Derek Hoiem

Harris Detector: Workflow

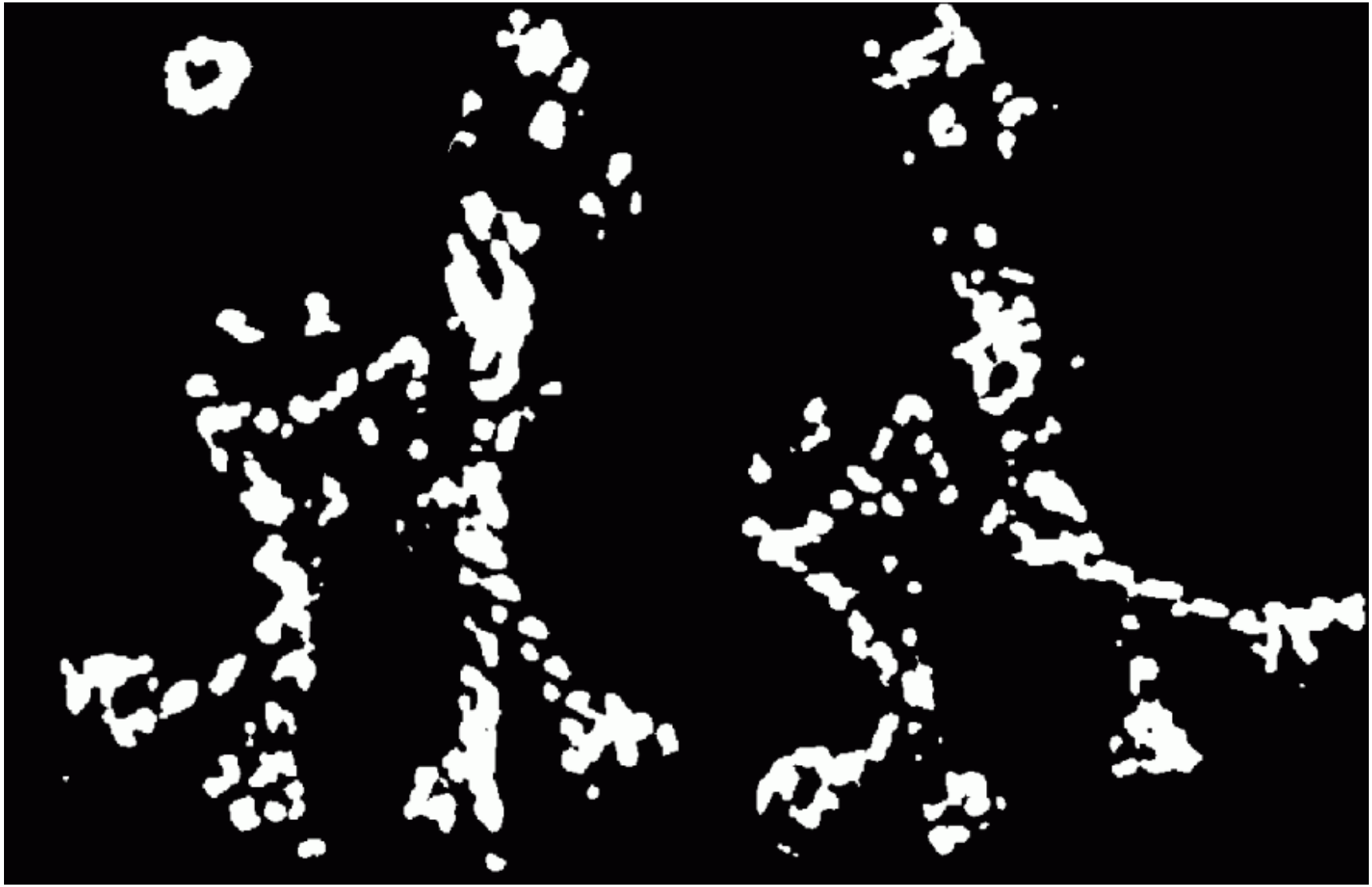
Harris Detector: Workflow

Compute corner response R



Harris Detector: Workflow

Find points with large corner response: R>threshold



Harris Detector: Workflow

Tracking

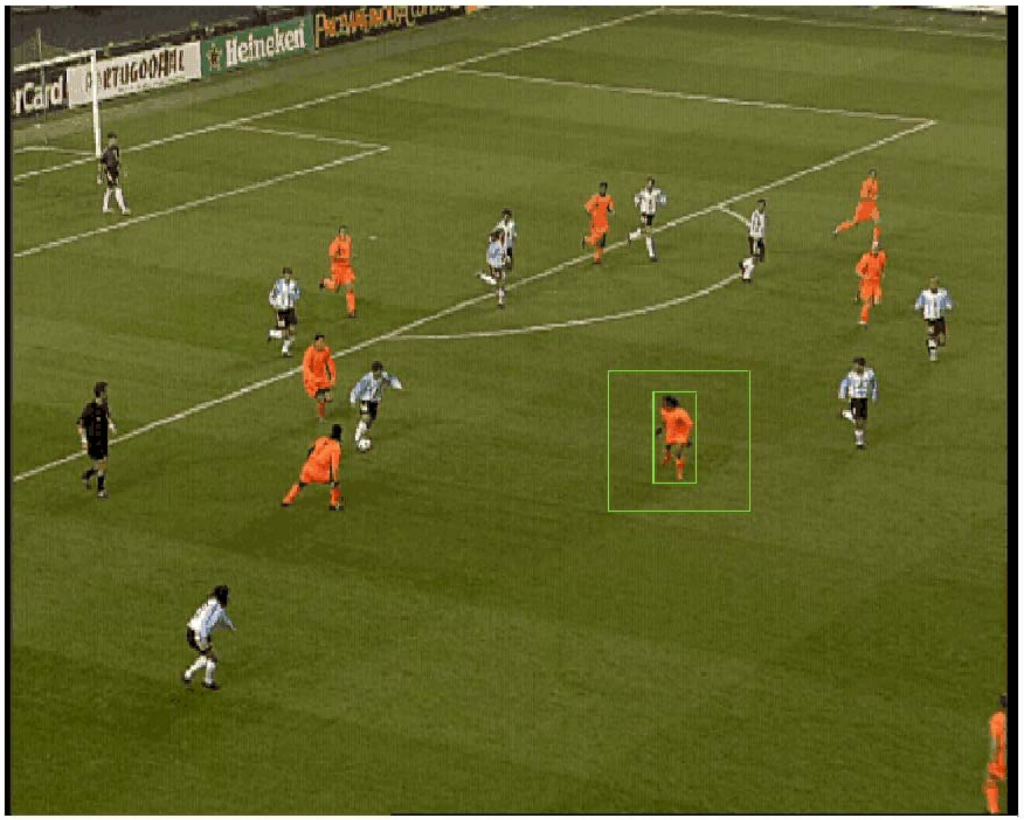
Visual tracking

Standard tracking algorithms

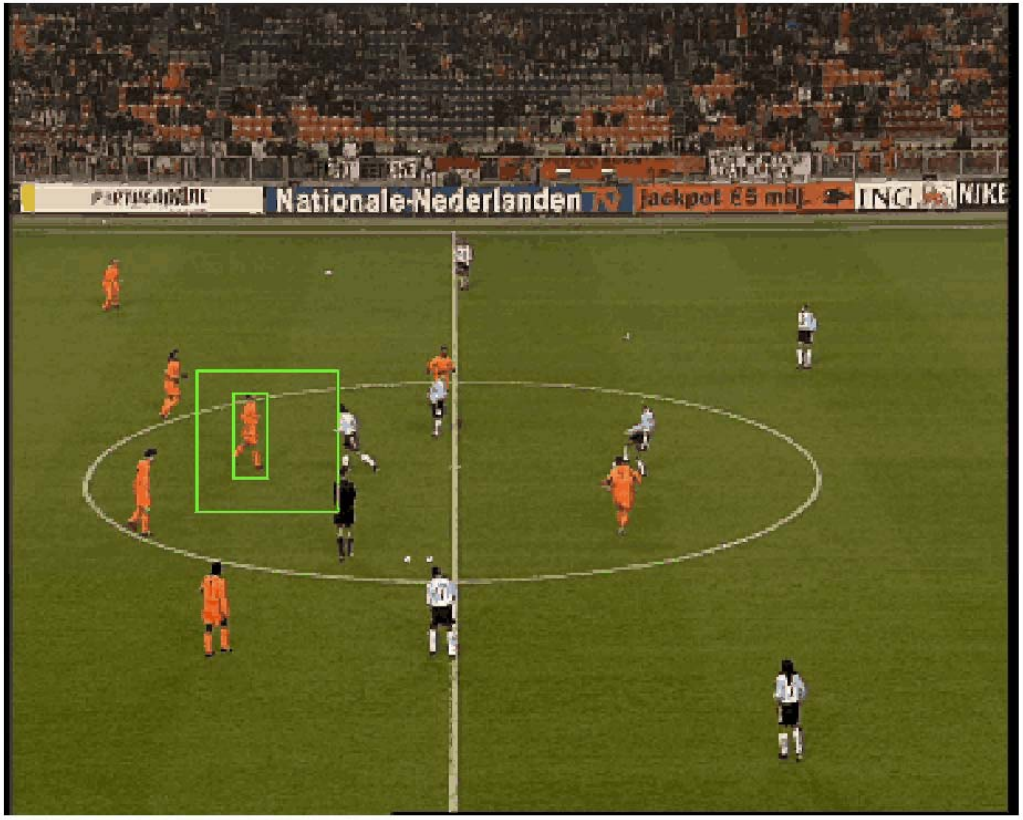
- **Motion segmentation based tracking**
- **Template tracking**
- **Mean-shift tracking**
- **Kalman filter**
- **Particle filtering**

Tracking

Tracking



Tracking



Mosaics

Object localization

Techniques:

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

Mocaics

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

Lecture 5
Learning and Object Recognition

Object/Scene Categories


Aircraft

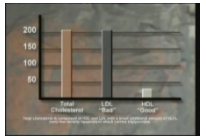

Animal


Boat


Building


Bus


Car



Chart


Corp. leader

Court


Crowd

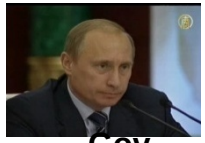

Desert


Entertainment


Explosion


Face

Flag USA


Gov. leader

Map


Meeting


Military

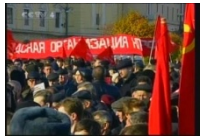

Mountain


Natural disaster


Office

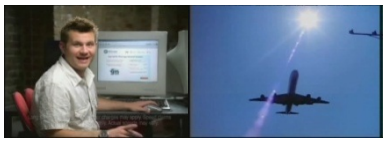

Outdoor


People


People marching


Police / security


Prisoner


Screen

Sky


Sports

Studio


Truck


Urban

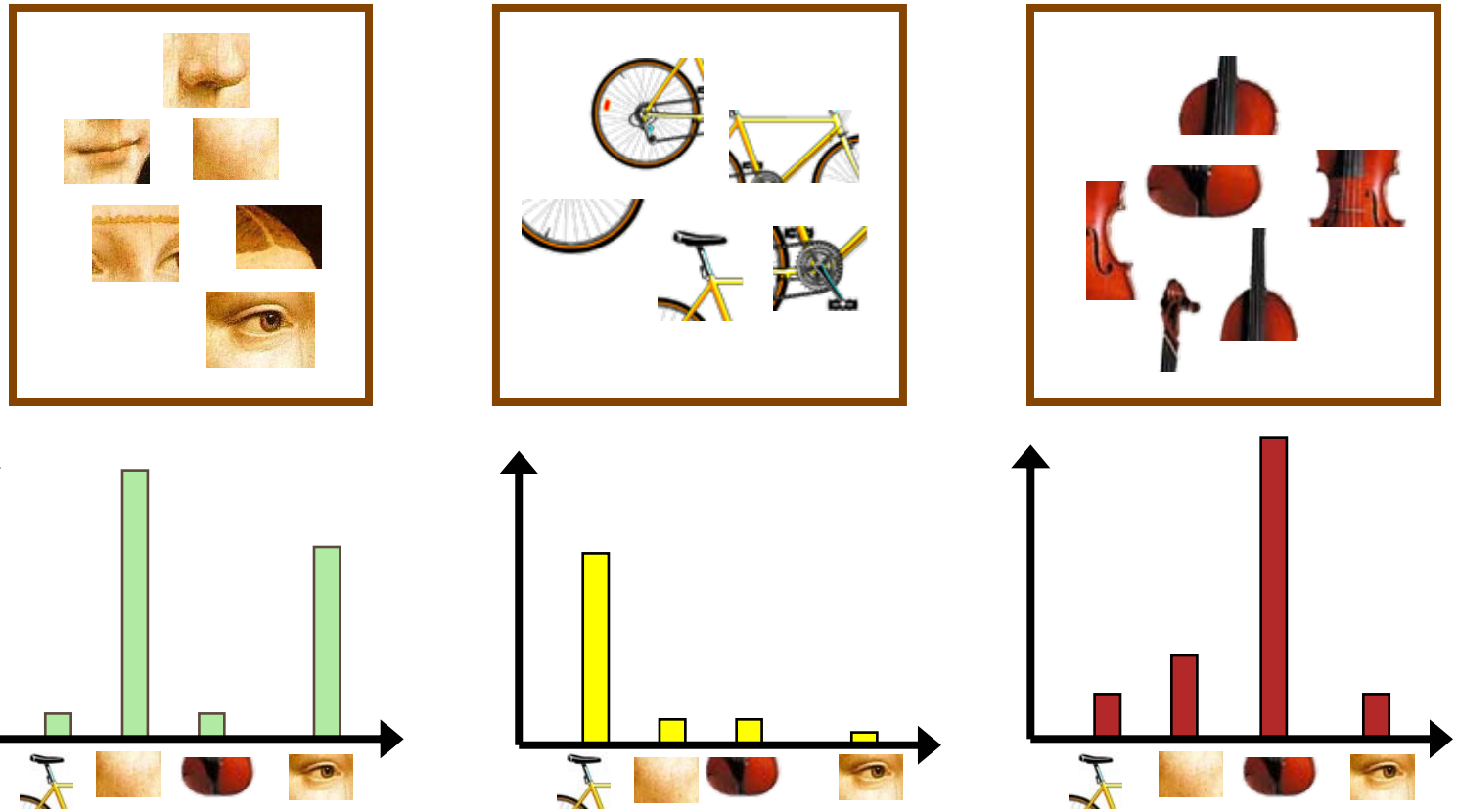

Vegetation


Vehicle

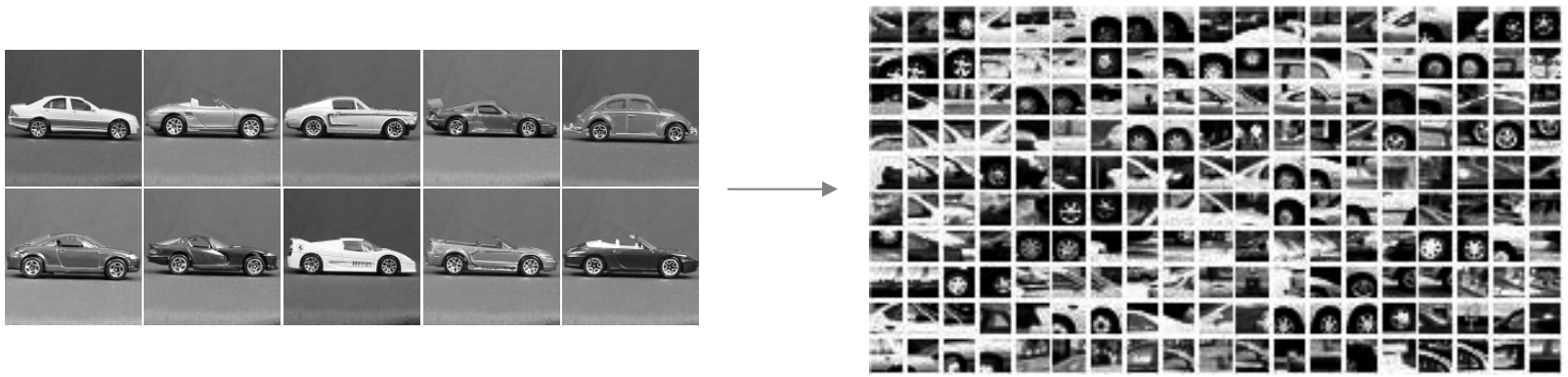

Violence

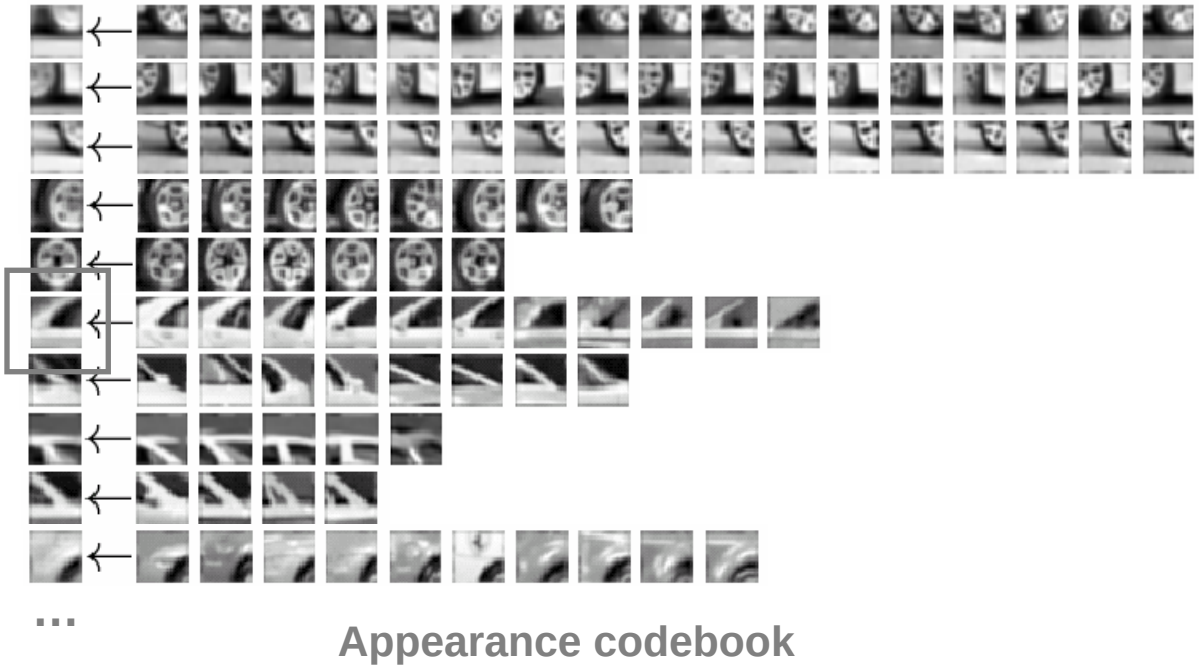
Bag-of-features Steps

1. Extract features
2. Learn “visual vocabulary”
3. Quantize features using visual vocabulary
4. Represent images by frequencies of “visual words”



Example codebook

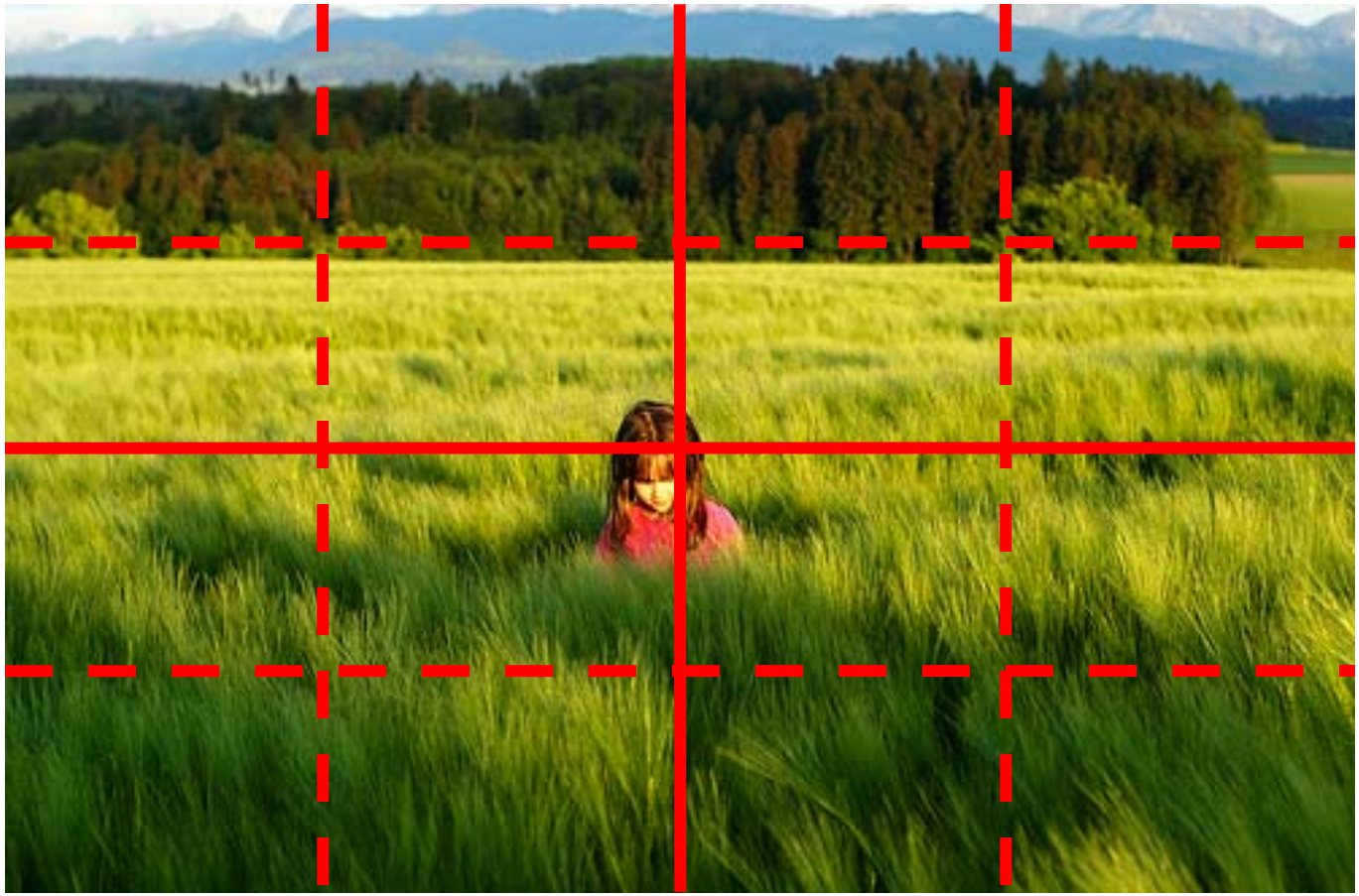




Appearance codebook

Source: B. Leibe

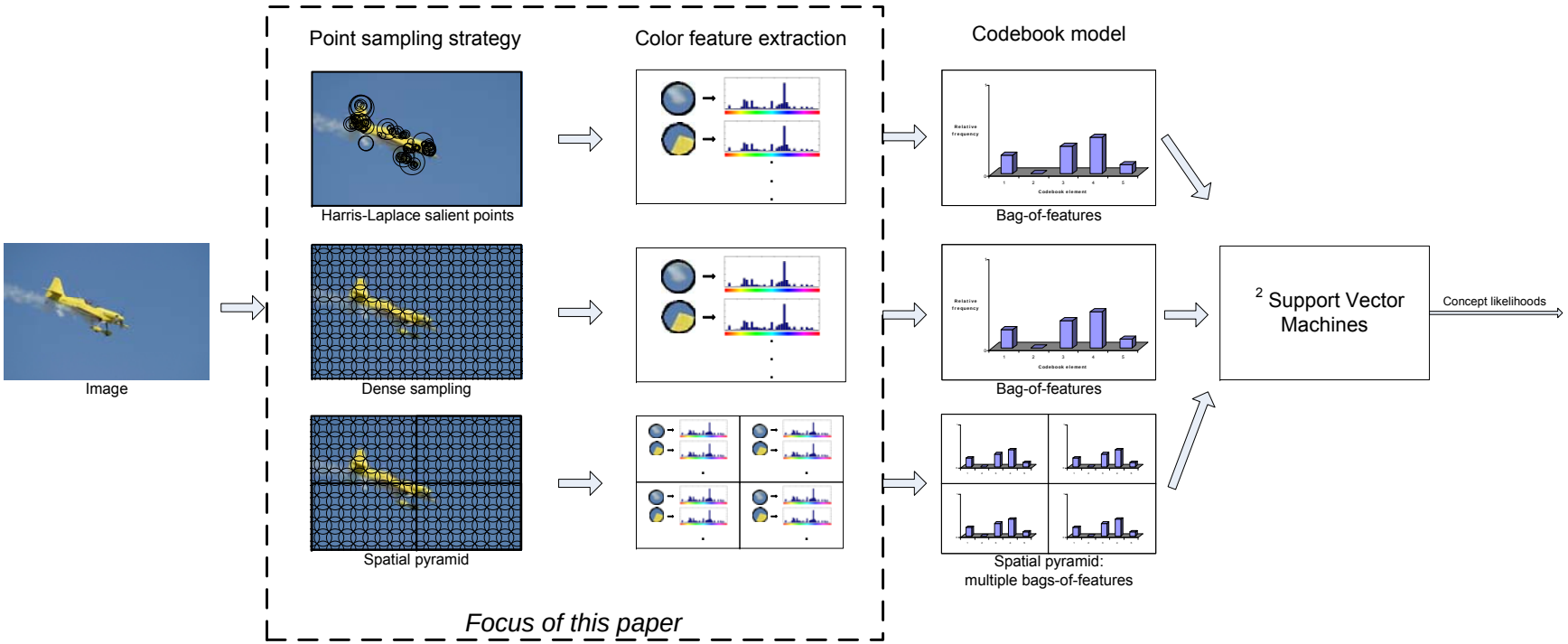
Spatial Pyramid



Compute histogram in each spatial bin

PASCAL VOC 2007/2008

Codebook size=4000



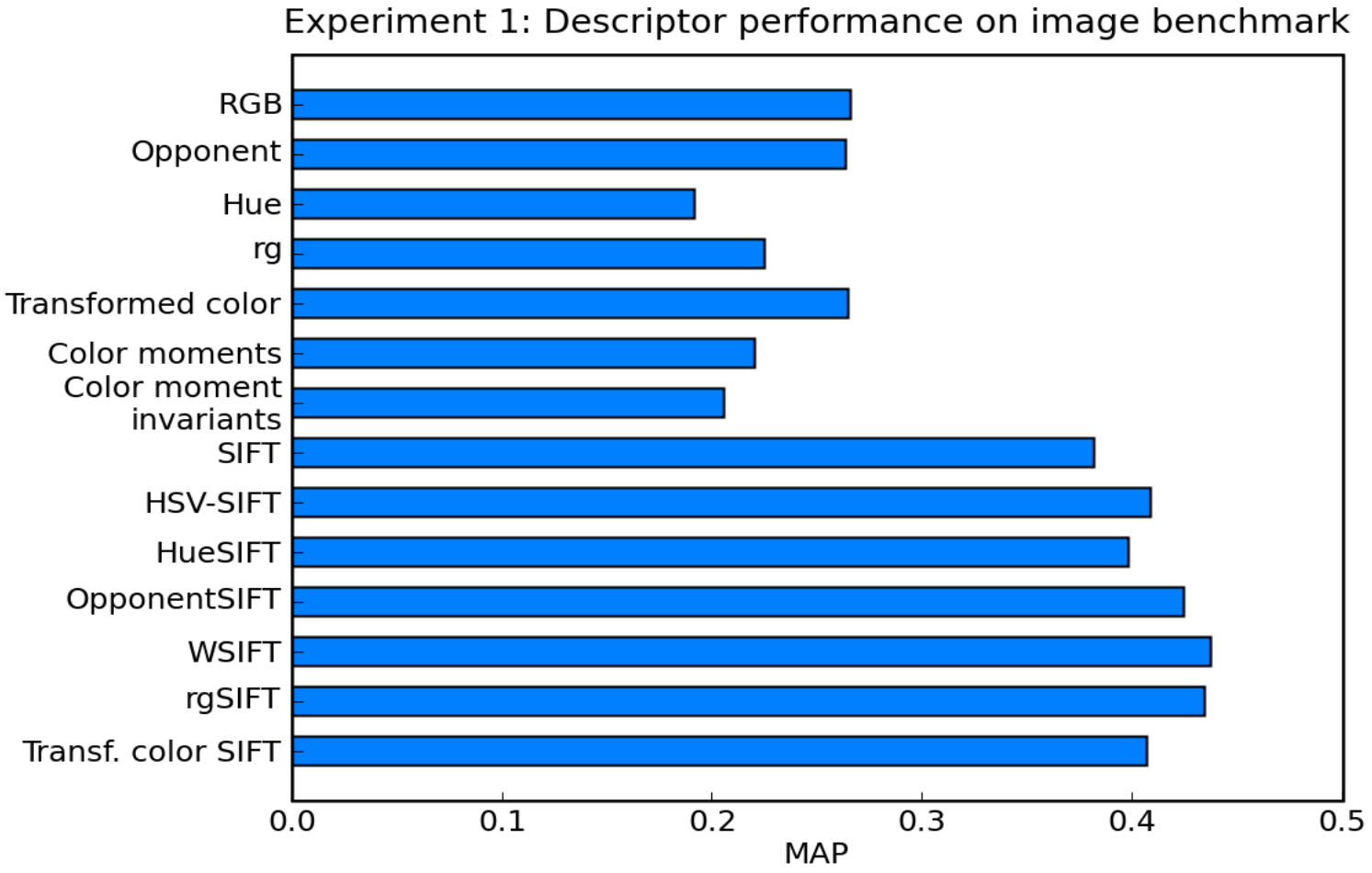
Point sampling
Harris-Laplace
Dense sampling

Spatial Pyramid
1x1
2x2
1x3

Color Descriptor
SIFT
OpponentSIFT
WSIFT
rgSIFT
Transformed color SIFT

[van de Sande, IEEE PAMI, 09]

Results on PASCAL VOC 2007



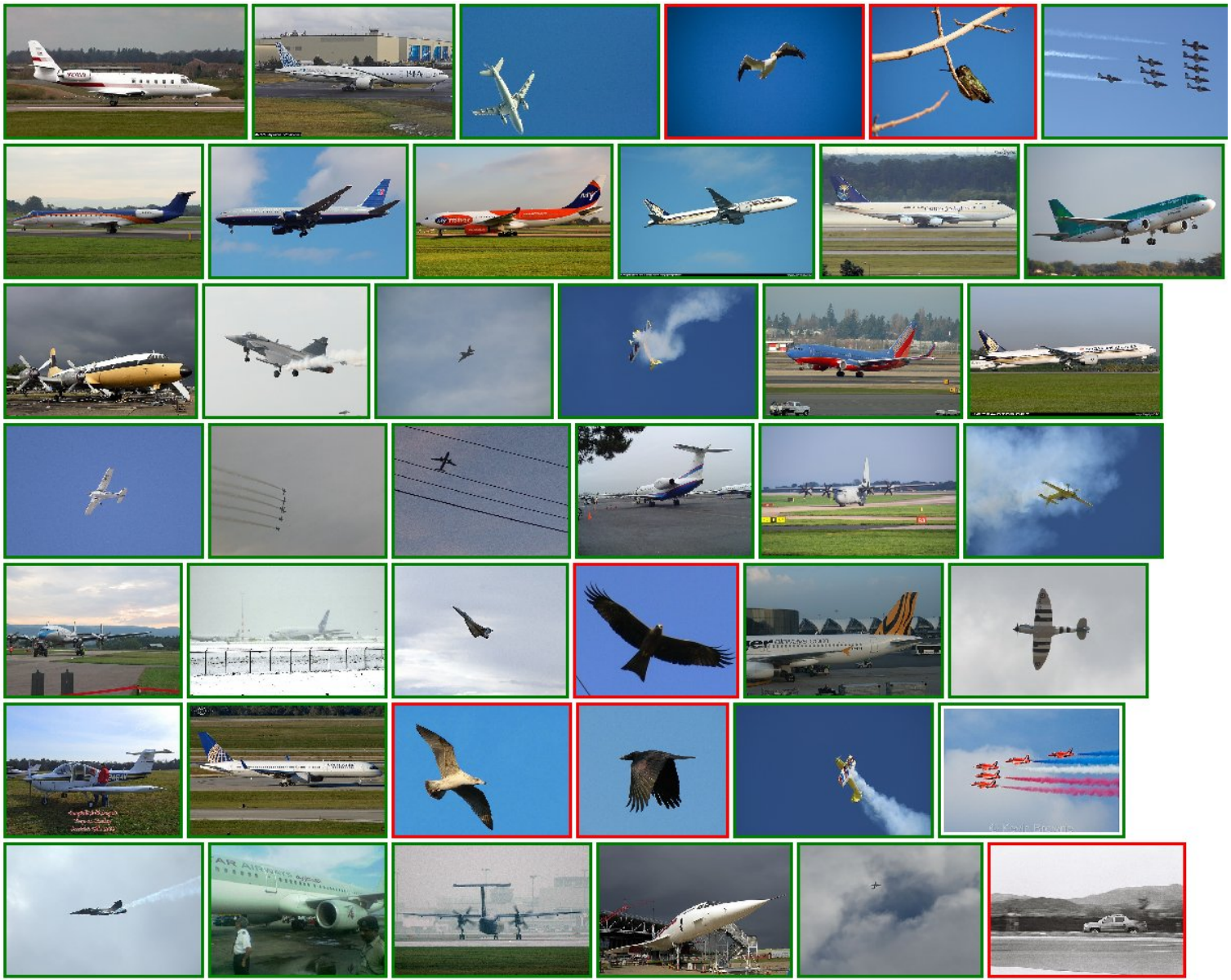
Color Descriptor Taxonomy

[van de Sande, IEEE PAMI, 09]

	Light intensity change $\begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$	Light intensity shift $\begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} o_1 \\ o_1 \\ o_1 \end{pmatrix}$	Light intensity change and shift $\begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} o_1 \\ o_1 \\ o_1 \end{pmatrix}$	Light color change $\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$	Light color change and shift $\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} o_1 \\ o_2 \\ o_3 \end{pmatrix}$
RGB Histogram	-	-	-	-	-
O_1, O_2	-	+	-	-	-
O_3, Intensity	-	-	-	-	-
Hue	+	+	+	-	-
Saturation	+	+	+	-	-
r, g	+	-	-	-	-
Transformed color	+	+	+	+	+
Color moments	-	+	-	-	-
Moment invariants	+	+	+	+	+
SIFT (∇I)	+	+	+	+	+
HSV-SIFT	+	+	+	+/-	+/-
HueSIFT	+	+	+	+/-	+/-
OpponentSIFT	+/-	+	+/-	+/-	+/-
W-SIFT	+	+	+	+/-	+/-
rgSIFT	+	+	+	+/-	+/-
Transf. color SIFT	+	+	+	+	+

VOC2007 results –airplanes (1)

VOC2007 results – airplanes (2)



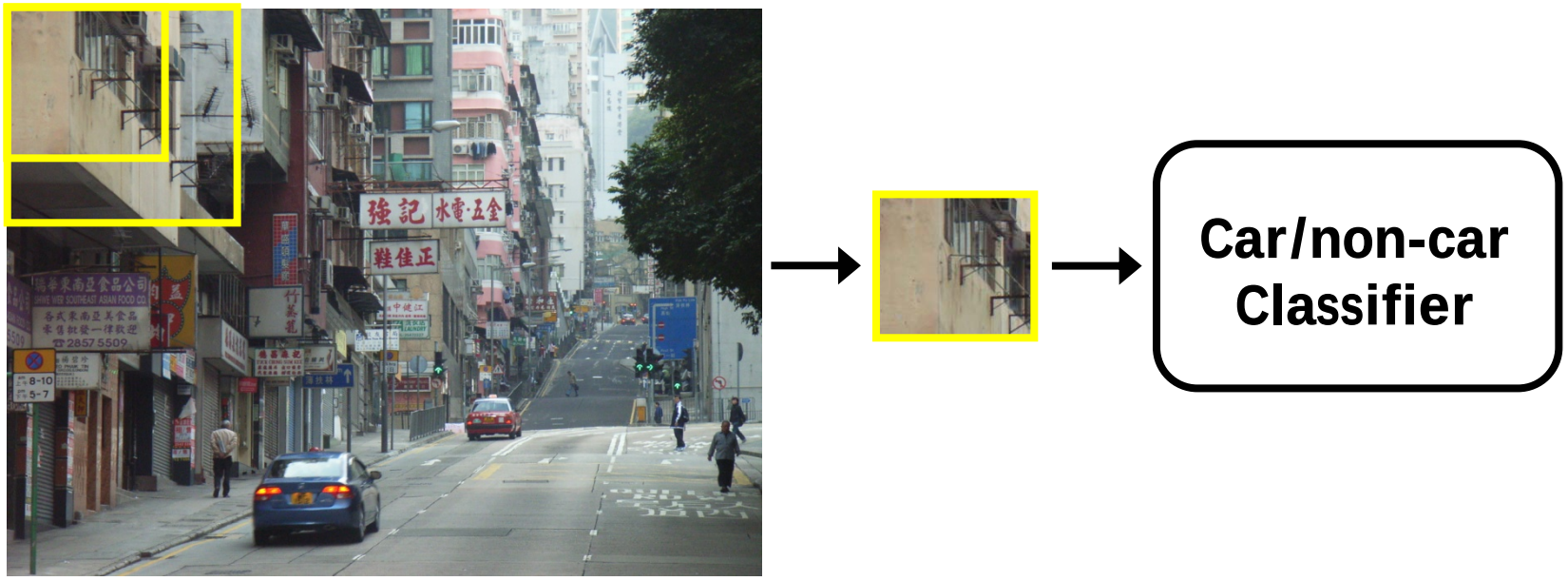
VOC2007 results –persons (1)

VOC2007 results – persons (2)

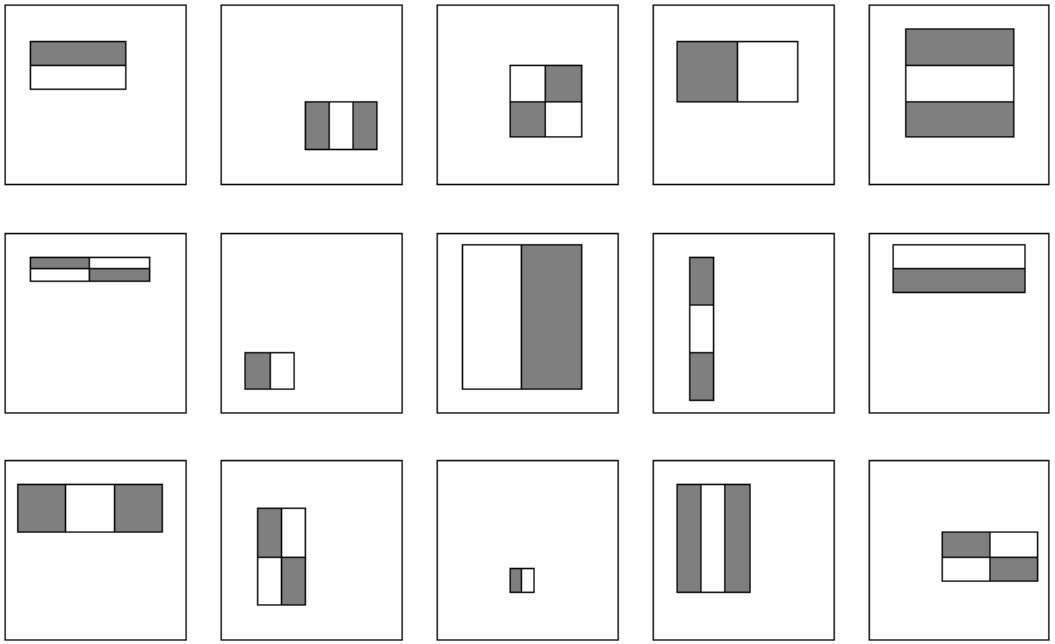
Lecture 6
Visual Attention and Affective Computing

Window-based models

Generating and scoring candidates



Viola-Jones Detector: Features



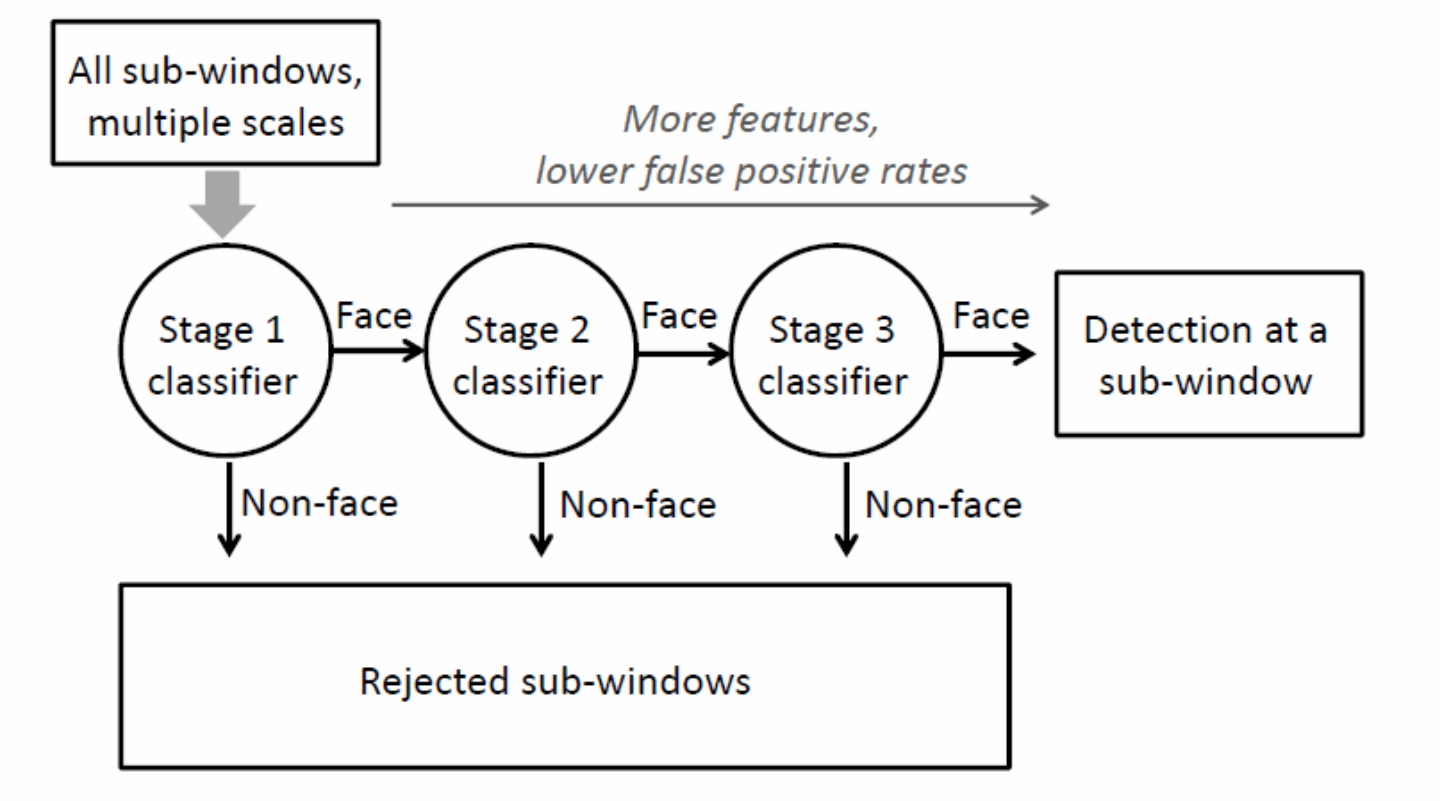
Considering all possible filter parameters: position, scale, and type:

180,000+ possible features associated with each 24 x 24 window

Which subset of these features should we use to determine if a window has a face?

Use AdaBoost both to select the informative features and to form the classifier

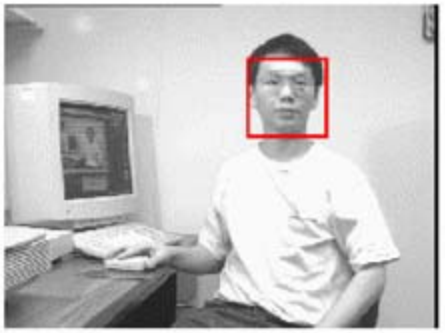
Cascading Classifiers for Detection

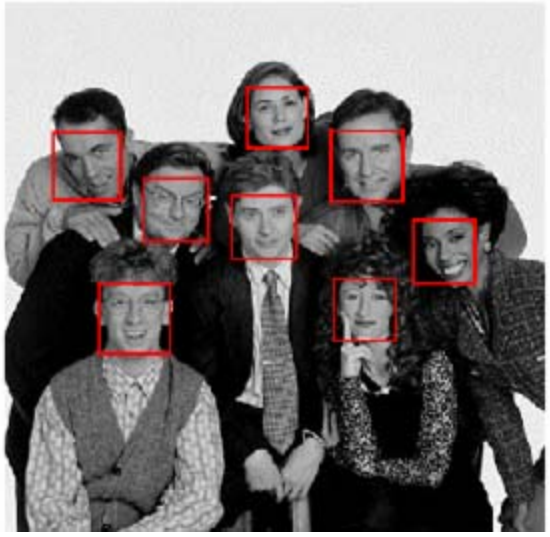


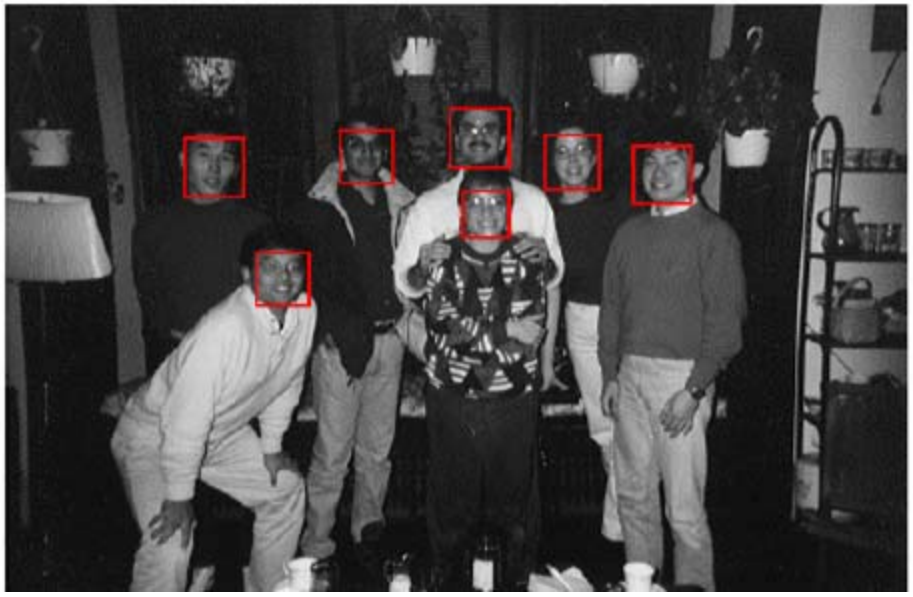
- Form a *cascade* with low false negative rates early on
- Apply less accurate but faster classifiers first to immediately discard windows that clearly appear to be negative

Kristen Grauman

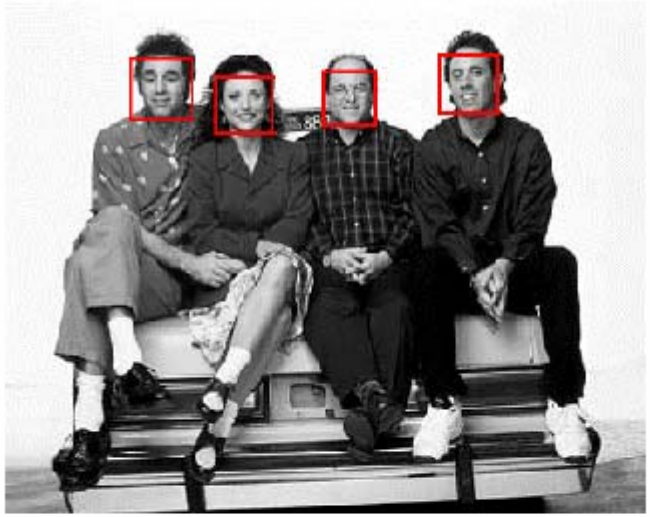
Viola-Jones Face Detector: Results

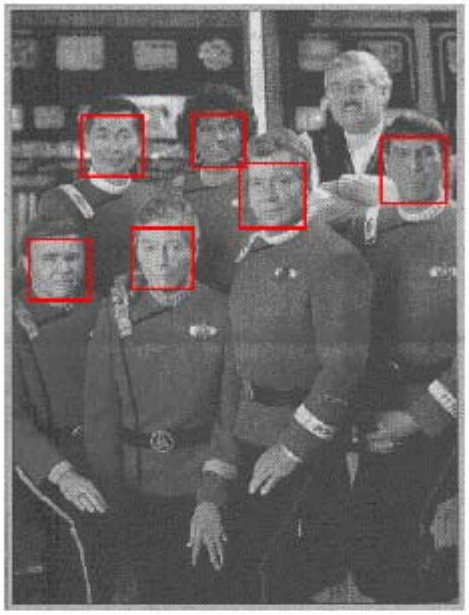


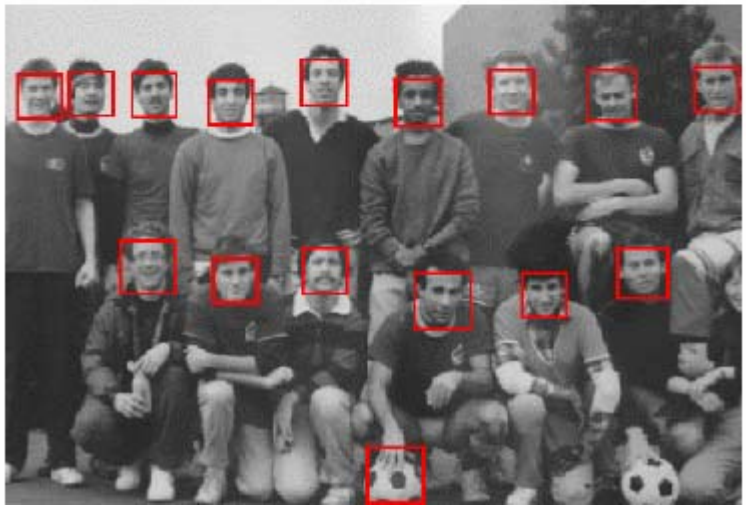


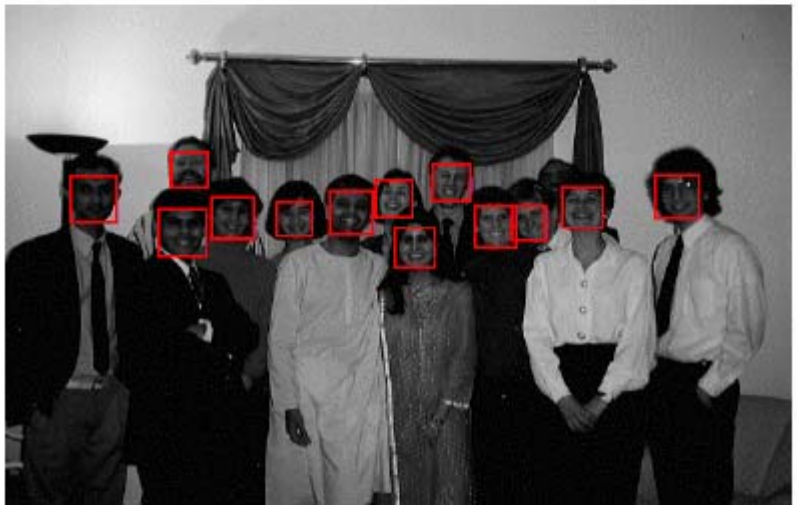


Viola-Jones Face Detector: Results



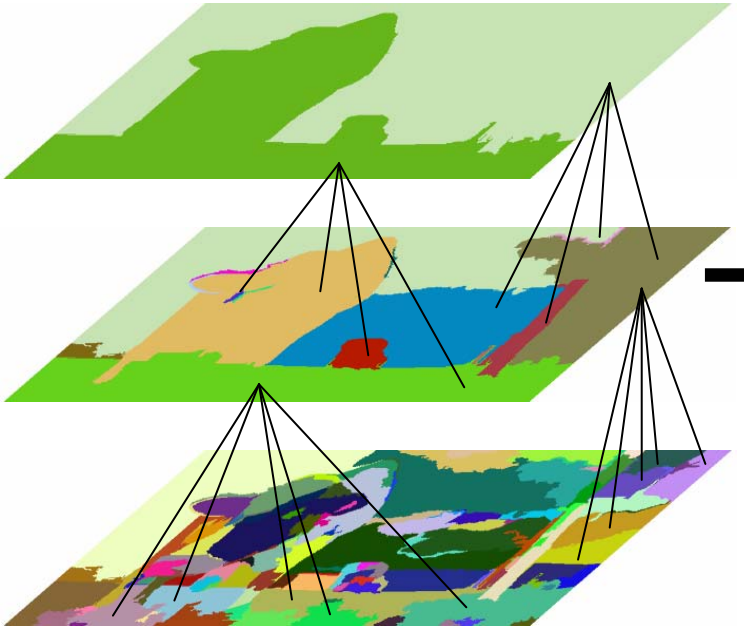


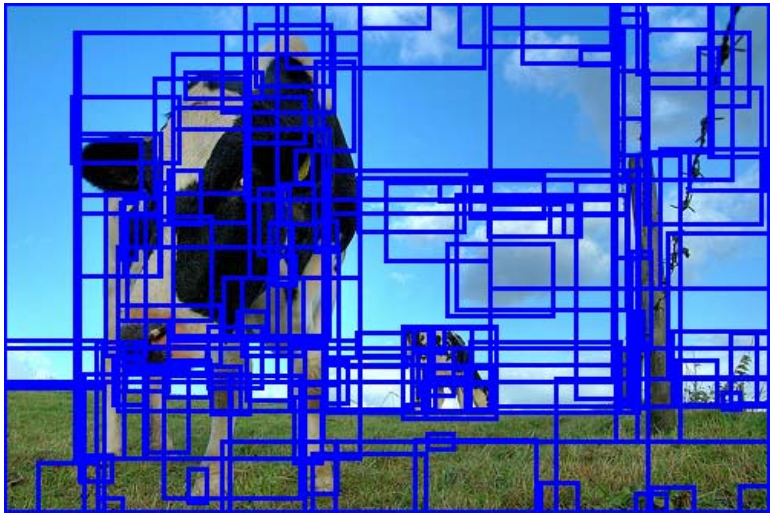




Selective Search: Approach

- Hypotheses based on hierarchical grouping

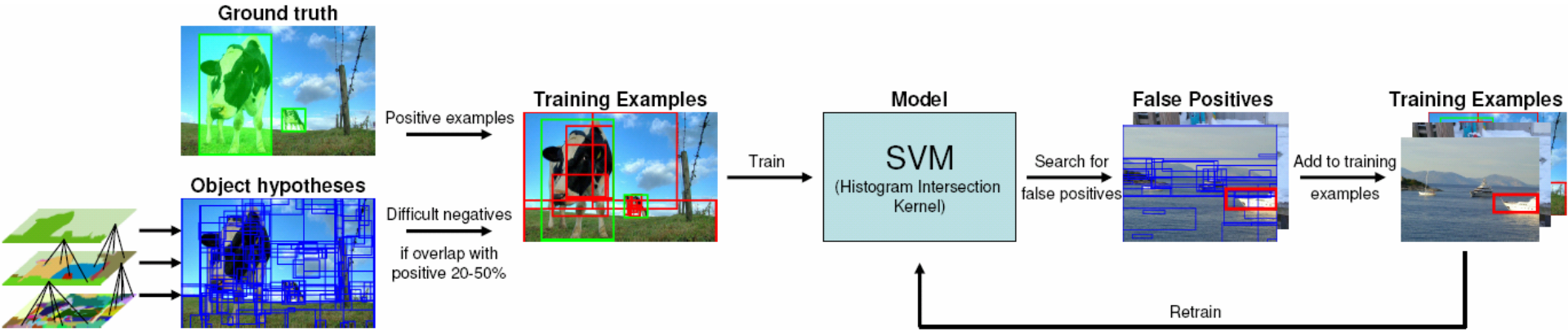




Object hypotheses from
all hierarchy levels

Localisation System Training

- Use positives and mirrored positives
- Use object hypotheses to create difficult initial negatives (at most 7,500)
- Add 2 iterations of false positives (from 4,000 images)



- Features: Bag-of-words, sample every pixel, SIFT, “ColorSIFT” and RGB-SIFT, pyramid up to level 3, codebook size 4096
- Histogram Intersection Kernel with Fast Approximation

Facial emotion recognition

Facial Expression Recognition

- Face model: 16 surface patches embedded in Bezier volumes.
- Piecewise Bezier Volume Deformation (PBVD) tracker is used to trace the motion of the facial features.

Temporal Phase Segmentation

- A facial expression is composed of three main phases:
  - **Onset**: Neutral state to expressive face
  - **Apex**: Stable period of the expressive face
  - **Offset**: Expressive state to neutral face

Definitions of Dynamic Features

Feature	Definition
Duration	$\left[\frac{\eta(\mathcal{D}^+)}{\omega}, \frac{\eta(\mathcal{D}^-)}{\omega}, \frac{\eta(\mathcal{D})}{\omega}\right]$
Duration Ratio	$\left[\frac{\eta(\mathcal{D}^+)}{\eta(\mathcal{D})}, \frac{\eta(\mathcal{D}^-)}{\eta(\mathcal{D})}\right]$
Maximum Amplitude	$\max(\mathcal{D})$
Mean Amplitude	$\left[\frac{\sum \mathcal{D}}{\eta(\mathcal{D})}, \frac{\sum \mathcal{D}^+}{\eta(\mathcal{D}^+)}, \frac{\sum \lvert\mathcal{D}^-\rvert}{\eta(\mathcal{D}^-)}\right]$
STD of Amplitude	$\mathrm{std}(\mathcal{D})$
Total Amplitude	$\left[\sum \mathcal{D}^+, \sum \lvert\mathcal{D}^-\rvert\right]$
Net Amplitude	$\sum \mathcal{D}^+ - \sum \lvert\mathcal{D}^-\rvert$
Amplitude Ratio	$\left[\frac{\sum \mathcal{D}^+}{\sum \mathcal{D}^+ + \sum \lvert\mathcal{D}^-\rvert}, \frac{\sum \lvert\mathcal{D}^-\rvert}{\sum \mathcal{D}^+ + \sum \lvert\mathcal{D}^-\rvert}\right]$
Maximum Speed	$[\max(\mathcal{V}^+), \max(\lvert\mathcal{V}^-\rvert)]$
Mean Speed	$\left[\frac{\sum \mathcal{V}^+}{\eta(\mathcal{V}^+)}, \frac{\sum \lvert\mathcal{V}^-\rvert}{\eta(\mathcal{V}^-)}\right]$
Maximum Acceleration	$[\max(\mathcal{A}^+), \max(\lvert\mathcal{A}^-\rvert)]$
Mean Acceleration	$\left[\frac{\sum \mathcal{A}^+}{\eta(\mathcal{A}^+)}, \frac{\sum \lvert\mathcal{A}^-\rvert}{\eta(\mathcal{A}^-)}\right]$
Net Ampl., Duration Ratio	$\frac{(\sum \mathcal{D}^+ - \sum \lvert\mathcal{D}^-\rvert)\omega}{\eta(\mathcal{D})}$
Left/Right Ampl. Difference	$\frac{\lvert\sum \mathcal{D}_L - \sum \mathcal{D}_R\rvert}{\eta(\mathcal{D})}$

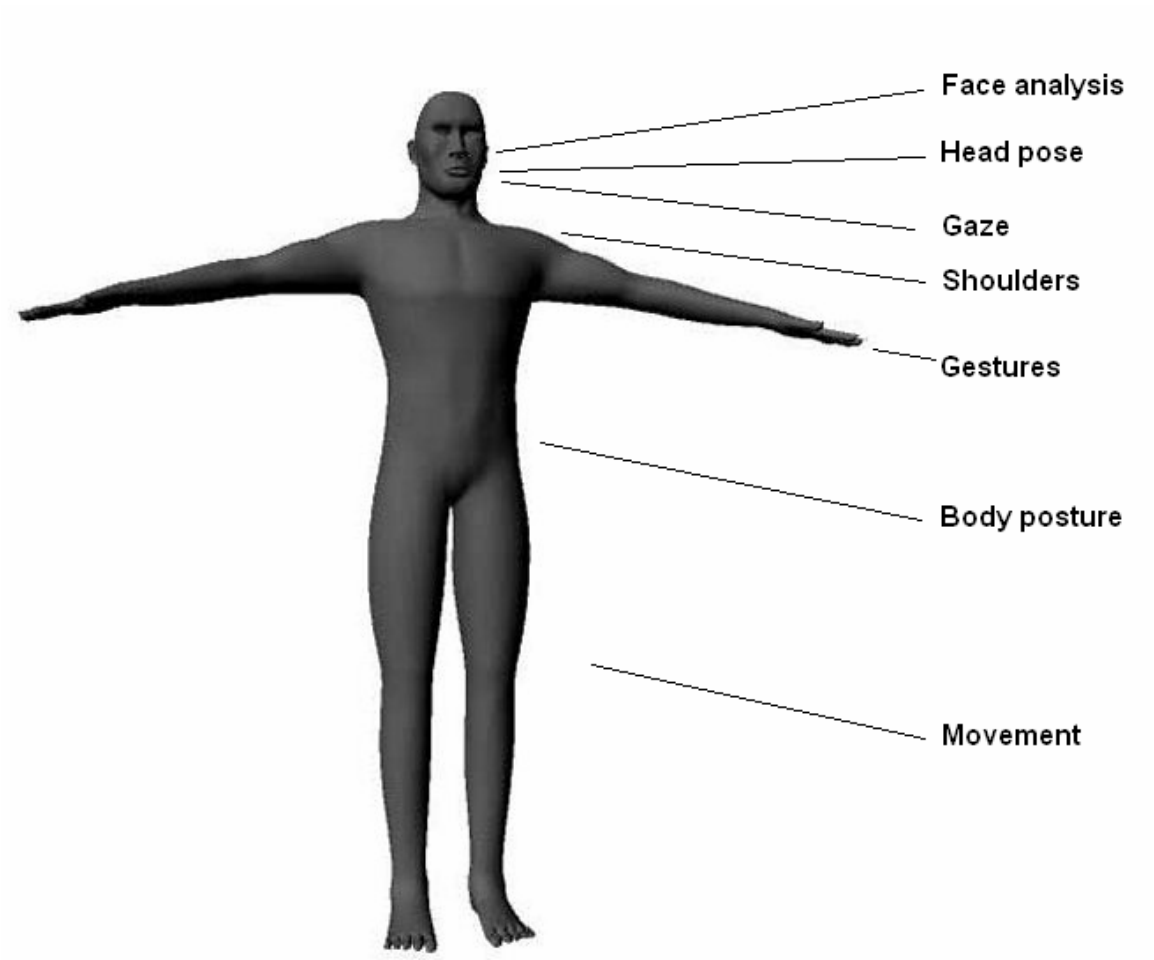
- $\mathcal{D}$: Amplitude signal
- $\mathcal{V}$: Speed signal
- $\mathcal{A}$: Acceleration signal
- η: Signal length
- ω: Frame rate

- 25-dimensional feature vectors are extracted for each facial region on onset, apex, and offset phases, separately.

Human behaviour analysis

Activity Recognition

Visual analysis of the human body



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 Tracking
- 26-11-2011, Monday, 15:00-17:00, Science Park A1.04 - Learning and Object Recognition
- 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

The end