Intelligent Multimedia Systems

Master AI, 2012, Lecture 6

Lecturers: Theo Gevers

Lab: Intelligent Systems Lab Amsterdam (ISLA)

Email: th.gevers@uva.nl

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

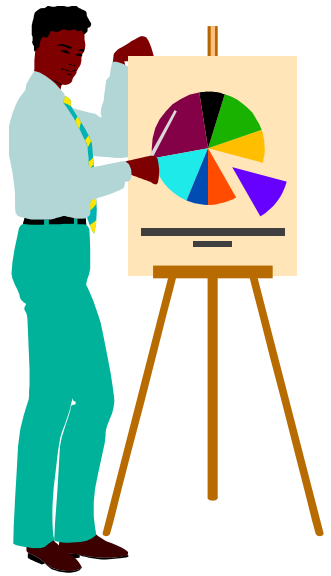
Lectures

- 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

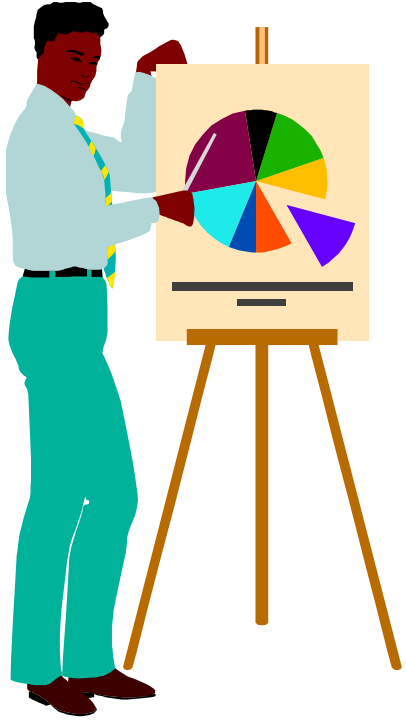
Today's class

Part I: Visual Attention: Object Localization

Part II: Affective Computing (Hamdi Dibeklioglu)



Today's class



PART I (before the break)

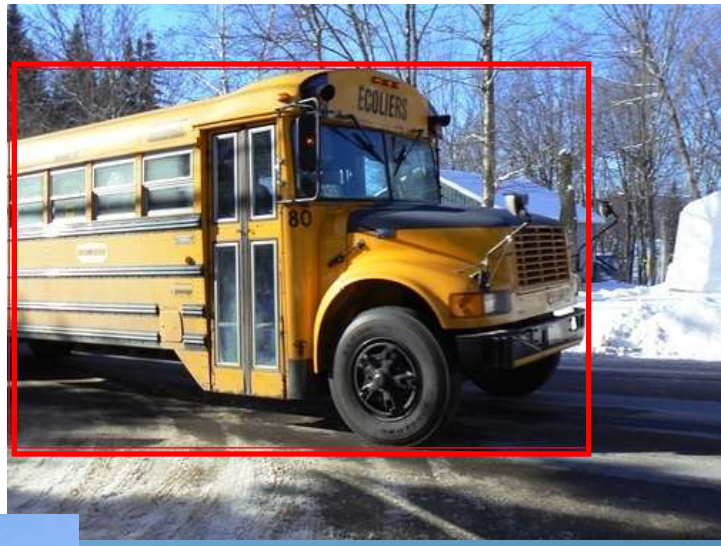
Sliding Window Approach

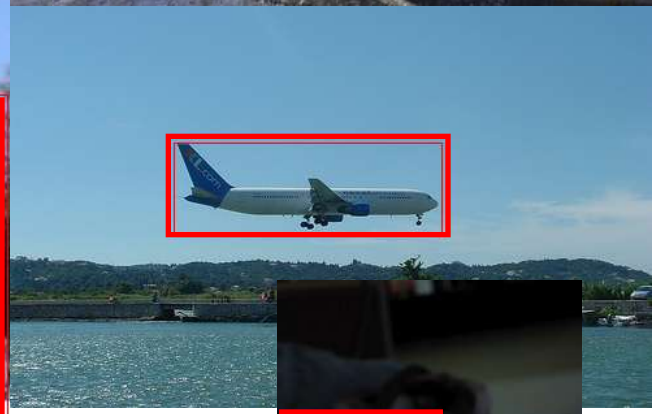
Sliding Windows for Face Detection

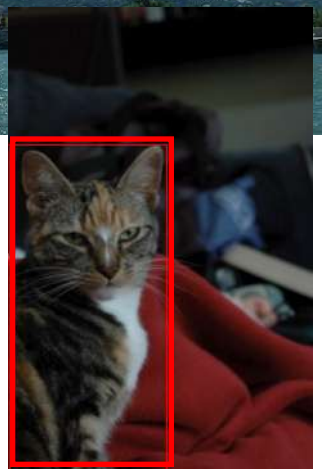
Image Segmentation for Object Localization

Object localization

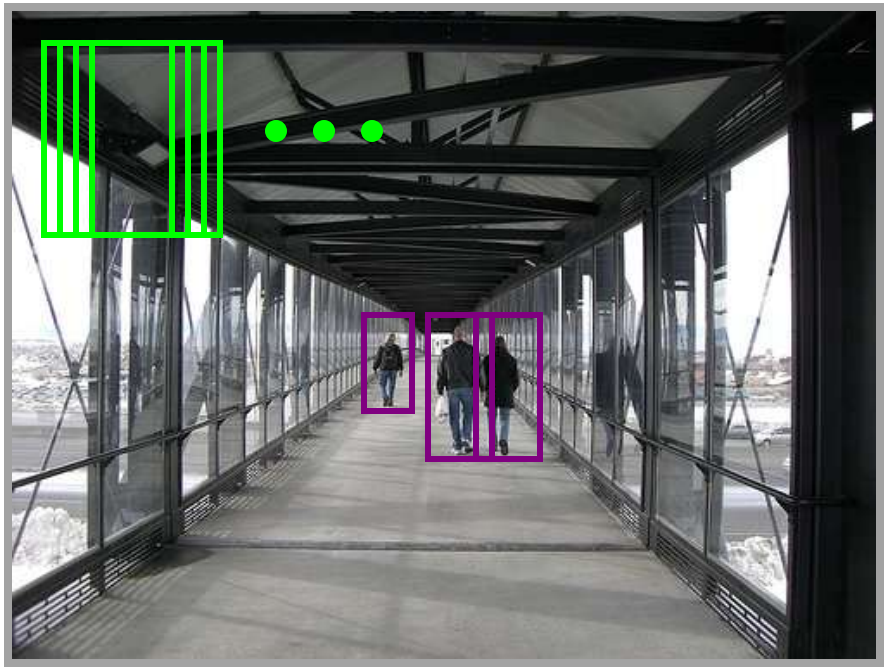
– Where is the object located?







Detection: Is this an X?



Boosted dec. trees, cascades
+ Very fast evaluation
- Slow training (esp. multi-class)

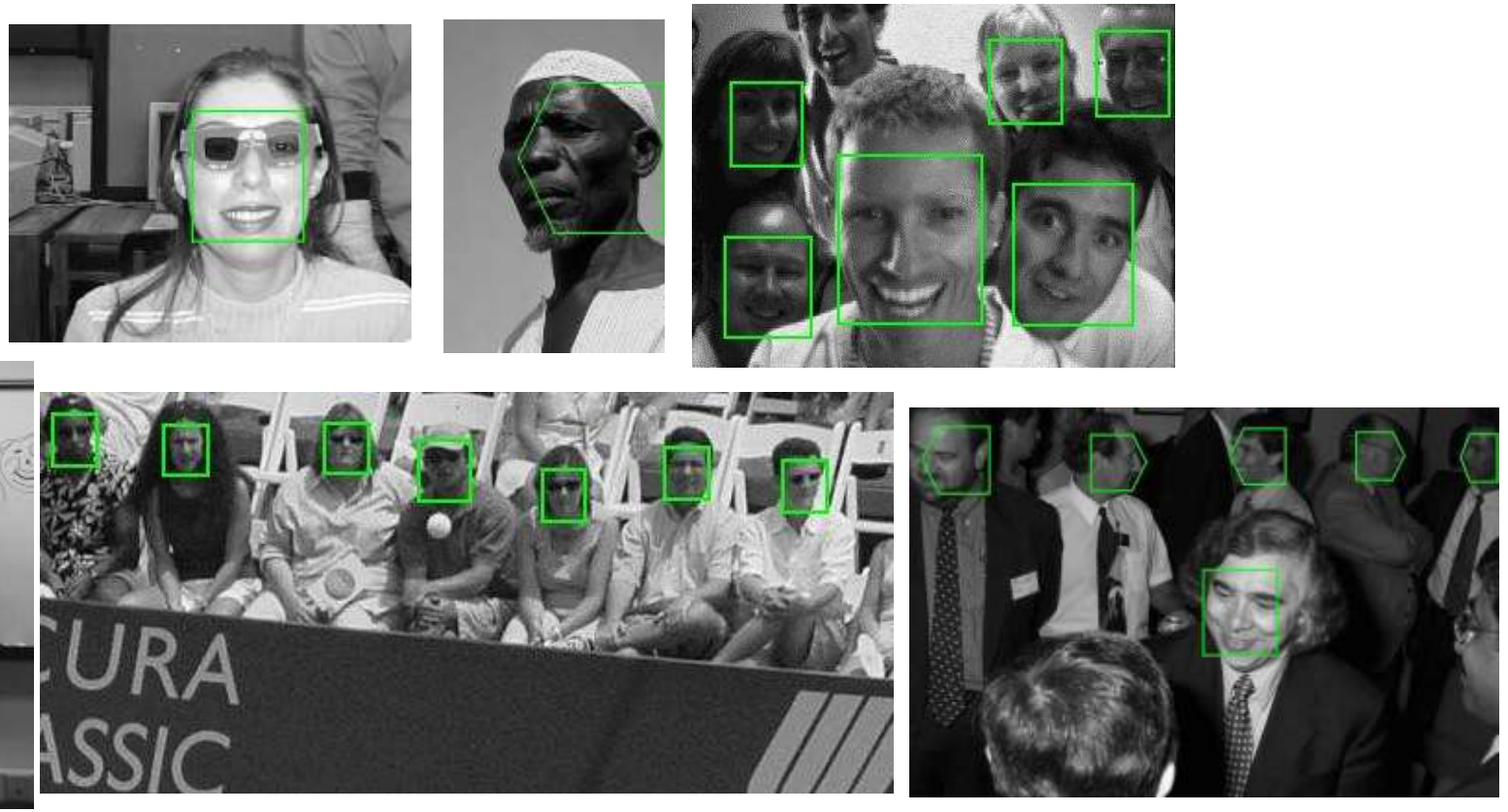
Linear SVM
+ Fast evaluation
+ Fast training
- Need to find good features

Non-linear kernelized SVM
+ Better class. acc. than linear
. Medium training
- Slow evaluation

Ask this question over and over again,
varying position, scale, multiple categories…
Speedups: hierarchical, early reject, feature sharing,
but same underlying question!

Face Detection

Schneiderman & Kanade (CMU), 2000…

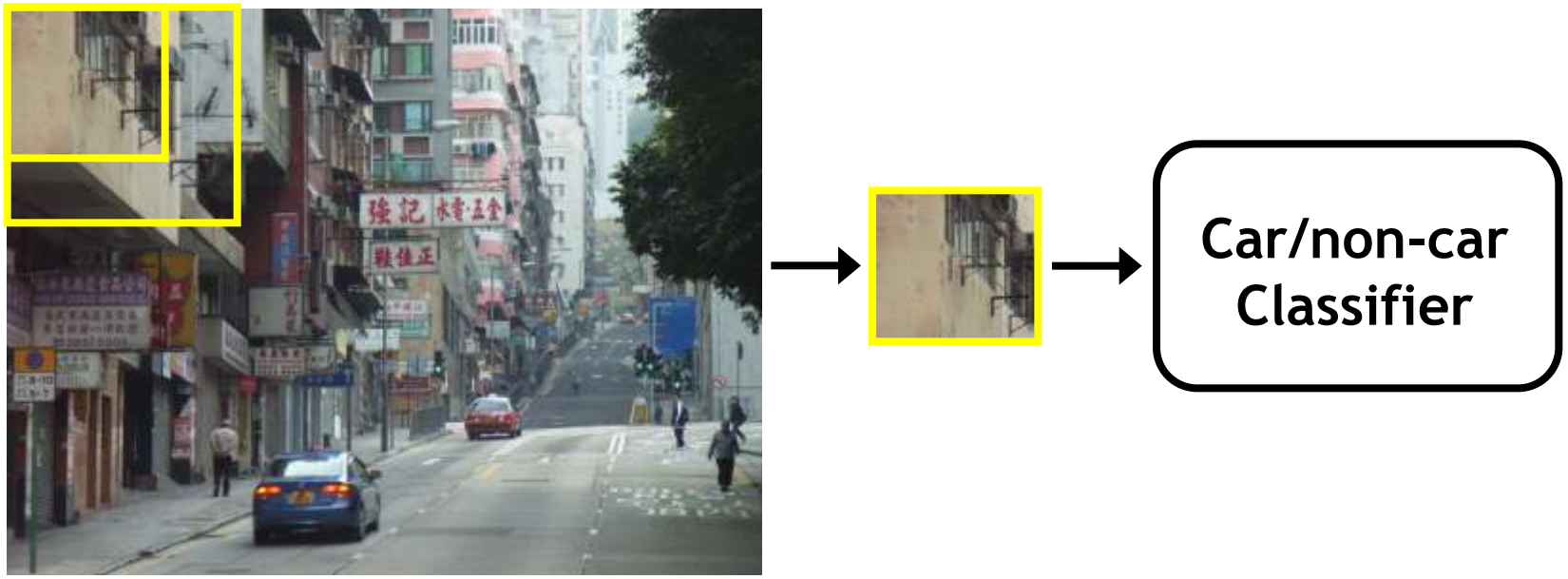


Generic Category Recognition: Basic Framework

- Build/train object model
  - Choose a representation
  - Learn or fit parameters of model / classifier
- Generate candidates in new image
- Score the candidates

Window-based models

Generating and scoring candidates



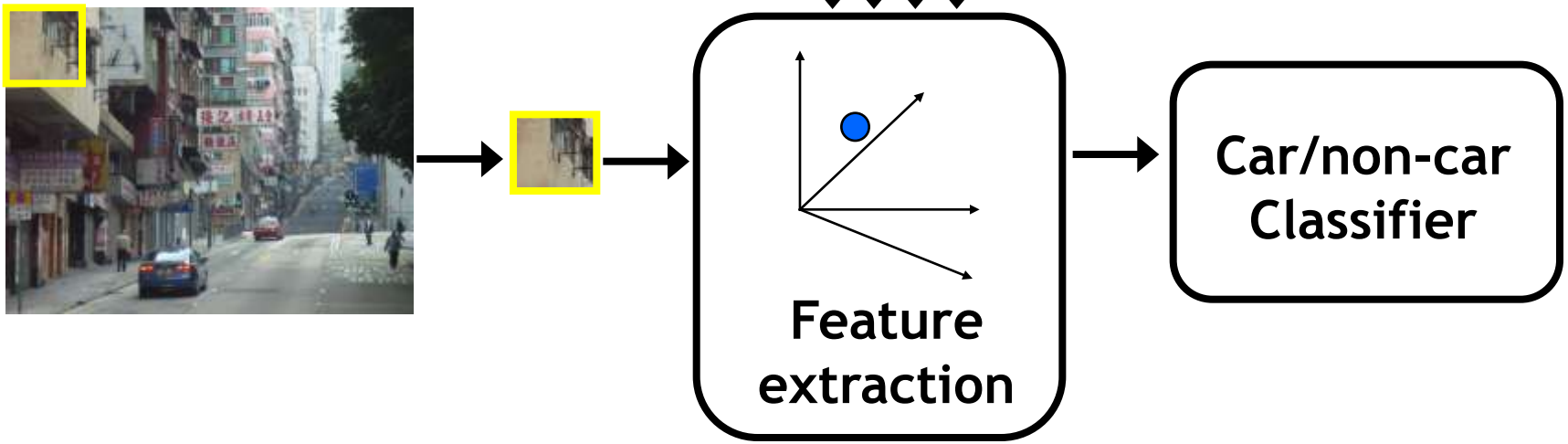
Window-based Object Detection

Training:

1. Obtain training data
2. Define features
3. Define classifier

Given new image:

1. Slide window
2. Score by classifier



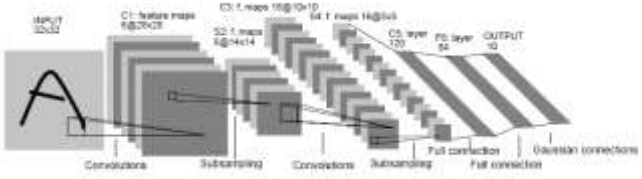
Discriminative Classifier Construction

Nearest neighbor

10^6 examples

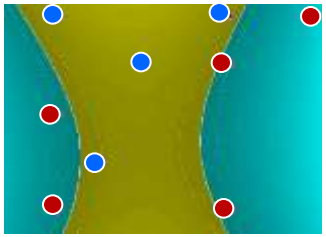
Shakhnarovich, Viola, Darrell 2003
Berg, Berg, Malik 2005...

Neural networks



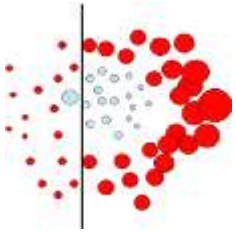
LeCun, Bottou, Bengio, Haffner 1998
Rowley, Baluja, Kanade 1998
…

Support Vector Machines



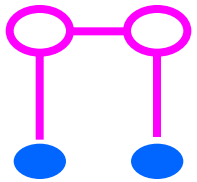
Guyon, Vapnik
Heisele, Serre, Poggio, 2001,…

Boosting



Viola, Jones 2001, Torralba et al. 2004, Opelt et al. 2006,…

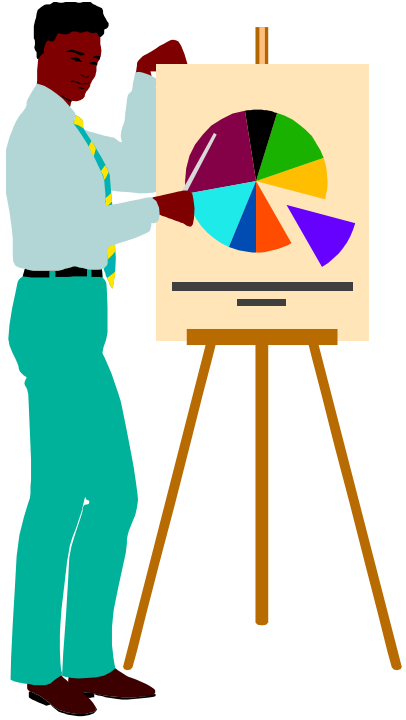
Conditional Random Fields



McCallum, Freitag, Pereira 2000; Kumar, Hebert 2003
…

Slide adapted from Antonio Torralba

Today's class



Sliding Window Approach

Sliding Windows for Face Detection

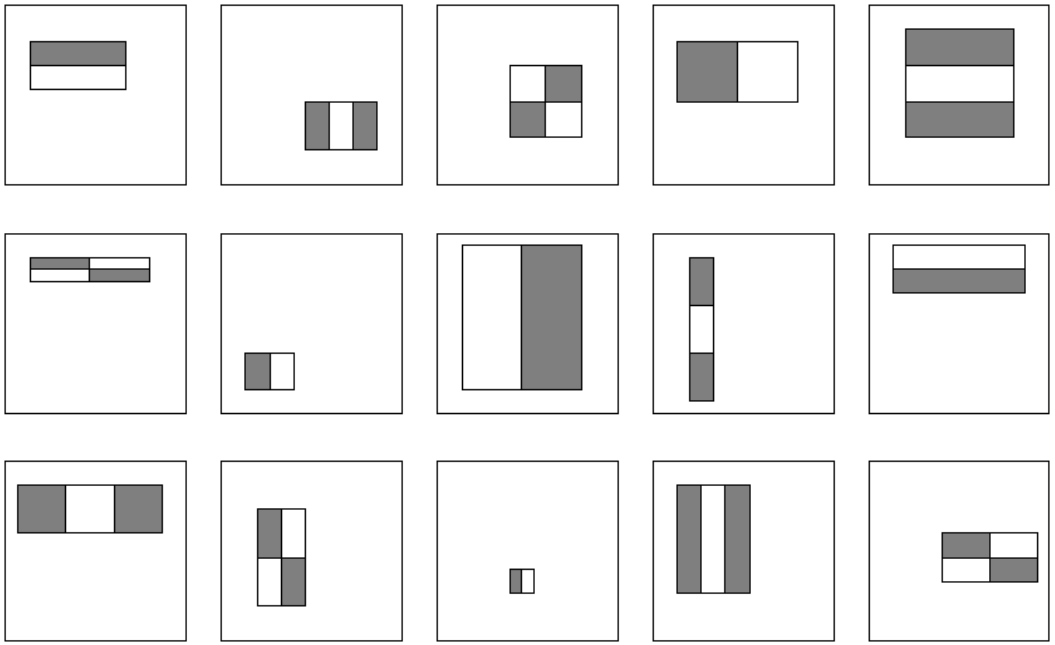
Image Segmentation for Object Localization

Viola-Jones Face Detector

Main idea:

- Represent local texture with efficiently computable "rectangular" features within window of interest
- Select discriminative features to be weak classifiers
- Use boosted combination of them as final classifier
- Form a cascade of such classifiers, rejecting clear negatives quickly

Viola-Jones Detector: Features



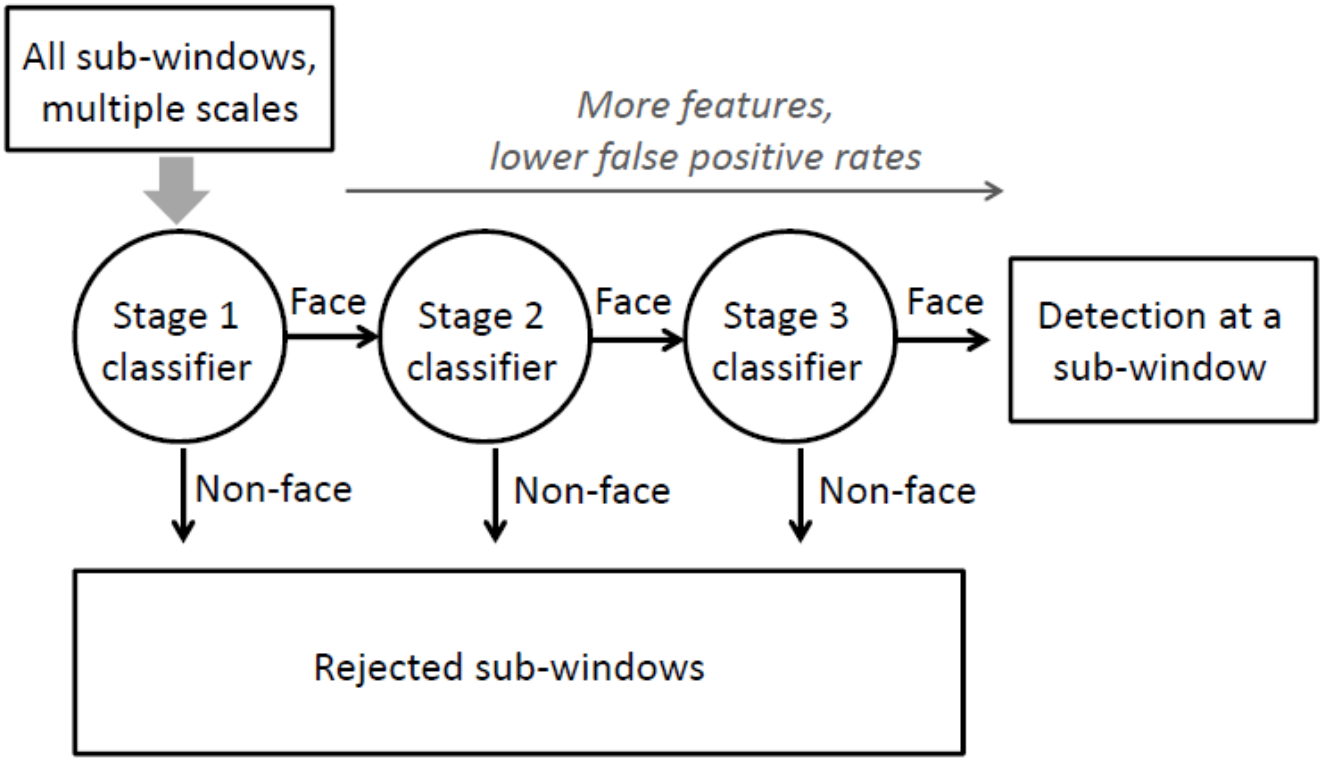
Considering all possible filter parameters: position, scale, and type:

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?

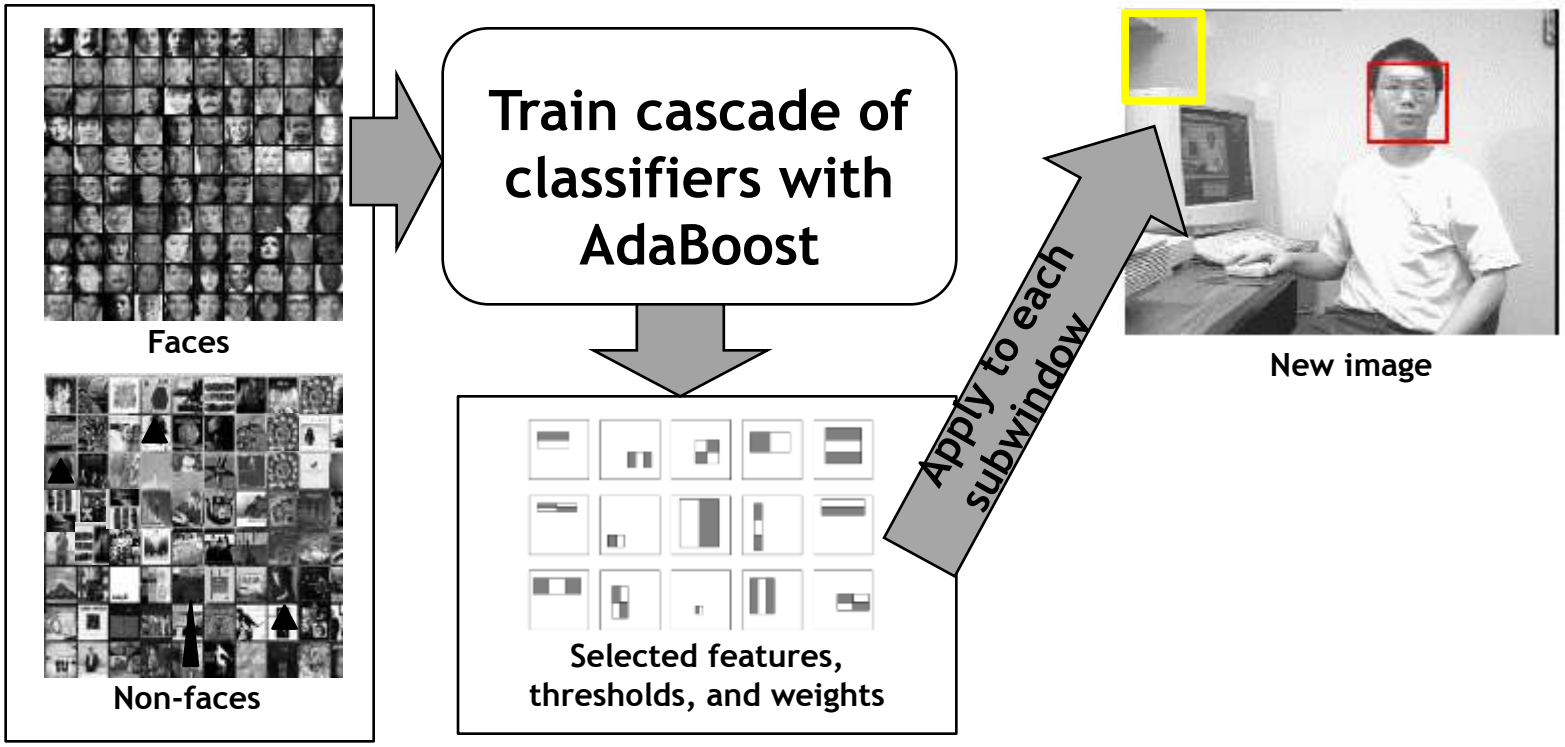
Use AdaBoost both to select the informative features and to form the classifier

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

Viola-Jones Detector: Summary



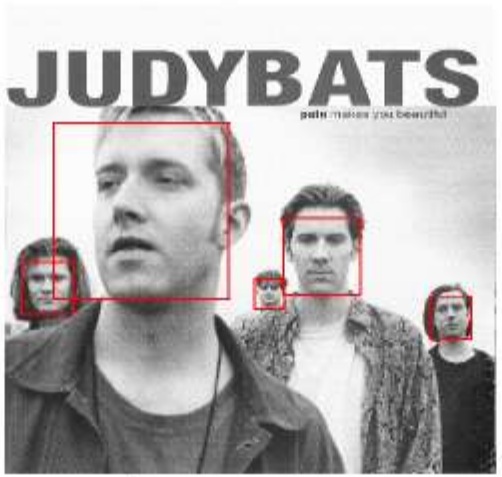
Train with 5K positives, 350M negatives

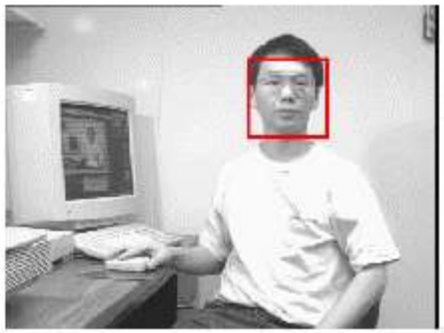
Real-time detector using 38 layer cascade

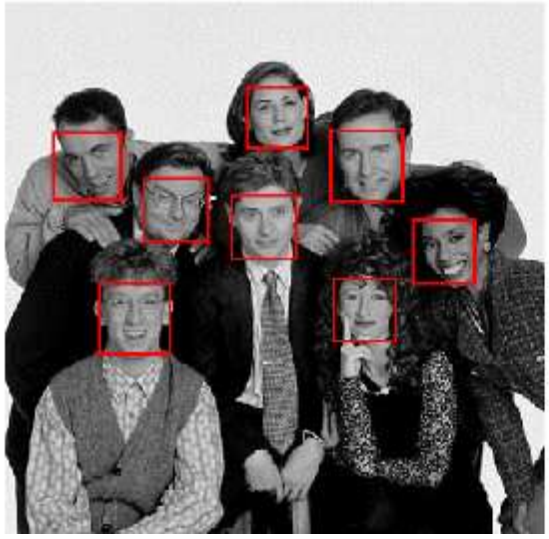
6061 features in all layers

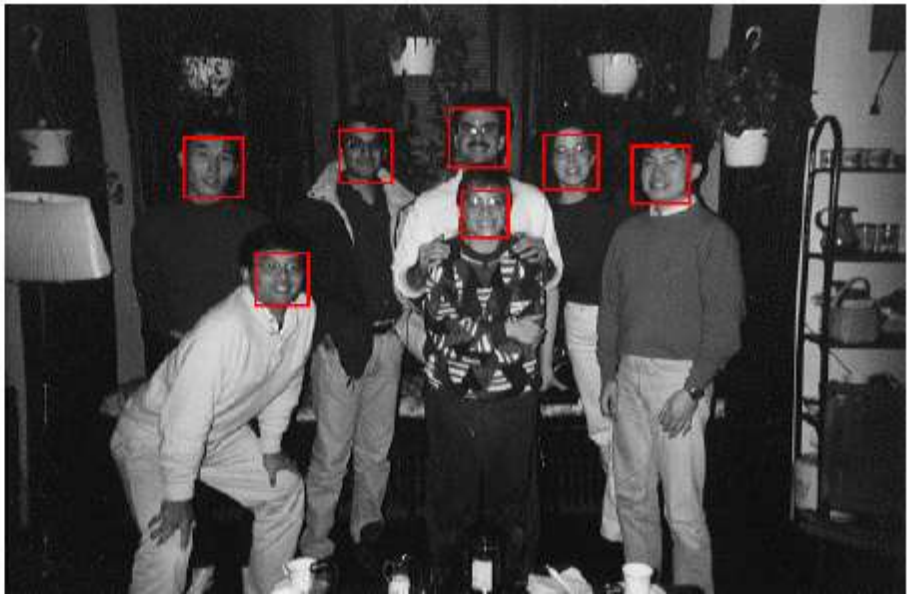
[Implementation available in OpenCV:
http://www.intel.com/technology/computing/opencv/]

Viola-Jones Face Detector: Results

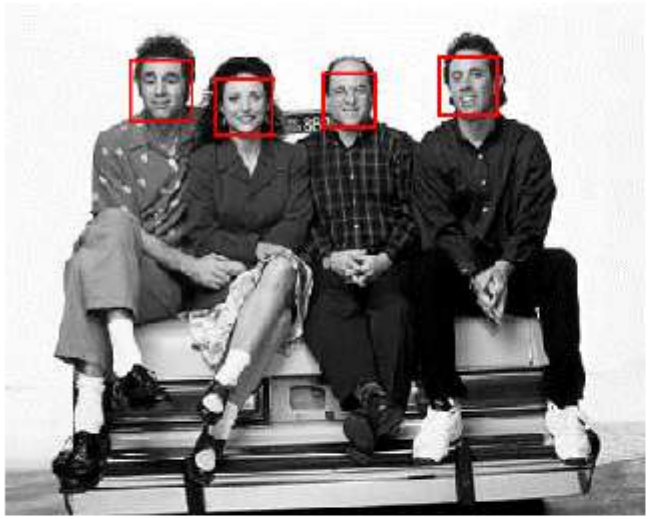


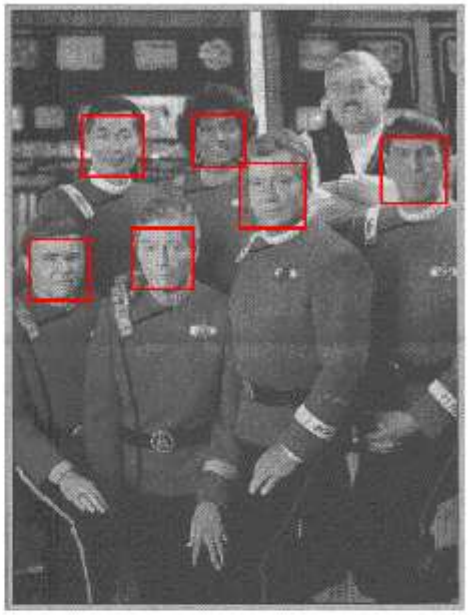


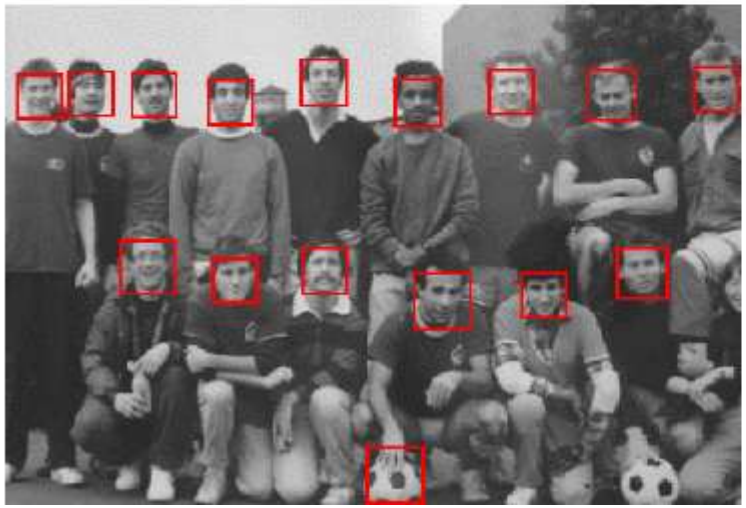


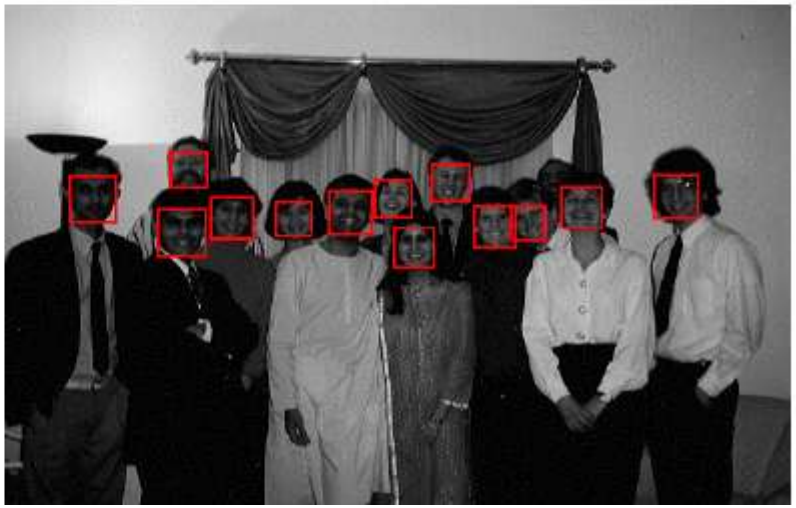


Viola-Jones Face Detector: Results



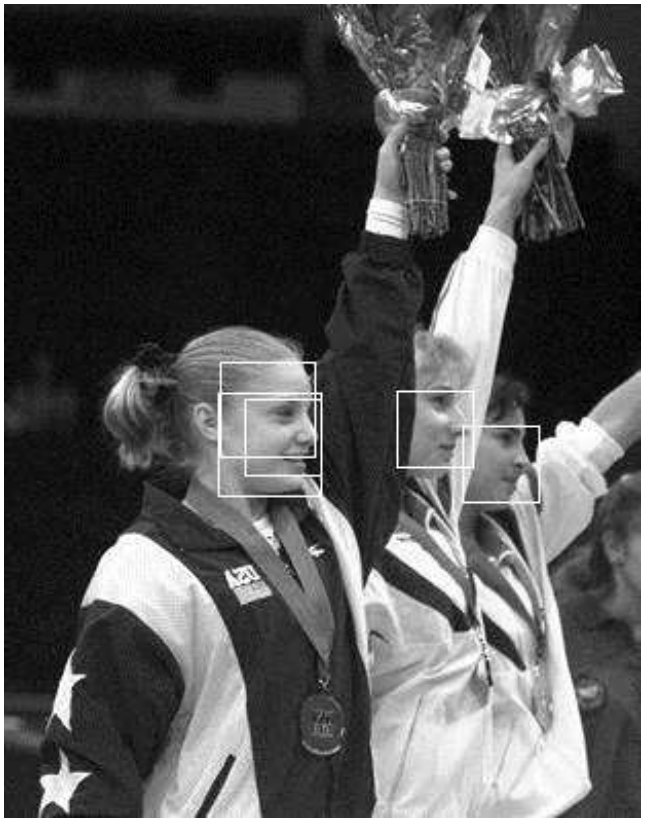






Viola-Jones Face Detector: Results

Viola-Jones Face Detector: Results



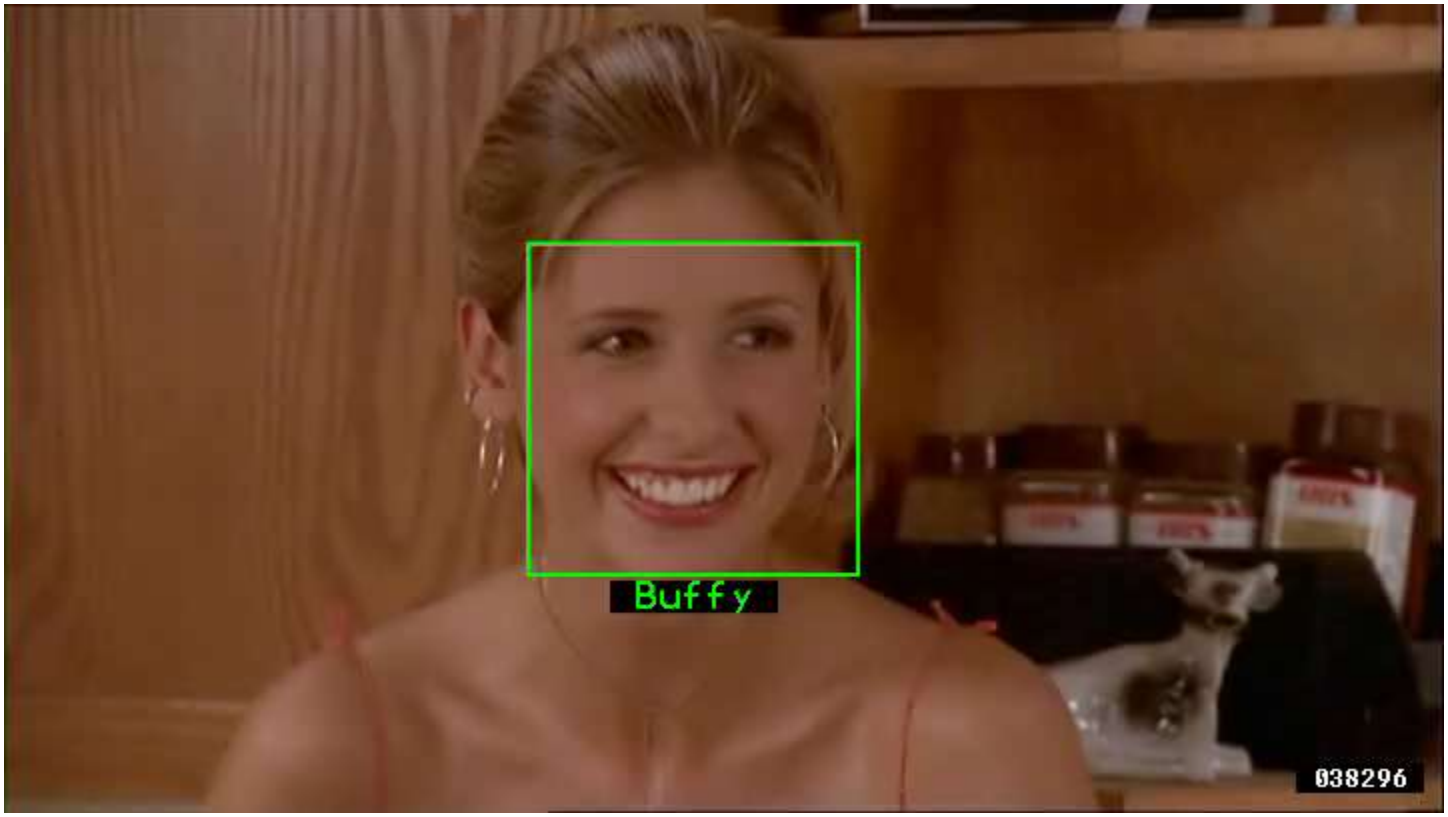
Speed: Frontal Face Detector

- Schneiderman-Kanade (2000): 5 seconds

- Viola-Jones (2001): 15 fps

Slide: Derek Hoiem

Example using Viola-Jones Detector



Frontal faces detected and then tracked, character names inferred with alignment of script and subtitles.

Everingham, M., Sivic, J. and Zisserman, A.
"Hello! My name is... Buffy" - Automatic naming of characters in TV video, BMVC 2006. http://www.robots.ox.ac.uk/~vgg/research/nface/index.html

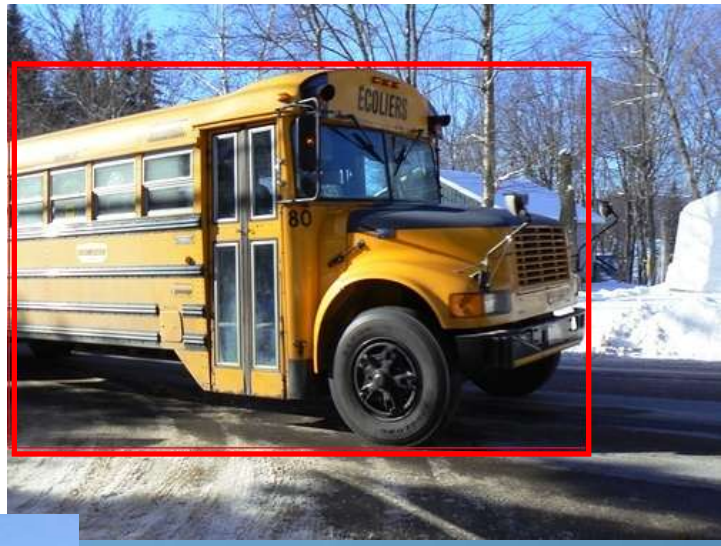
Consumer Application: iPhoto 2009

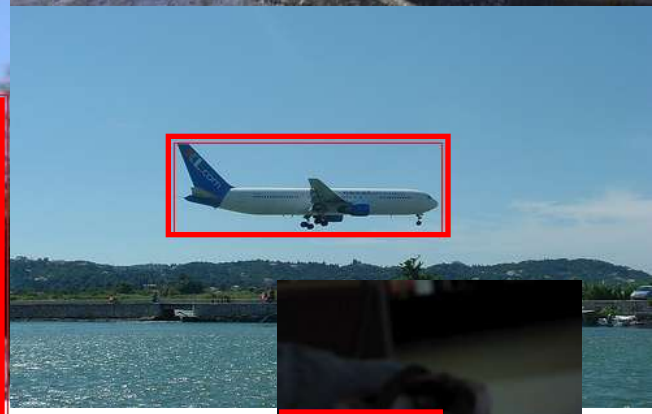
http://www.apple.com/ilife/iphoto/

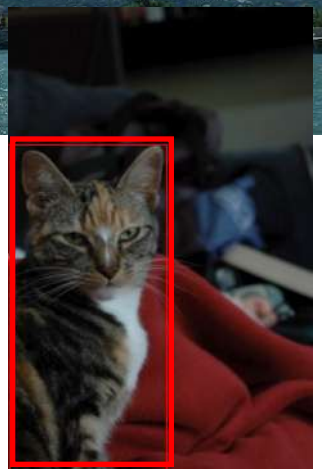
Slide credit: Lana Lazebnik

Object localization

– Where is the object located?







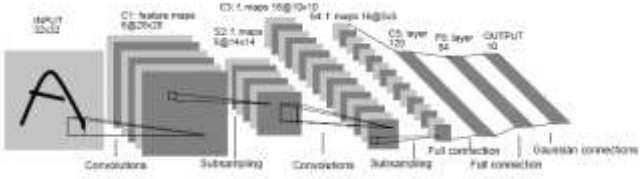
Discriminative Classifier Construction

Nearest neighbor

10^6 examples

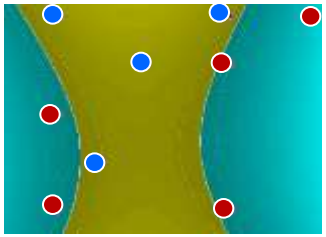
Shakhnarovich, Viola, Darrell 2003
Berg, Berg, Malik 2005...

Neural networks



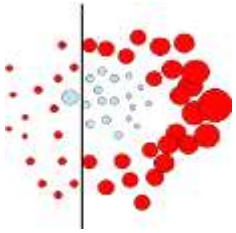
LeCun, Bottou, Bengio, Haffner 1998
Rowley, Baluja, Kanade 1998
…

Support Vector Machines



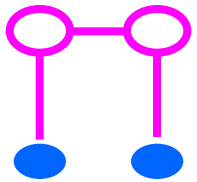
Guyon, Vapnik
Heisele, Serre, Poggio, 2001,…

Boosting



Viola, Jones 2001, Torralba et al. 2004, Opelt et al. 2006,…

Conditional Random Fields

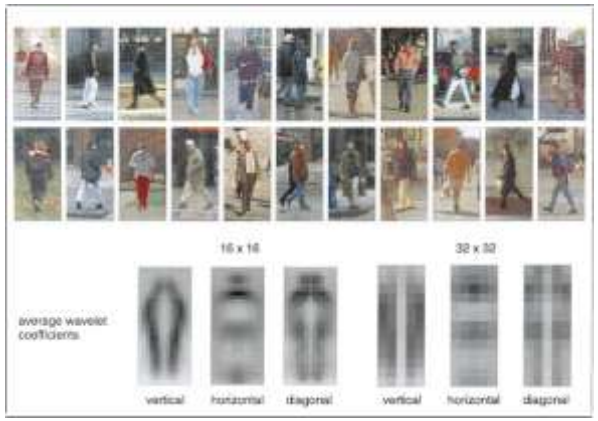


McCallum, Freitag, Pereira 2000; Kumar, Hebert 2003
…

Slide adapted from Antonio Torralba

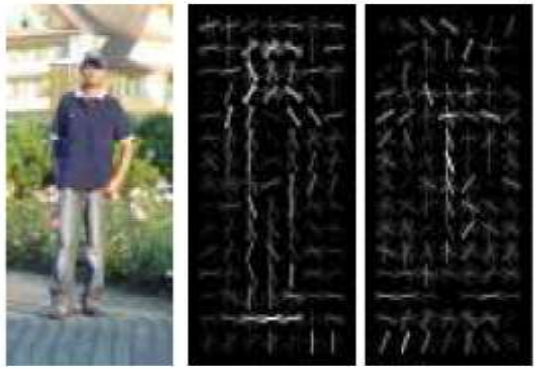
Pedestrian detection

- **Detecting upright, walking humans also possible using sliding window's appearance/texture; e.g.,**



SVM with Haar wavelets [Papageorgiou & Poggio, IJCV 2000]

Space-time rectangle features [Viola, Jones & Snow, ICCV 2003]



SVM with HoGs [Dalal & Triggs, CVPR 2005]

Window-based Detection: Strengths

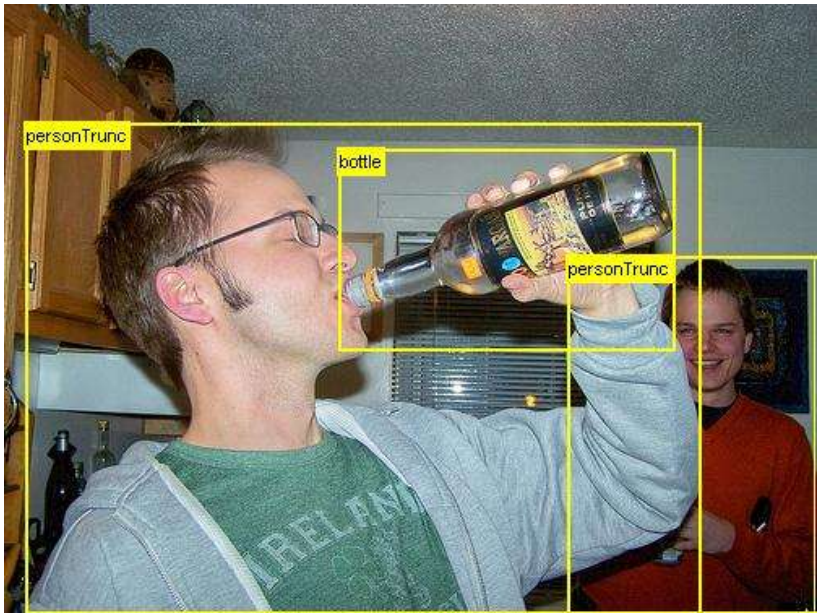
- **Sliding window detection and global appearance descriptors:**
  - **Simple detection protocol to implement**
  - **Good feature choices critical**
  - **Past successes for certain classes**

Window-based detection: Limitations

- **High computational complexity**
  - **For example: 250,000 locations x 30 orientations x 4 scales = 30,000,000 evaluations!**
  - **If training binary detectors independently, means cost increases linearly with number of classes**
- **With so many windows, false positive rate better be low**

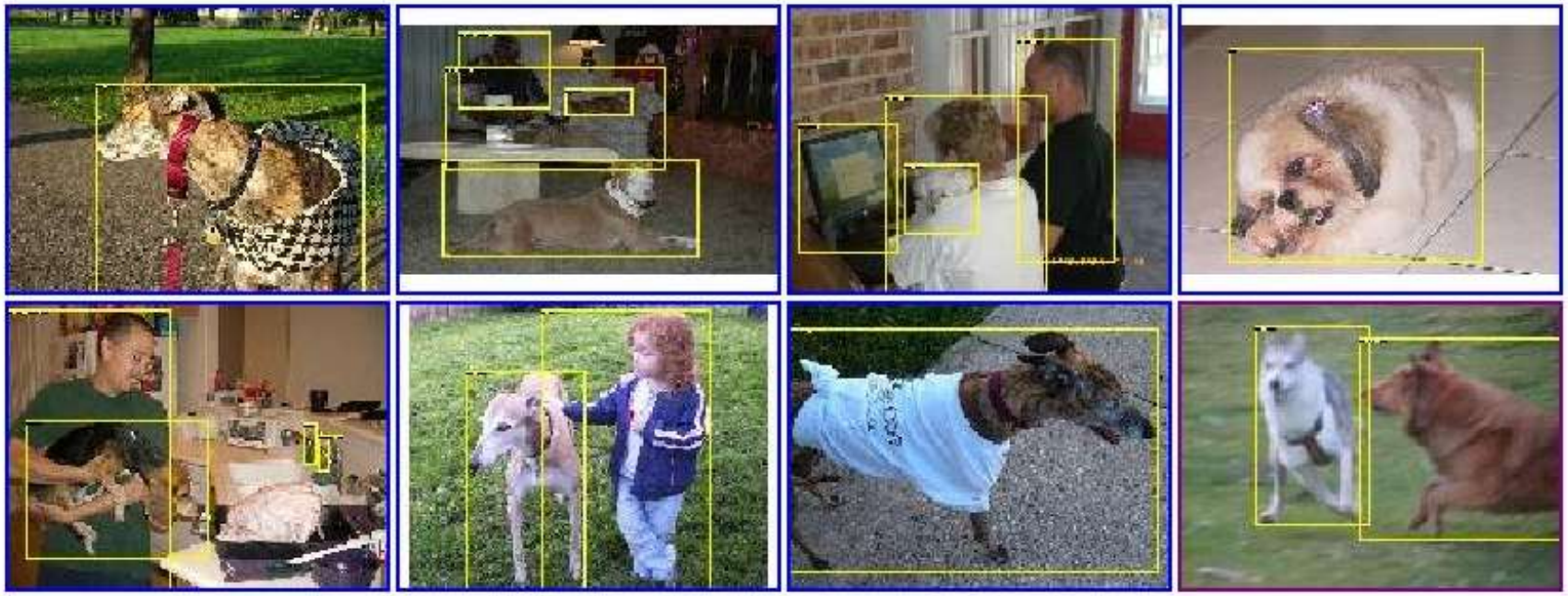
Limitations (continued)

- **Not all objects are “box” shaped**



Limitations (continued)

- **Non-rigid, deformable objects not captured well with representations assuming a fixed 2d structure; or must assume fixed viewpoint**
- **Objects with less-regular textures not captured well with holistic appearance-based descriptions**

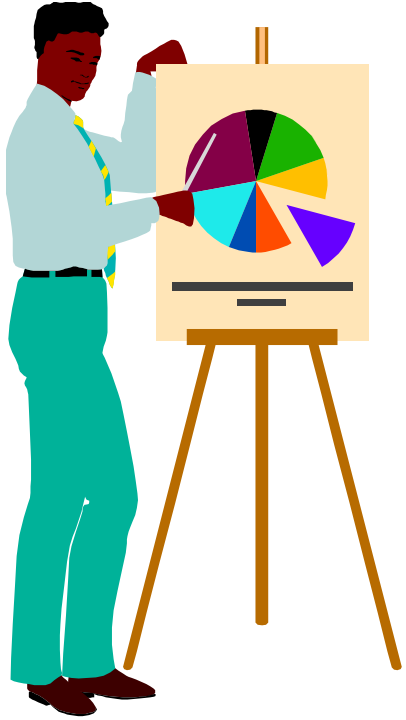


Kristen Grauman

Summary

- Basic pipeline for window-based detection
  - Model/representation/classifier choice
  - Sliding window and classifier scoring
- Viola-Jones face detector
  - Exemplar of basic paradigm
  - Plus key ideas: rectangular features, Adaboost for feature selection, cascade, hard negatives.
- Pros and cons of window-based detection

Today's class



Sliding Window Approach

Sliding Windows for Face Detection

Image Segmentation for Object Localization

Segmentation as Selective Search for Object Detection

Koen van de Sande
Jasper Uijlings
Arnold Smeulders
Theo Gevers

Selective Search for Recognition

- Design criteria
  - High recall
  - Coarse locations are sufficient
    - ⇨ Bounding boxes
  - Fast to compute
    - ⇨ Efficient low-level features
    - ⇨ <10s per image

Selective Search: High Recall

- Image is intrinsically hierarchical

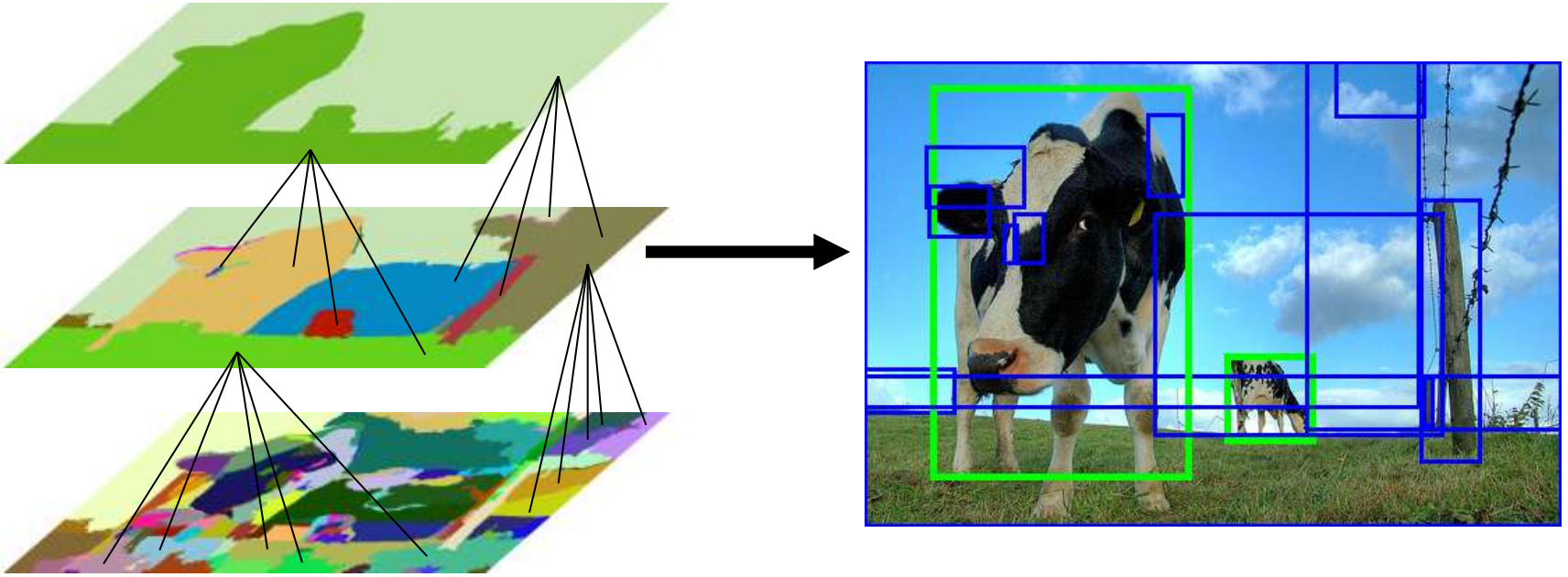
- Segmentation at a single scale won't find all objects

Selective Search: Approach

- Hypotheses based on hierarchical grouping

Selective Search: Approach

- Hypotheses based on hierarchical grouping

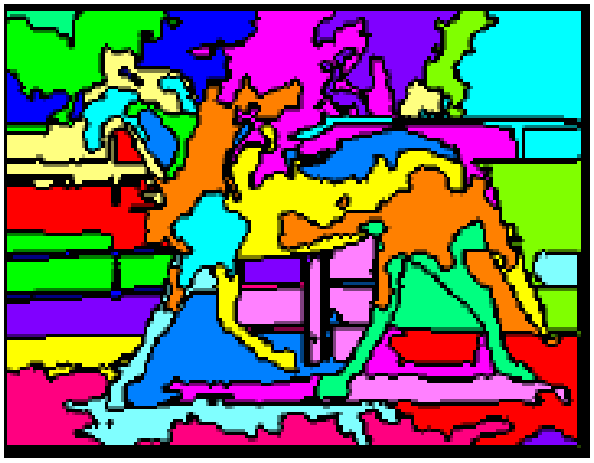


Group adjacent regions on color/texture cues

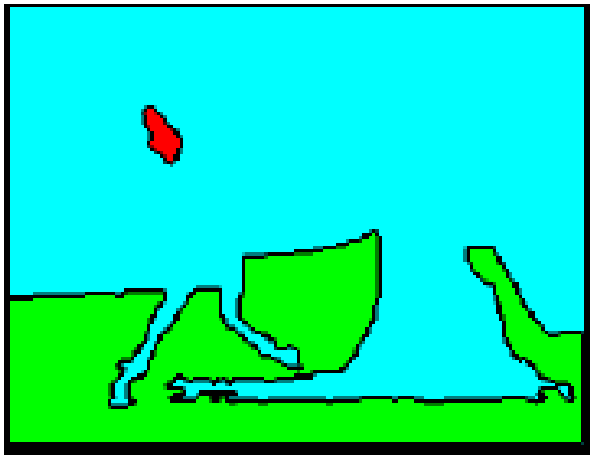
Image Segmentation

Goal: Break up the image into meaningful or perceptually similar regions

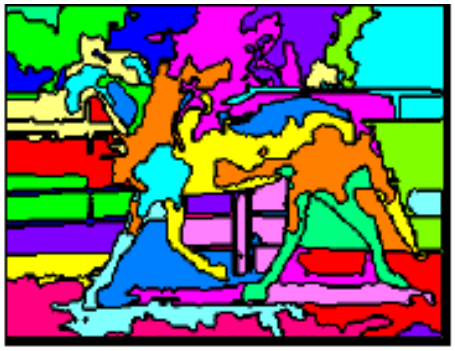
Types of segmentations

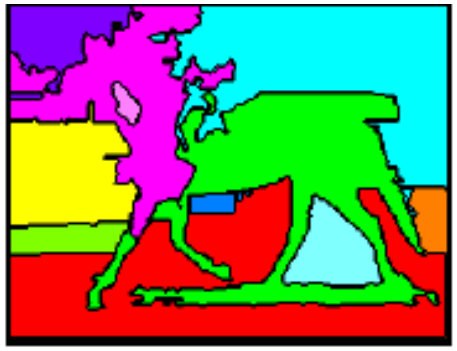


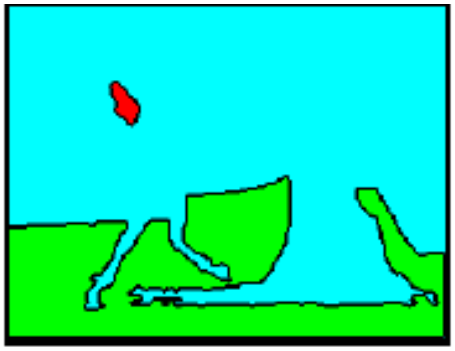
Oversegmentation



Undersegmentation



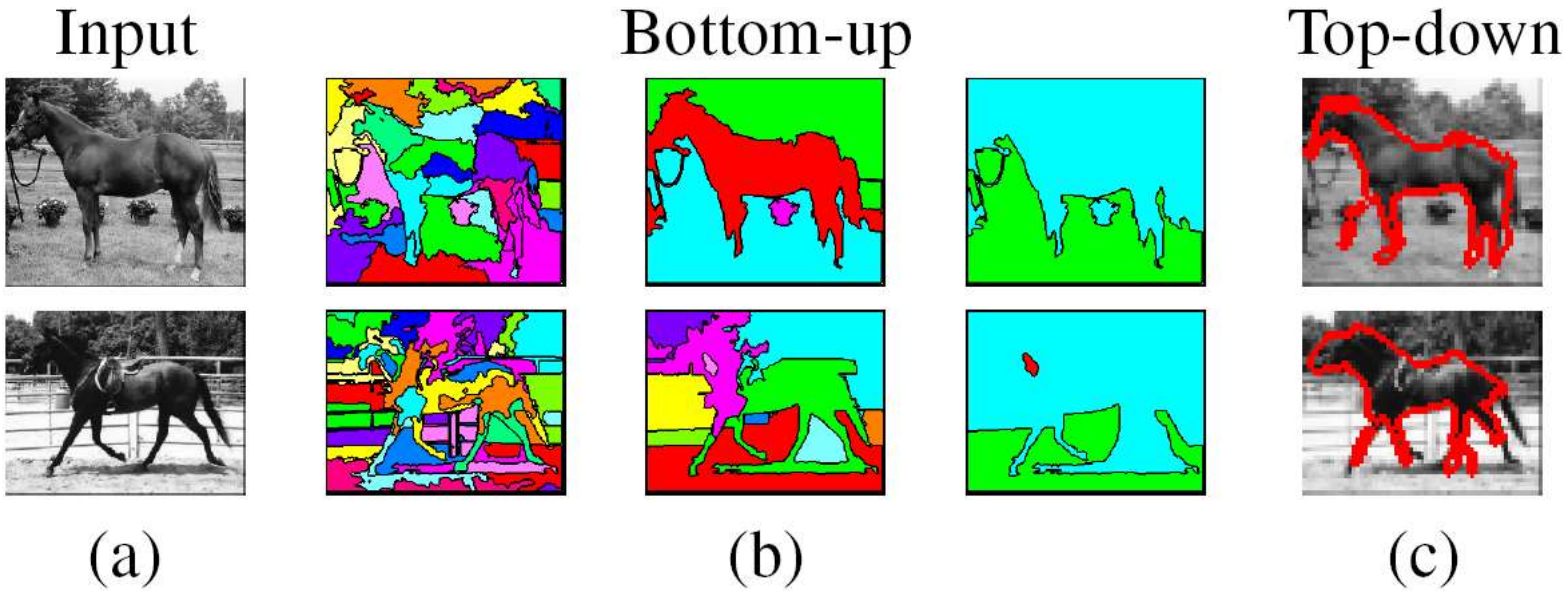




Multiple Segmentations

Major Processes for Segmentation

- Bottom-up: group tokens with similar features
- Top-down: group tokens that likely belong to the same object



[Levin and Weiss 2006]

Region-based Image Segmentation

How do we Segment?

- K-means
  - Iteratively re-assign points to the nearest cluster center
- Split-and-Merge
  - Iteratively split and merge regions
- Agglomerative clustering
  - Start with each point as its own cluster and iteratively merge the closest clusters
- Mean-shift clustering
  - Estimate modes of pdf
- Spectral clustering
  - Split the nodes in a graph based on assigned links with similarity weights

K-means

1. Initialize cluster centers: $\mathbf{c}^0$; t=0

2. Assign each point to the closest center

$$\boldsymbol{\delta}^t = \underset{\boldsymbol{\delta}}{\operatorname{argmin}} \frac{1}{N} \sum_j^N \sum_i^K \delta_{ij} \left(\mathbf{c}_i^{t-1} - \mathbf{x}_j \right)^2$$

3. Update cluster centers as the mean of the points

$$\mathbf{c}^t = \underset{\mathbf{c}}{\operatorname{argmin}} \frac{1}{N} \sum_j^N \sum_i^K \delta_{ij}^t \left(\mathbf{c}_i - \mathbf{x}_j \right)^2$$

4. Repeat 2-3 until no points are re-assigned (t=t+1)

Slide: Derek Hoiem

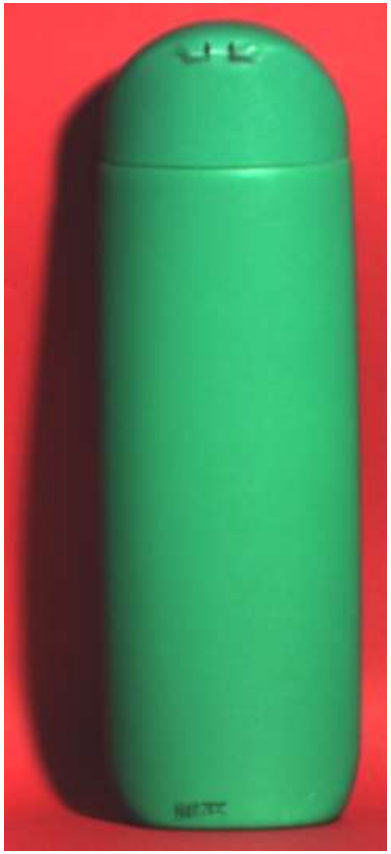
K-means: Design Choices

- Initialization
  - Randomly select K points as initial cluster center
  - Or greedily choose K points to minimize residual
- Distance measures
  - Traditionally Euclidean, could be others

K-means

Problem: partitioning a colour image into disjoint regions corresponding to objects in the image

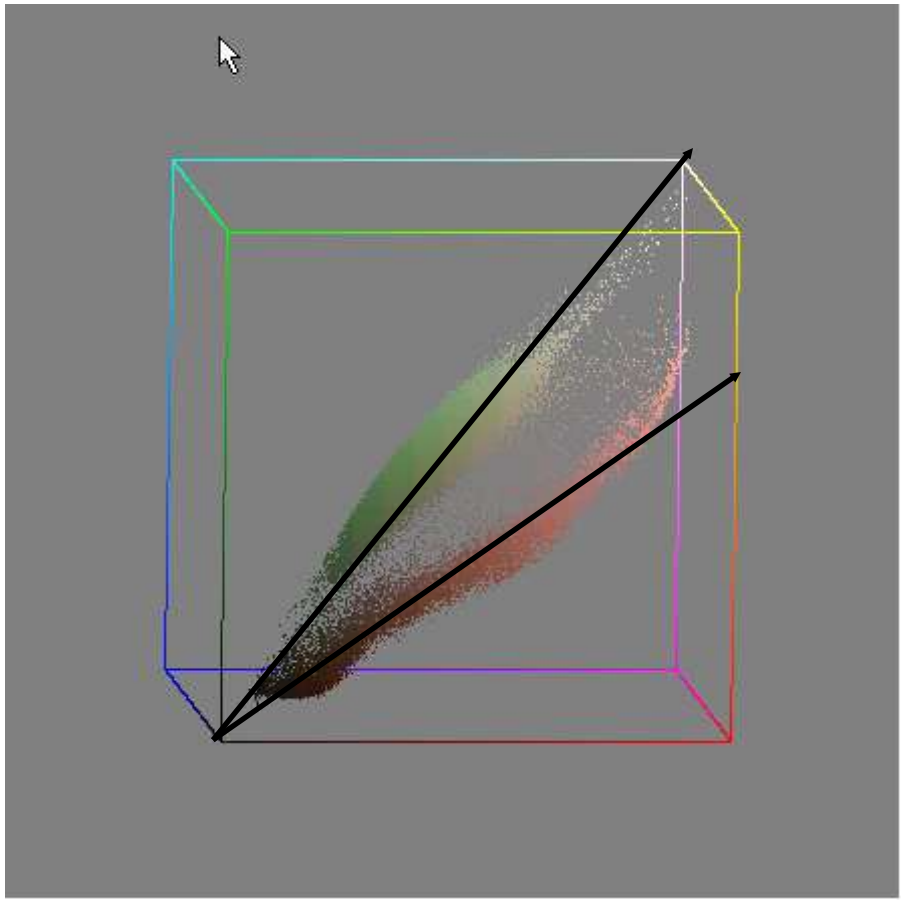
Homogeneity criterion: mean and variance



K-means: Matte Objects

A.1 Features: normalized rgb. A.2 Homogeneity criterion: mean and variance

B.1 Features: RGB. B2. Homogeneity criterion: fitting lines

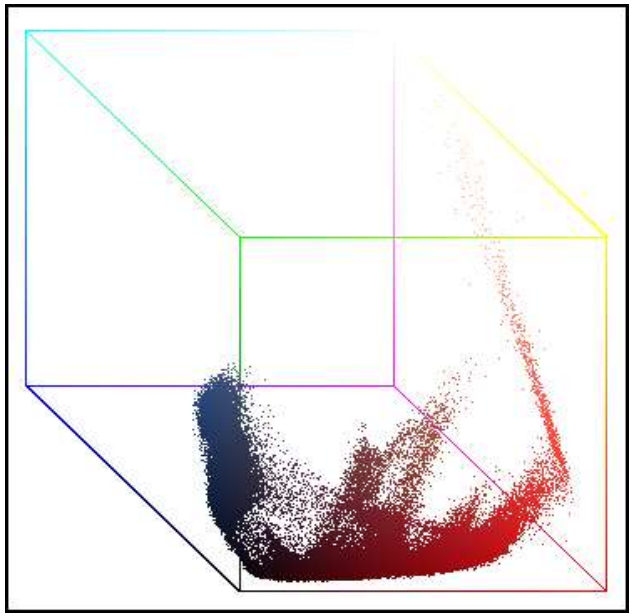


K-means: Shiny Objects

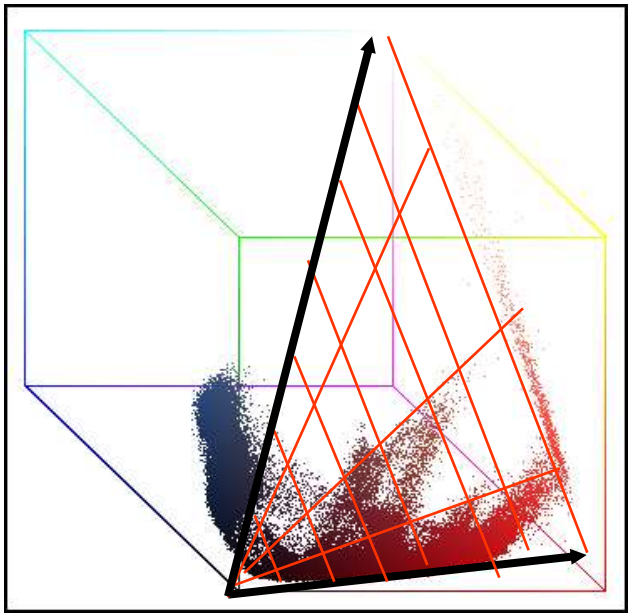
A.1 Features: normalized H. A.2 Homogeneity criterion: mean and variance

B.1 Features: RGB. B2. Homogeneity criterion: fitting planes

Original image



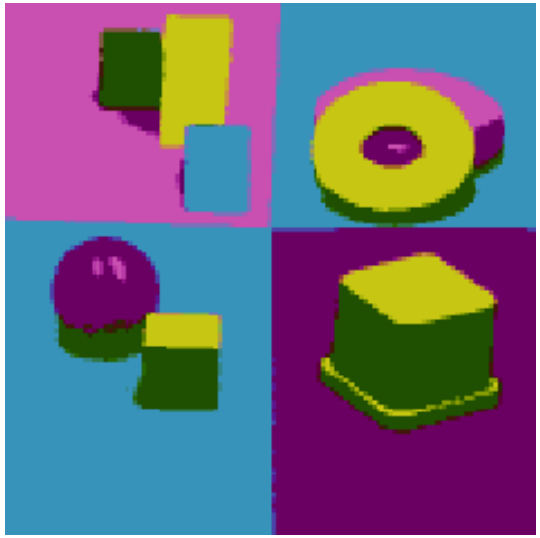
3D plot of whole image



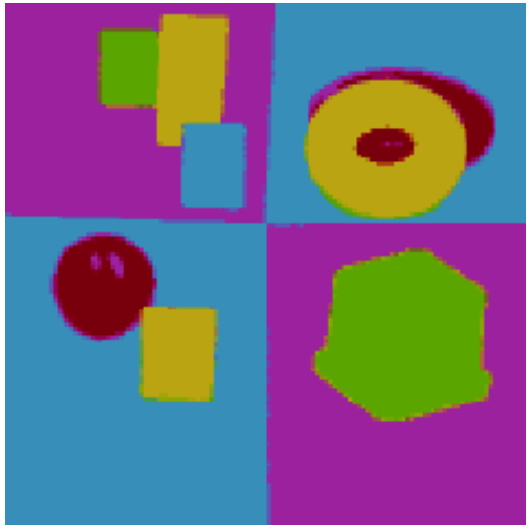
3D plot of highlight with plane fitted

K-means

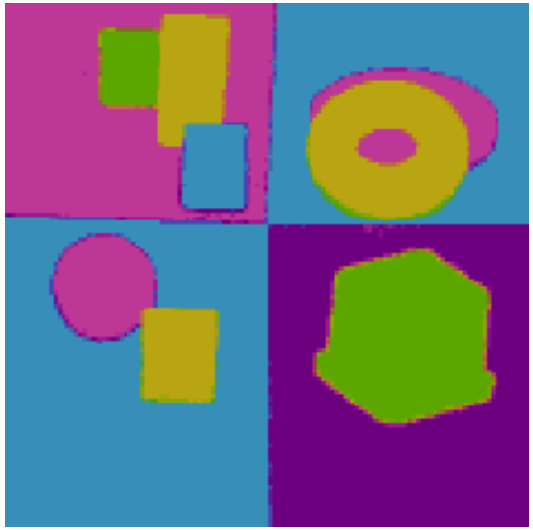
Original image



Mean&var in RGB

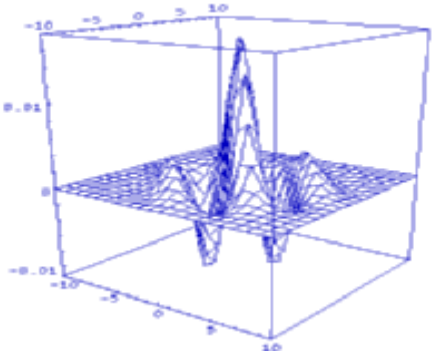


Fitting lines in RGB



Fitting planes in RGB

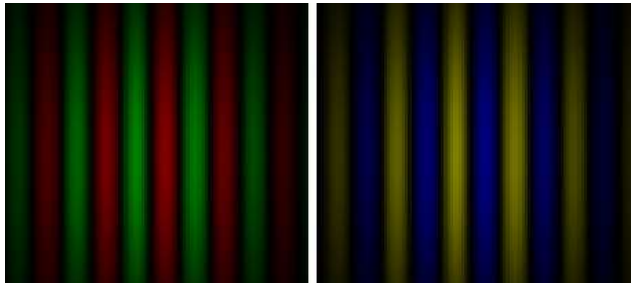
Kmeans: Texture

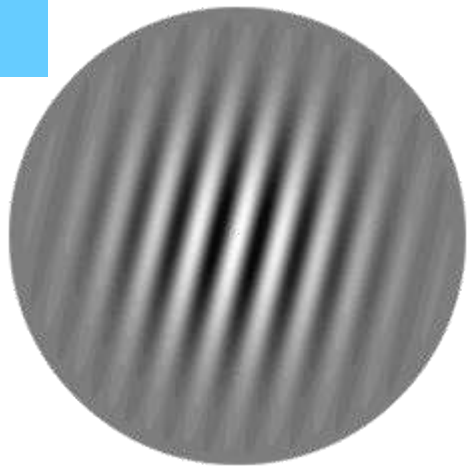


The 2D Gabor function is:

$$h(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\delta^2}} e^{2\pi j(ux+vy)}$$

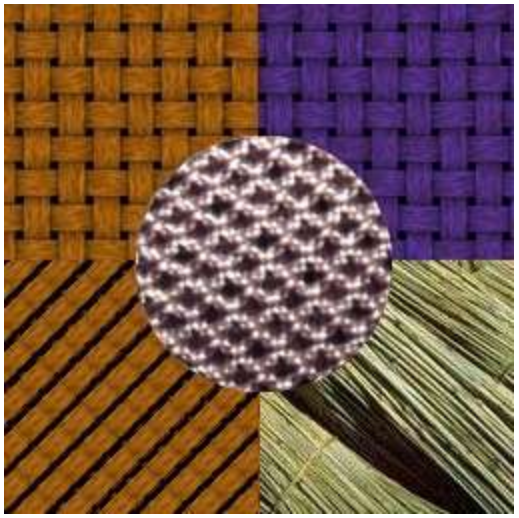
Tuning parameters: u, v, □
+ usual invariants by combination



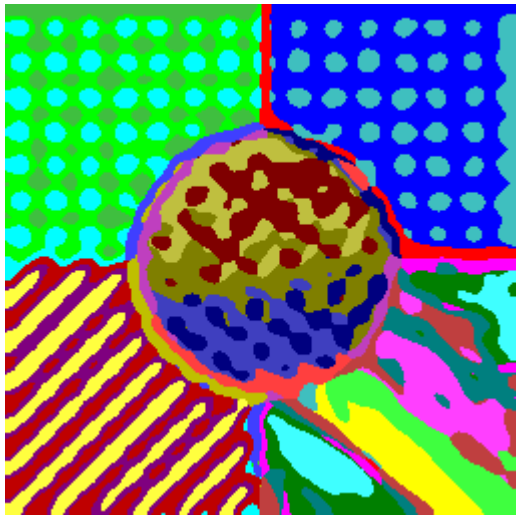


Minh SP 2005

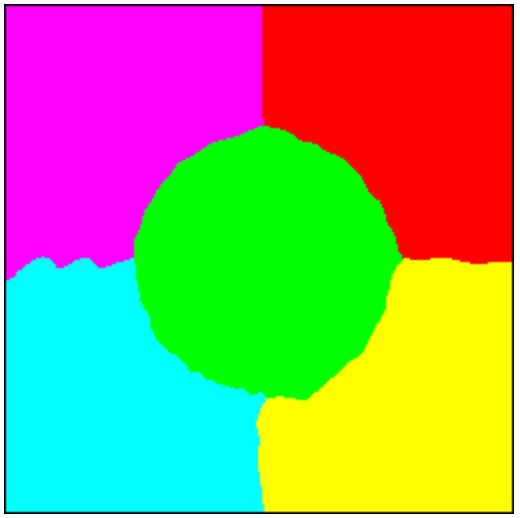
Gabor and K-means Clustering



Original image



K-means clustering



Segmentation

Minh, ECCV, 2002

Gabor and K-means Clustering

How do we Segment?

- K-means
  - Iteratively re-assign points to the nearest cluster center
- Split-and-Merge
  - Iteratively split and merge regions
- Agglomerative clustering
  - Start with each point as its own cluster and iteratively merge the closest clusters
- Mean-shift clustering
  - Estimate modes of pdf
- Spectral clustering
  - Split the nodes in a graph based on assigned links with similarity weights

Split-and-Merge

Split regions until patch is homogeneous wrt query image

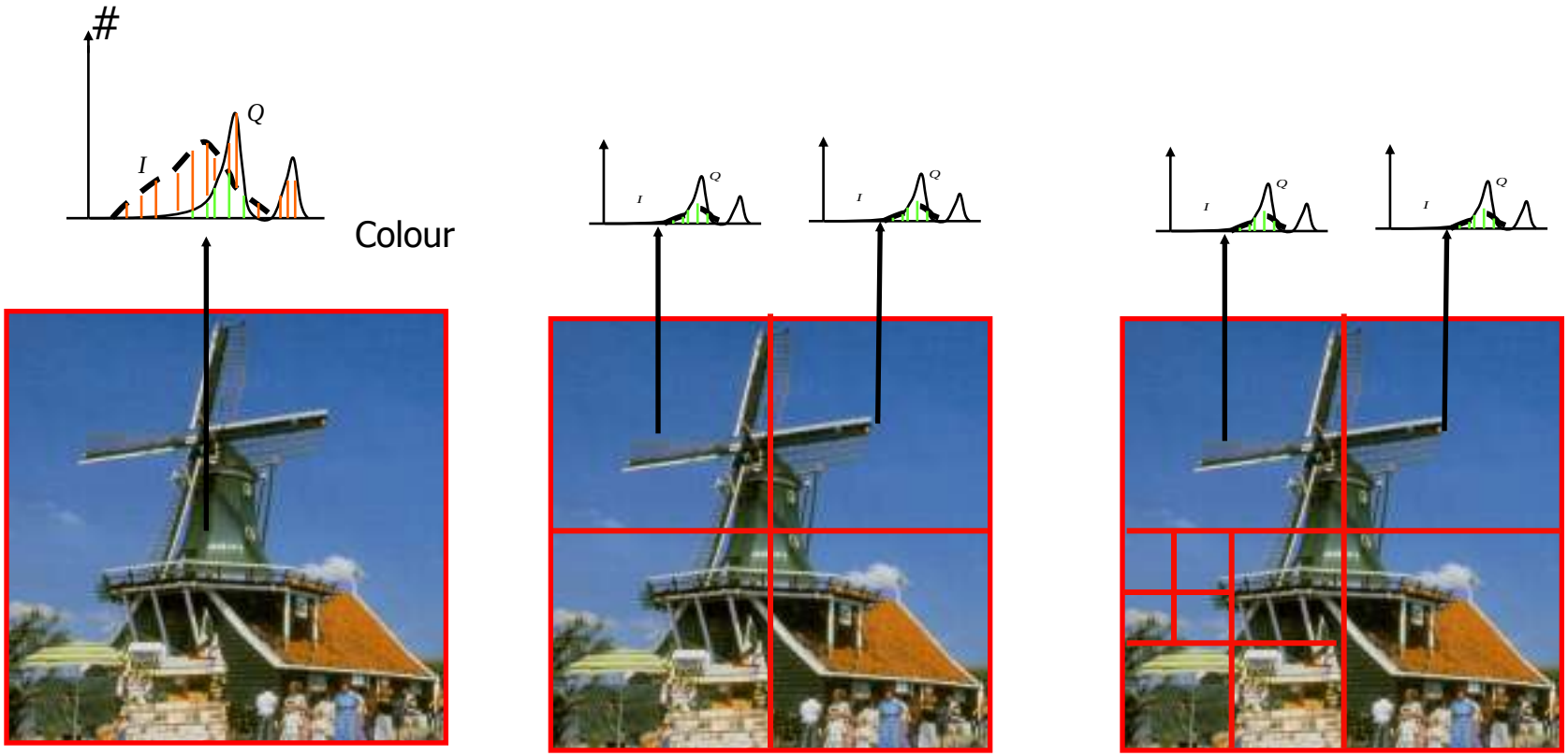
Split-and-Merge

... and merge patches which are alike.

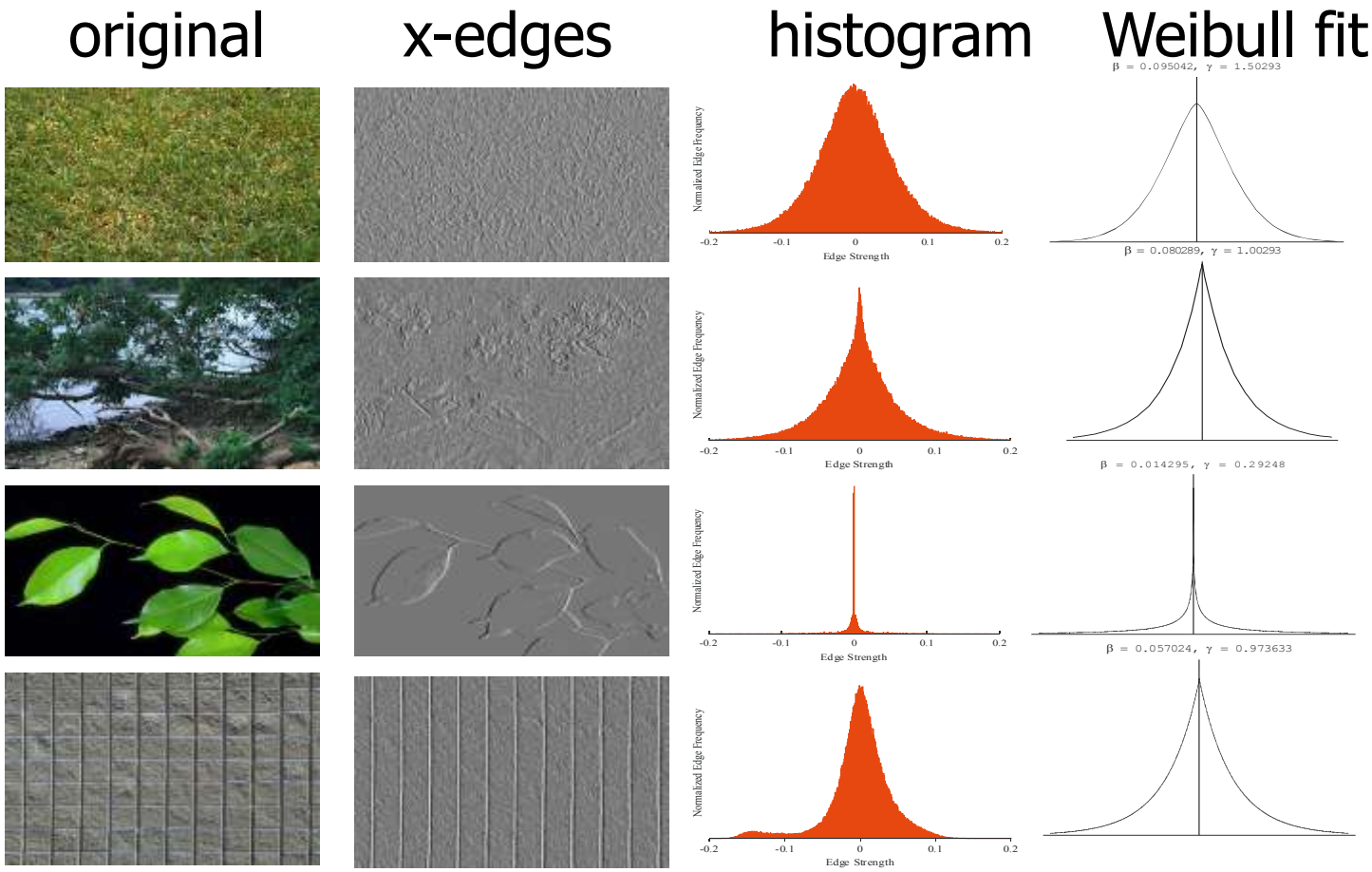
It works because of spatial coherence.

Split-and-Merge

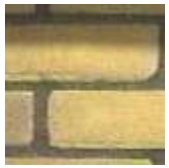
Split regions until patch is homogeneous wrt query image



Split-and-Merge Weibull



Split-and-Merge



texture

Original image

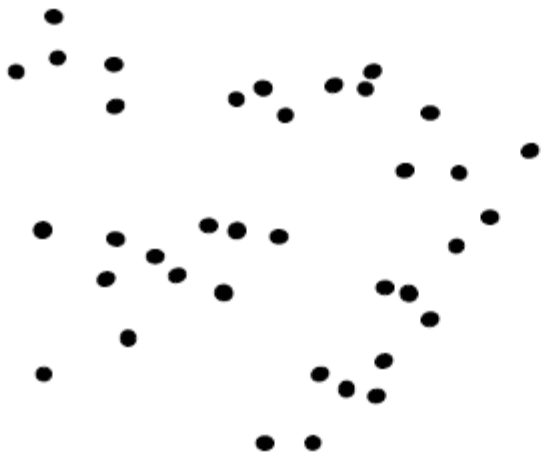
Result

Split-and-Merge

How do we Segment?

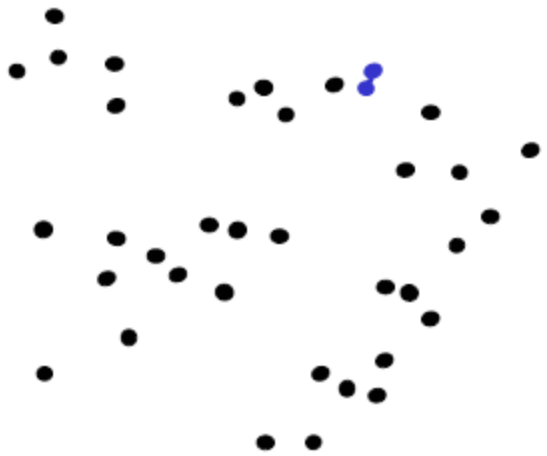
- K-means
  - Iteratively re-assign points to the nearest cluster center
- Split-and-Merge
  - Iteratively split and merge regions
- Agglomerative clustering
  - Start with each point as its own cluster and iteratively merge the closest clusters
- Mean-shift clustering
  - Estimate modes of pdf
- Spectral clustering
  - Split the nodes in a graph based on assigned links with similarity weights

Agglomerative Clustering



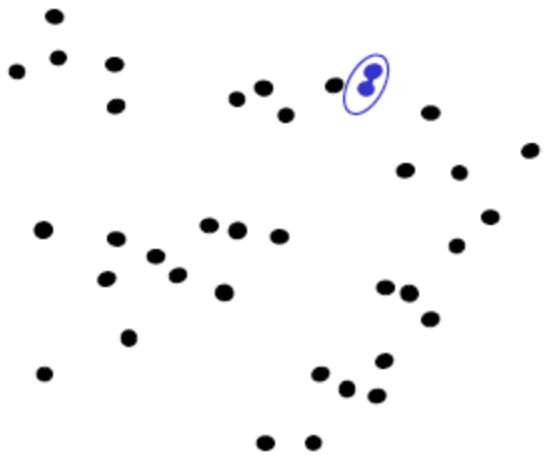
1. Say "Every point is its own cluster"

Agglomerative Clustering



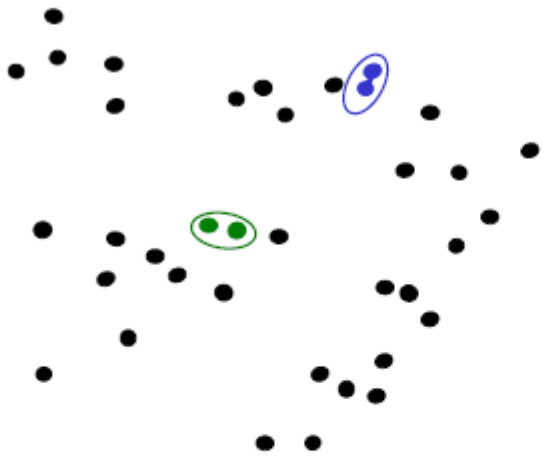
1. Say "Every point is its own cluster"
2. Find "most similar" pair of clusters

Agglomerative Clustering



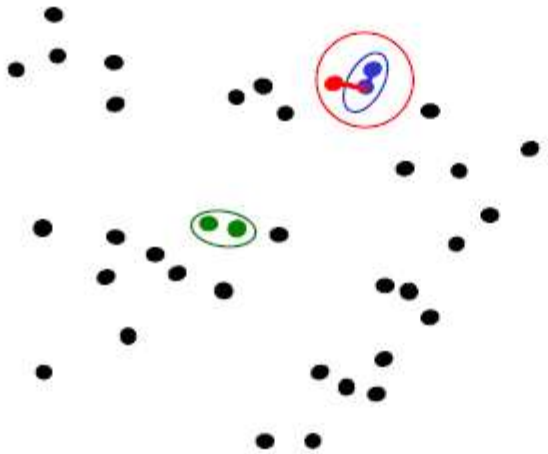
1. Say "Every point is its own cluster"
2. Find "most similar" pair of clusters
3. Merge it into a parent cluster

Agglomerative Clustering

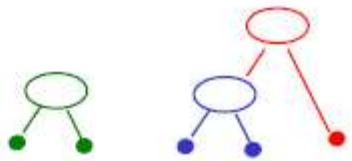


1. Say "Every point is its own cluster"
2. Find "most similar" pair of clusters
3. Merge it into a parent cluster
4. Repeat

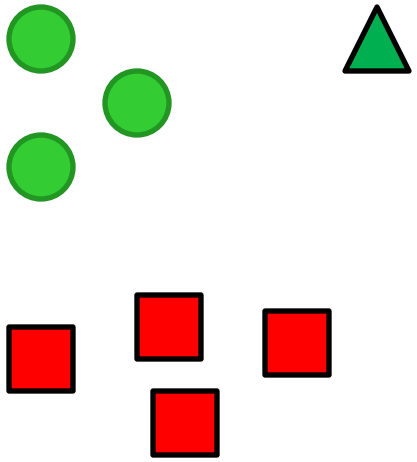
Agglomerative Clustering



1. Say "Every point is its own cluster"
2. Find "most similar" pair of clusters
3. Merge it into a parent cluster
4. Repeat



Agglomerative Clustering

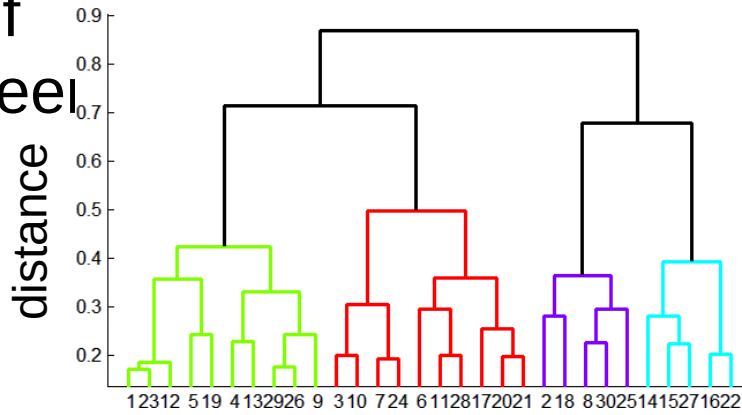


How to define cluster similarity?

- Average distance between points, maximum distance, minimum distance
- Distance between means or medoids

How many clusters?

- Clustering creates a dendrogram (a tree)
- Threshold based on max number of clusters or based on distance between merges



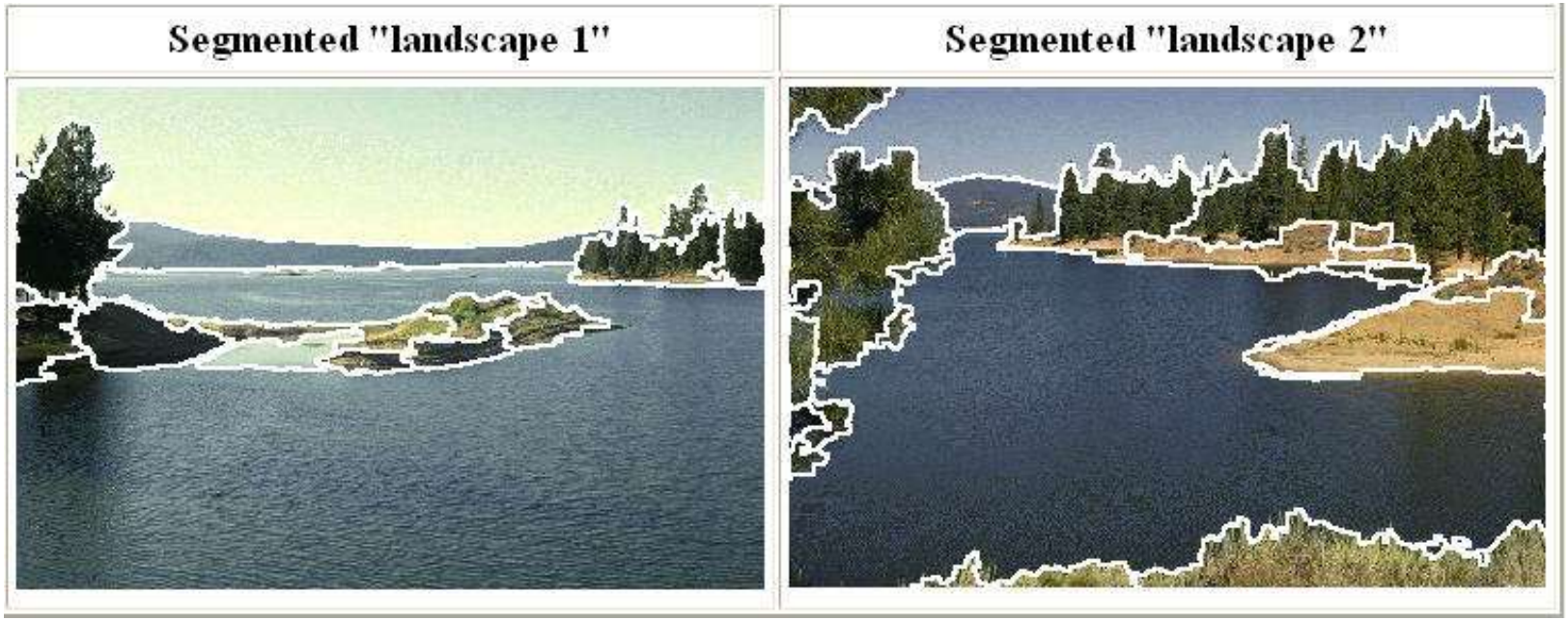
How do we Segment?

- K-means
  - Iteratively re-assign points to the nearest cluster center
- Split-and-Merge
  - Iteratively split and merge regions
- Agglomerative clustering
  - Start with each point as its own cluster and iteratively merge the closest clusters
- Mean-shift clustering
  - Estimate modes of pdf
- Spectral clustering
  - Split the nodes in a graph based on assigned links with similarity weights

Mean Shift Segmentation

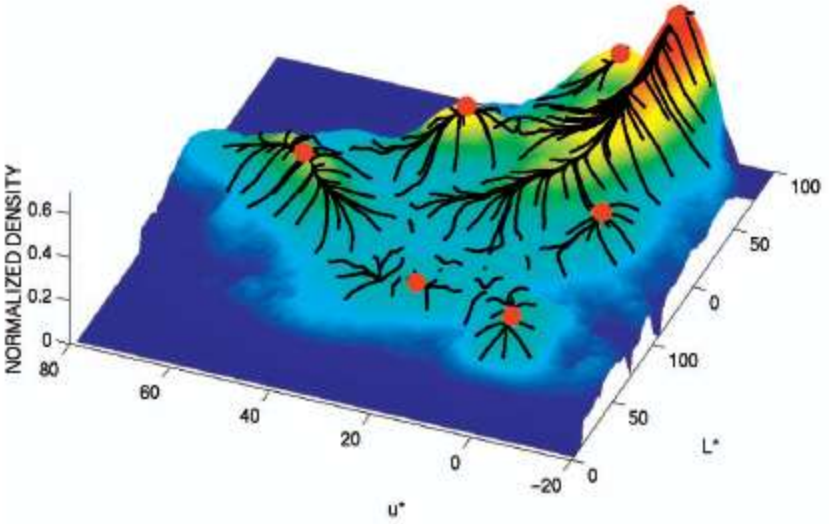
D. Comaniciu and P. Meer, Mean Shift: A Robust Approach toward Feature Space Analysis, PAMI 2002.

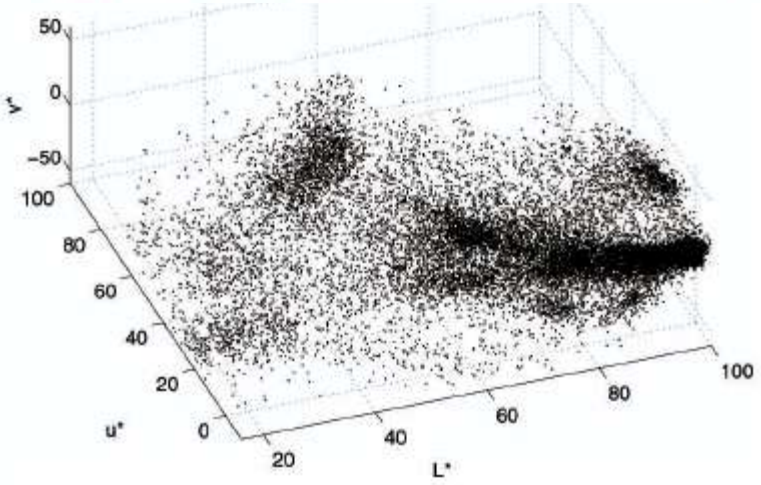
- Versatile technique for clustering-based segmentation

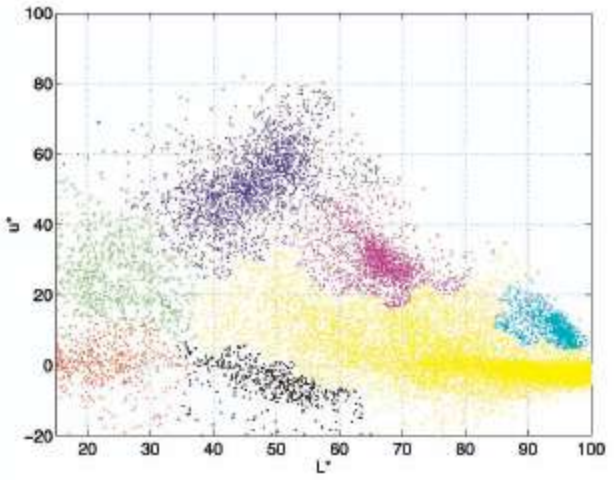


Mean Shift Algorithm

- Try to find *modes* of this non-parametric density







Mean shift clustering

- The mean shift algorithm seeks *modes* of the given set of points
  1. Choose kernel and bandwidth
  2. For each point:
     a) Center a window on that point
     b) Compute the mean of the data in the search window
     c) Center the search window at the new mean location
     d) Repeat (b,c) until convergence
  3. Assign points that lead to nearby modes to the same cluster

Kernel Density Estimation

Kernel density estimation function

$$\widehat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

Gaussian kernel

$$K\left(\frac{x - x_i}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2h^2}}.$$

Mean Shift Segmentation Results

http://www.caip.rutgers.edu/~comanici/MSPAMI/msPamiResults.html



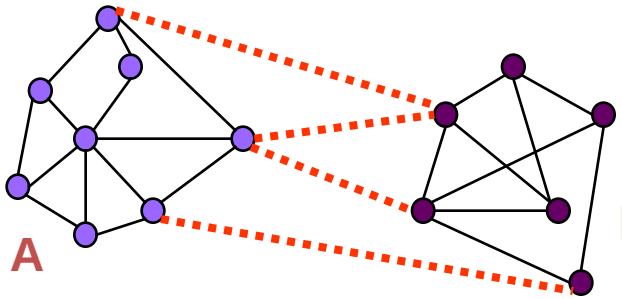

http://www.caip.rutgers.edu/~comanici/MSPAMI/msPamiResults.html

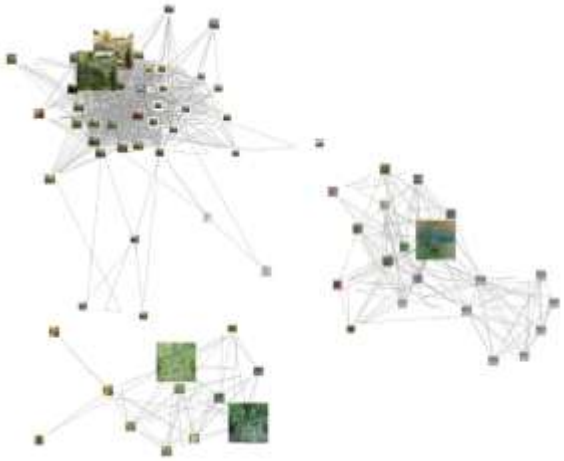
How do we Segment?

- K-means
  - Iteratively re-assign points to the nearest cluster center
- Split-and-Merge
  - Iteratively split and merge regions
- Agglomerative clustering
  - Start with each point as its own cluster and iteratively merge the closest clusters
- Mean-shift clustering
  - Estimate modes of pdf
- Spectral clustering
  - Split the nodes in a graph based on assigned links with similarity weights

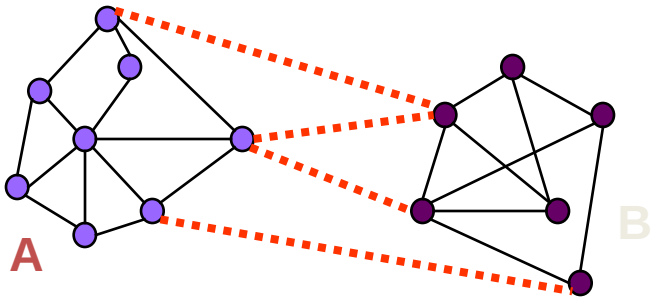
Spectral Clustering

Group points based on links in a graph





Cuts in a Graph



Normalized Cut

- a cut penalizes large segments
- fix by normalizing for size of segments

$$Ncut(A,B) = \frac{cut(A,B)}{volume(A)} + \frac{cut(A,B)}{volume(B)}$$

- volume(A) = sum of costs of all edges that touch A

Source: Seitz

Normalized Cuts for Segmentation

Selective Search: Approach

- Hypotheses based on hierarchical grouping

Selective Search: Approach

- Hypotheses based on hierarchical grouping

Ground truth

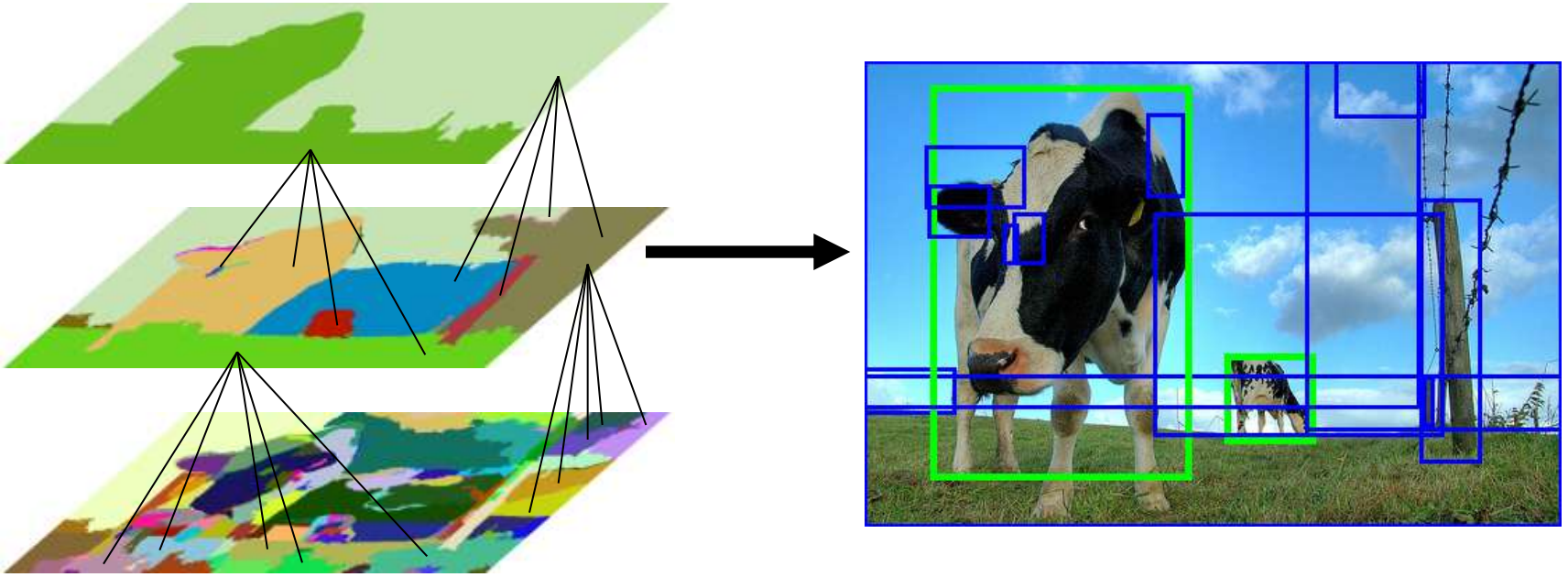
Selective Search: Approach

- Hypotheses based on hierarchical grouping

Initial segments from oversegmentation [Felzenszwalb2004]

Selective Search: Approach

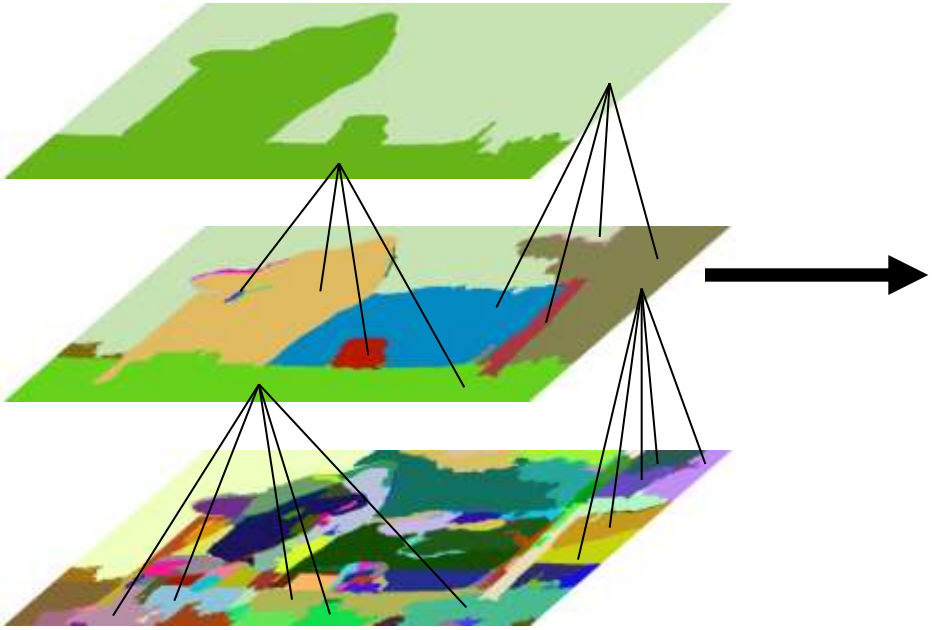
- Hypotheses based on hierarchical grouping

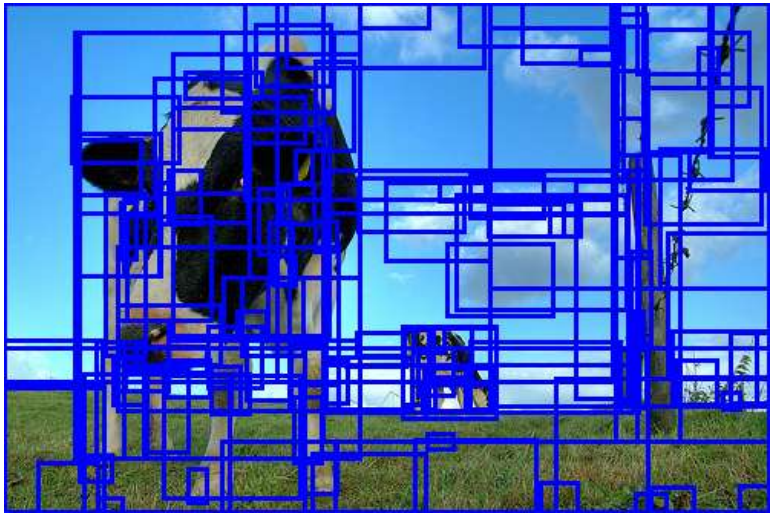


Group adjacent regions on color/texture cues

Selective Search: Approach

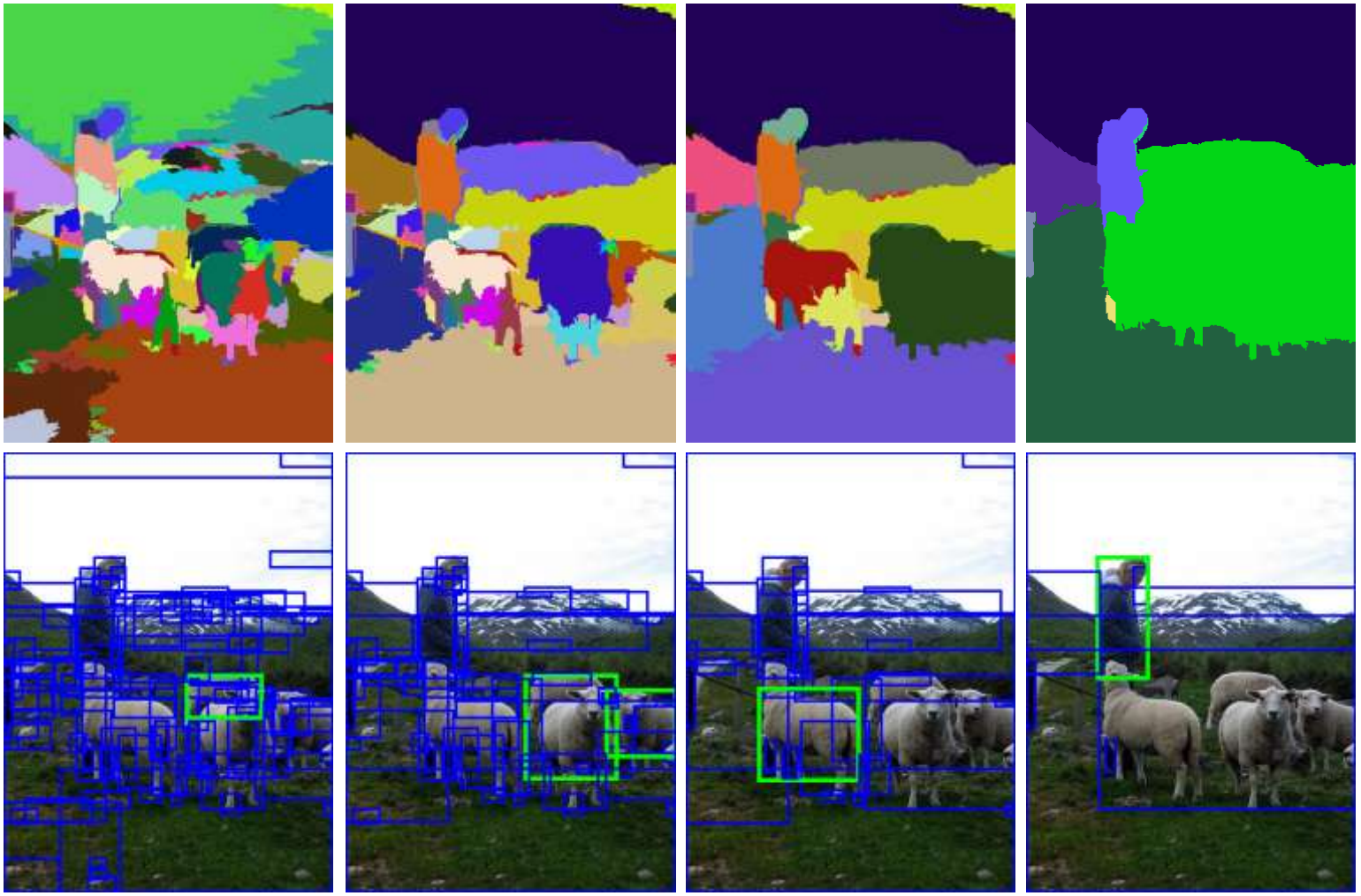
- Hypotheses based on hierarchical grouping





Object hypotheses from
all hierarchy levels

Example 1



Example 2

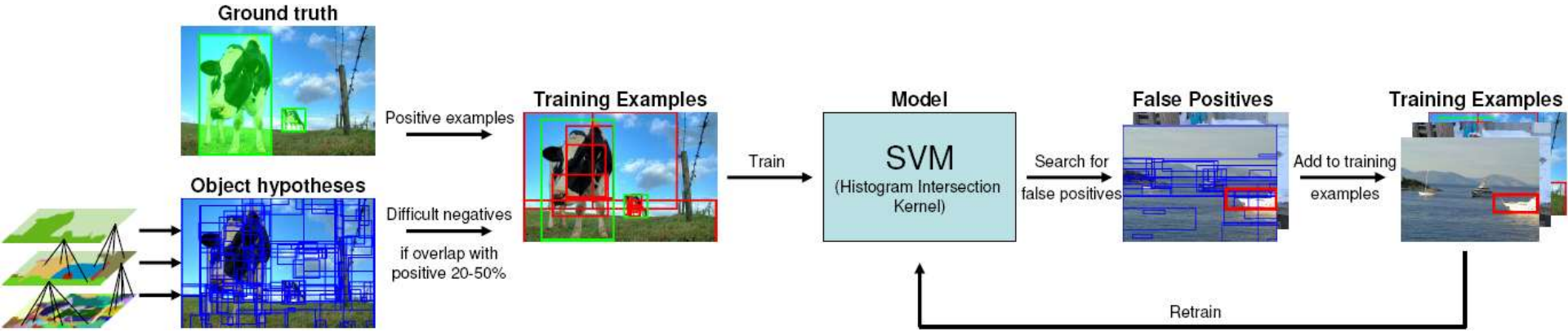
Selective Search on ILSVRC2011

- Apply to ILSVRC2011 train set
- Object hypotheses are class-independent

	ILSVRC2011 train
With bounding box annotations	315,525 images
Average #boxes/image	1,565
Average recall	98.5%

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

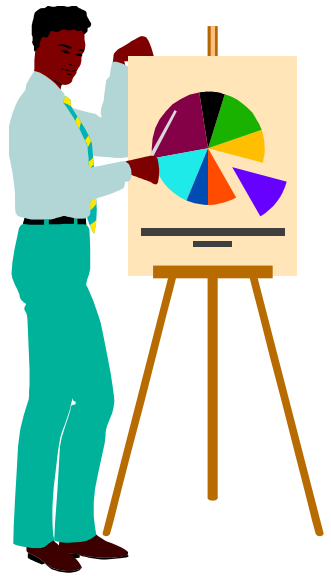
Summary

- Adopted segmentation as selective search strategy for object localisation:
  - High recall: >96% with ~1,500 locations
  - Coarse locations are sufficient: bounding boxes
  - Fast to compute: <10s per image
  - Class-independent
  - Enables the use of bag-of-words features

Today's class

Part I: Visual Attention: Object Localization

Part II: Affective Computing (Hamdi Dibeklioglu)



Part II:
Face Analysis for Affective Computing

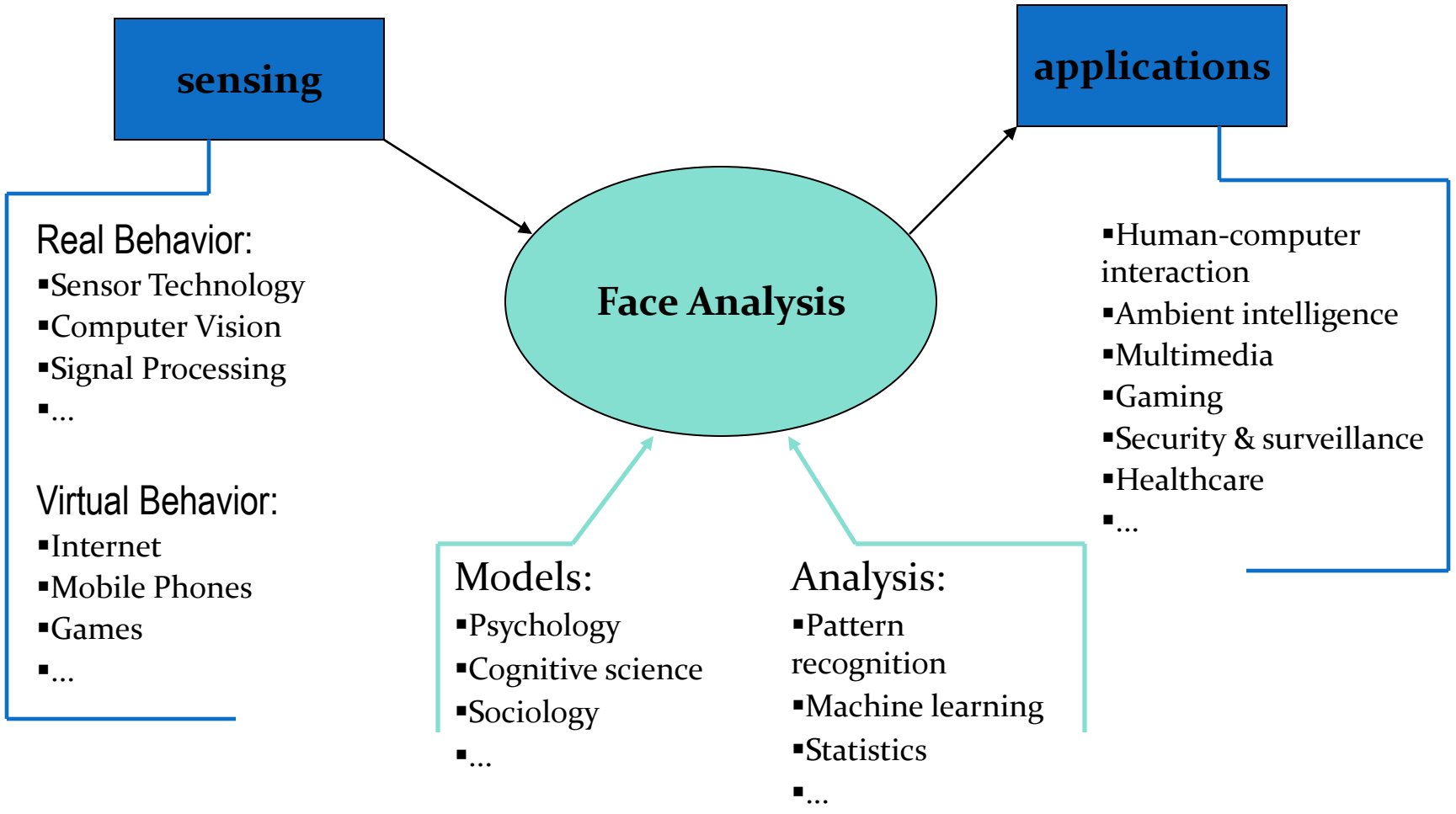
Outline

- Why & Where
- Facial Expression Analysis
- Spontaneous versus Posed Smile Classification
- Affective Human-Computer Interaction
- A New Smile Database: UvA-NEMO

Why & Where?

- Human-computer interaction: better interaction
- Ambient intelligence: smart environments
- Cognitive sciences: quantitative evaluation tools
- Education: responsive teaching and coaching systems
- Entertainment: more engaging games
- Healthcare: ambient assisted living
- Photography: cameras that click when everybody smiles

Computer analysis of face

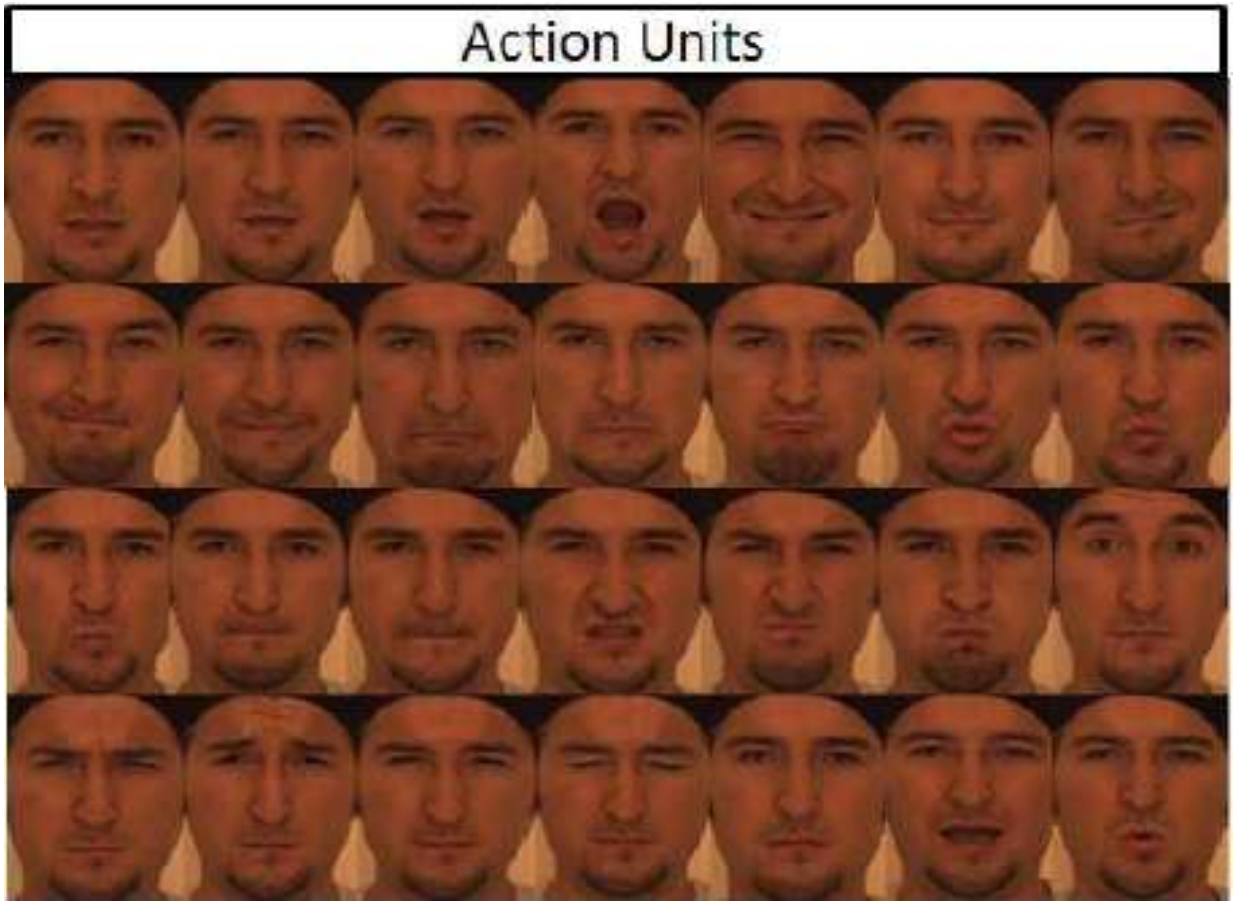


Facial Expression Analysis

Basic Facial Expression Categories

- Paul Ekman's basic facial expression categories:
  - Happiness, surprise, sadness, fear, anger, disgust

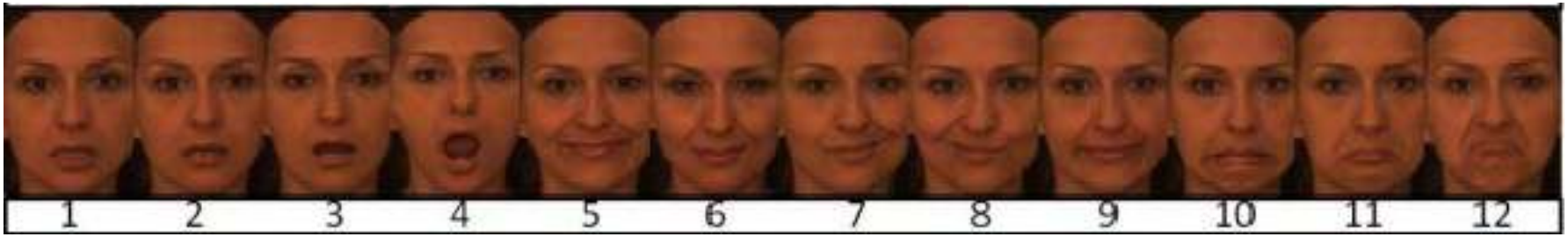
Facial Action Coding System



Ekman and Friesen, 1978

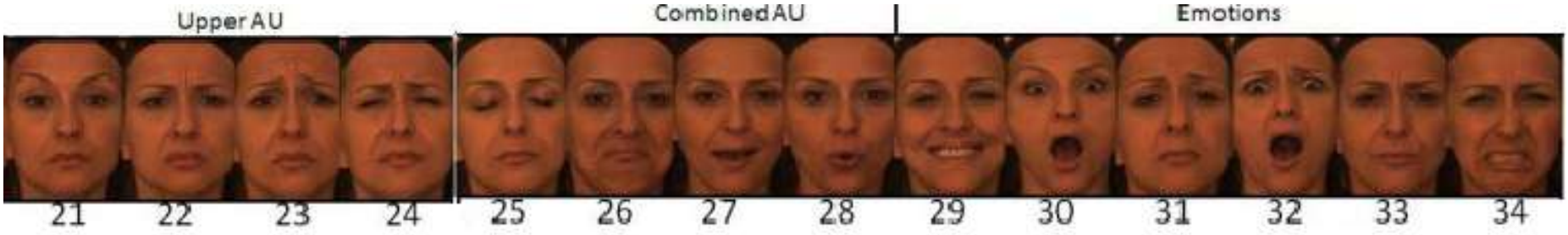
Lower Action Units

Expressions	Scan No	Explanation
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

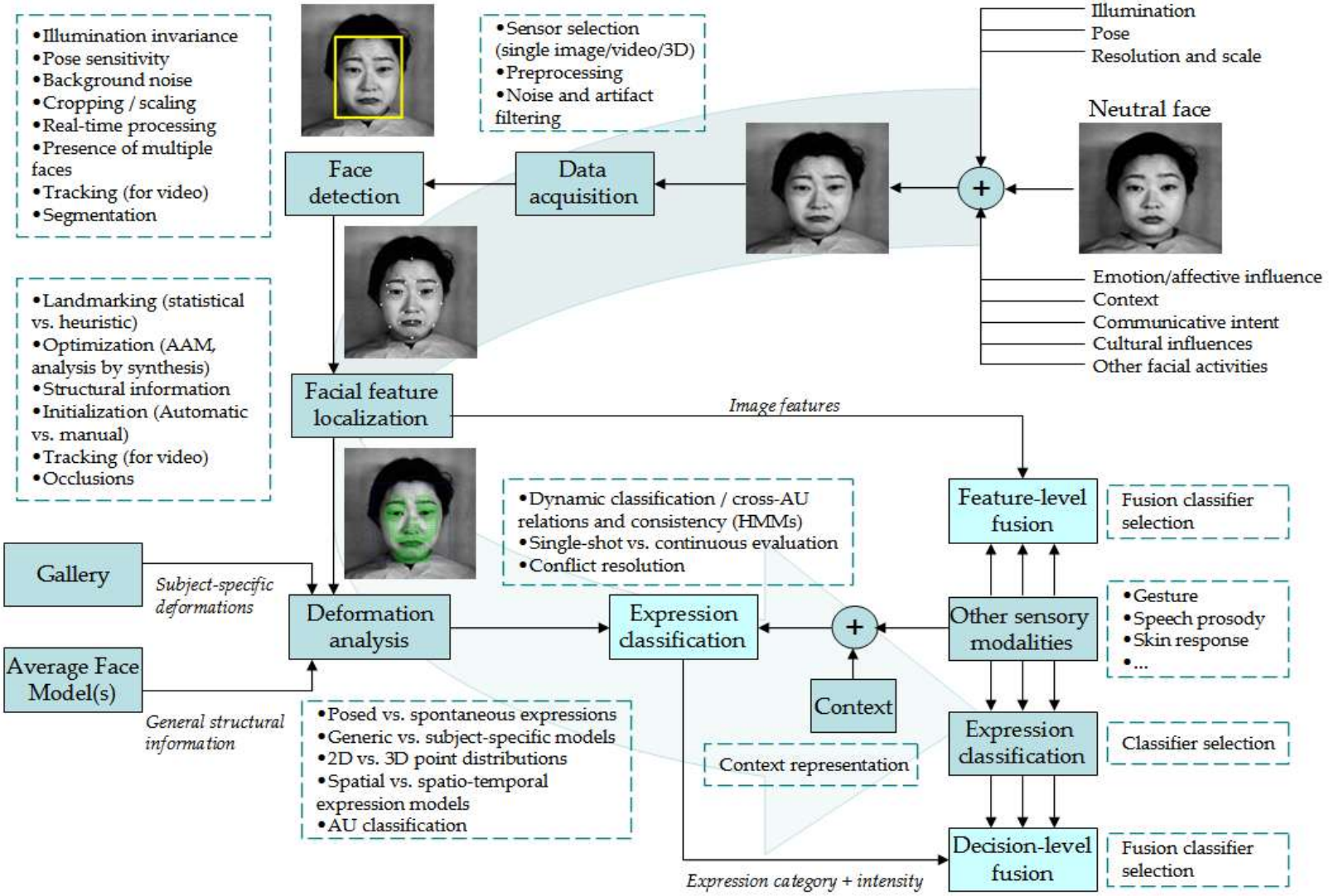


Upper/Combined AUs + Basic Expressions

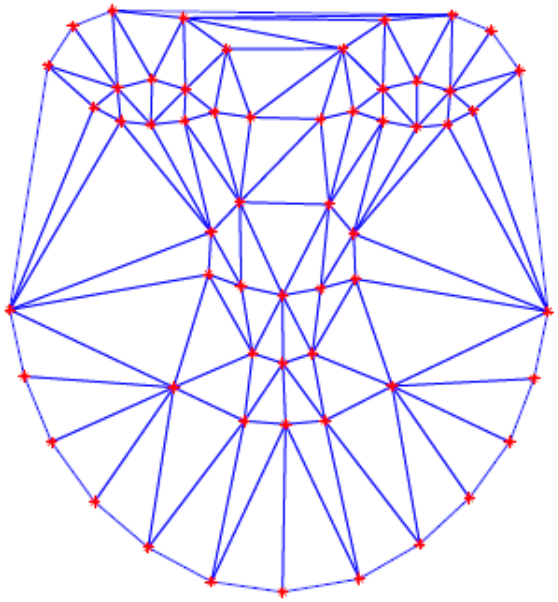
Expressions	Scan No	Explanation
Upper AUs	21	Outer Brow Raiser - AU2
	22	Brow Lowerer - AU4
	23	Inner Brow Raiser - AU1
	24	Squint - AU44
	25	Eyes Closed - AU43
Combined AUs	26	Jaw Drop (26) + Low Intensity Lip Corner Puller
	27	Lip Funneler (22) + Lips Part (25)
	28	Lip Corner Puller (12) + Lip Corner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



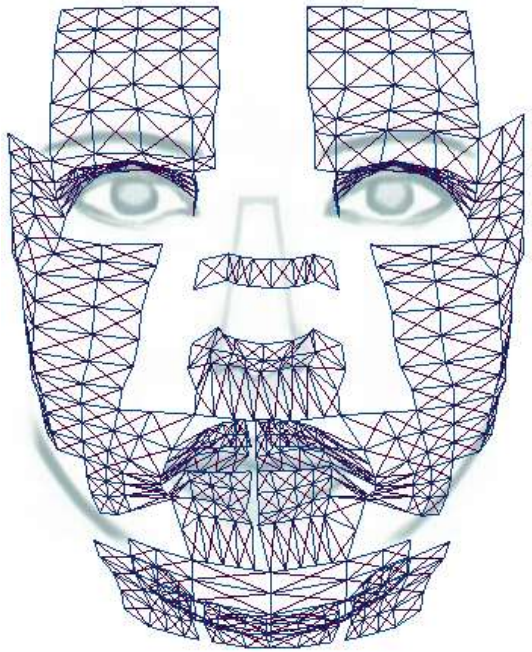
Facial Expression Analysis



Choosing a Representation



a)

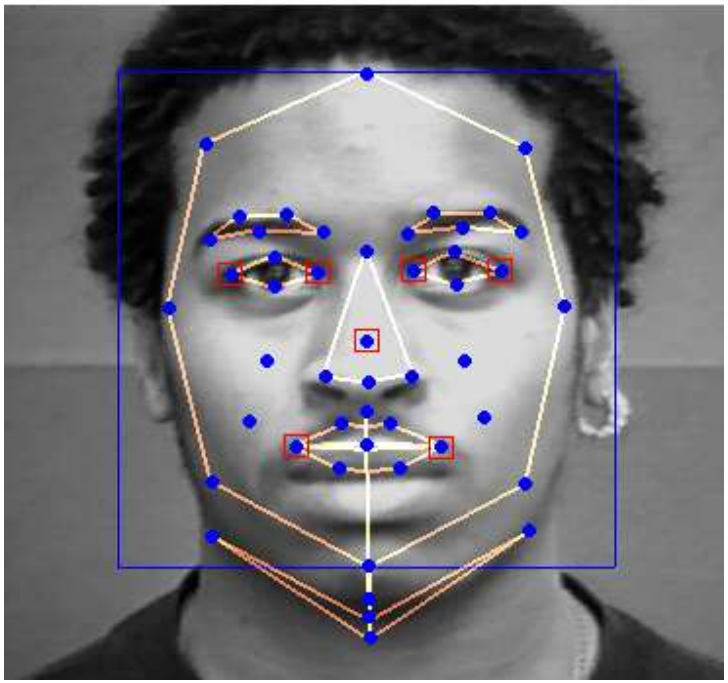


b)

(a) Landmarks and corresponding triangular mesh model for AAM (Stegmann, 2002).
(b) The wireframe model used in (Cohen et al., 2003)

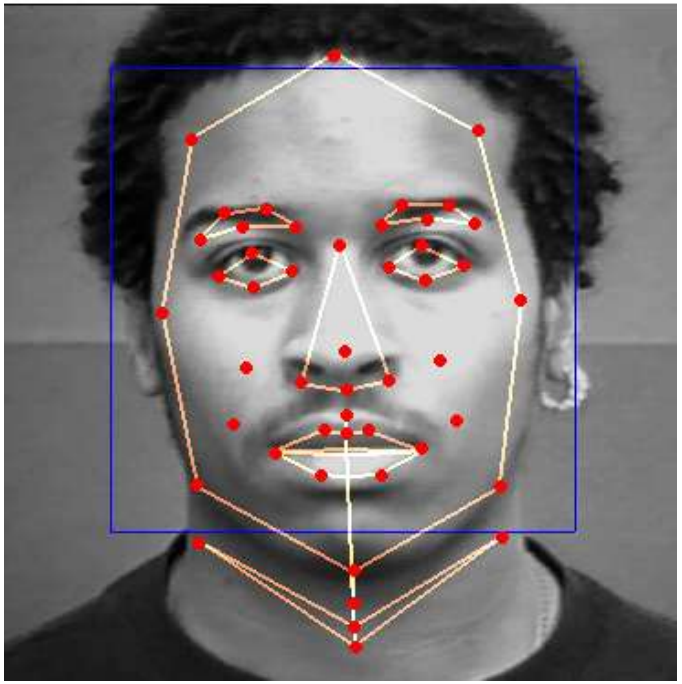
Tracker Initialization

Landmark Detection



Generic Model

TPS

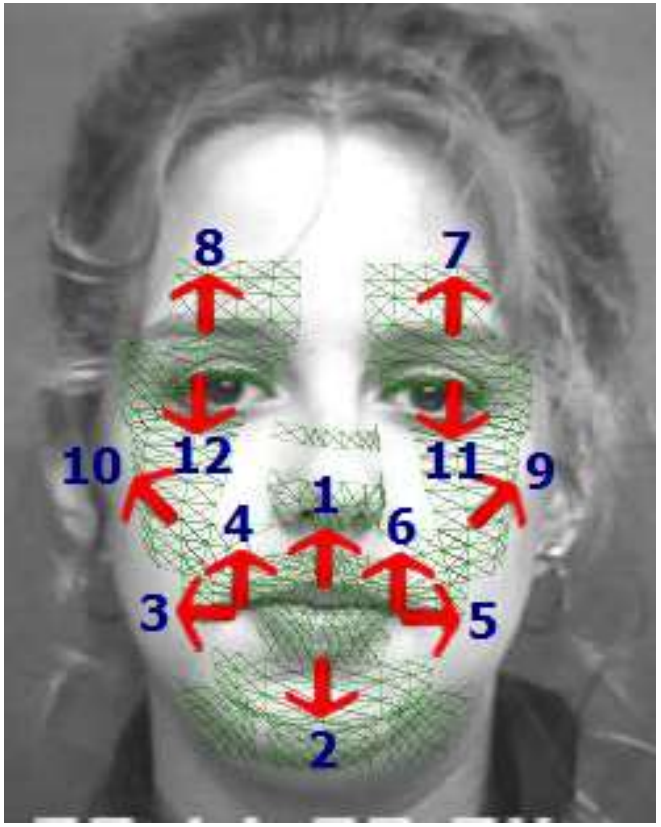


Warped Model

Facial Expression Recognition

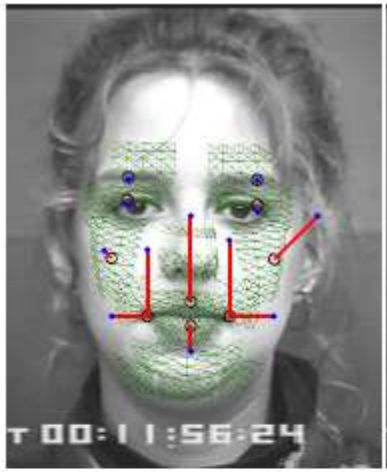
- Face model: 16 surface patches embedded in Bezier volumes.
- Piecewise Bezier Volume Deformation (PBVD) tracker is used to trace the motion of the facial features.

Facial Expression Recognition



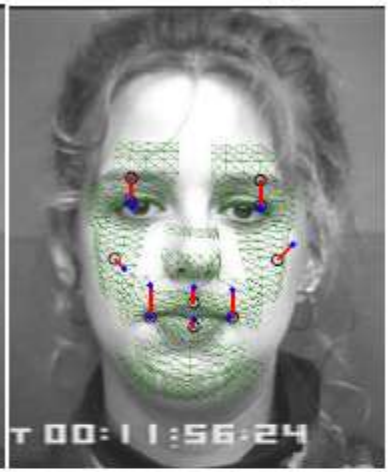
- 12 motion units
- Naive Bayes (NB) classifier for categorizing expressions
- NB Advantage: the posterior probabilities allow a soft output of the system

Average Motion Units

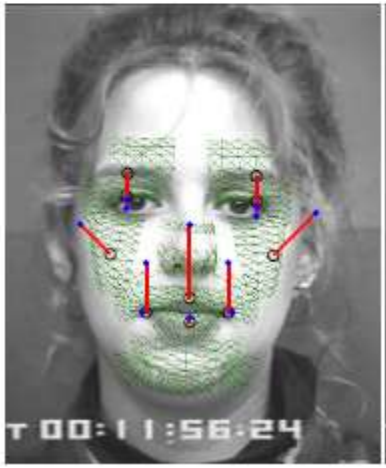


Happiness

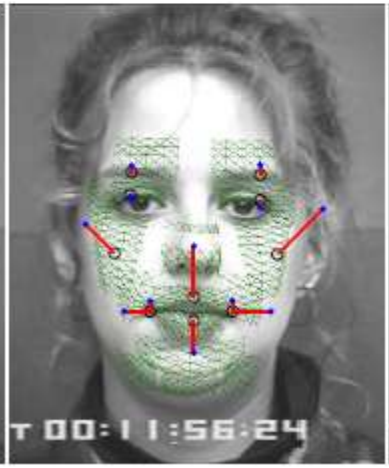
Surprise



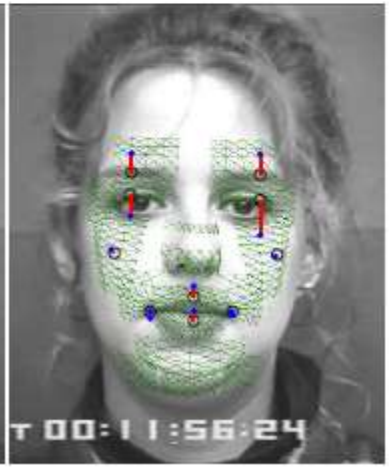
Anger



Disgust

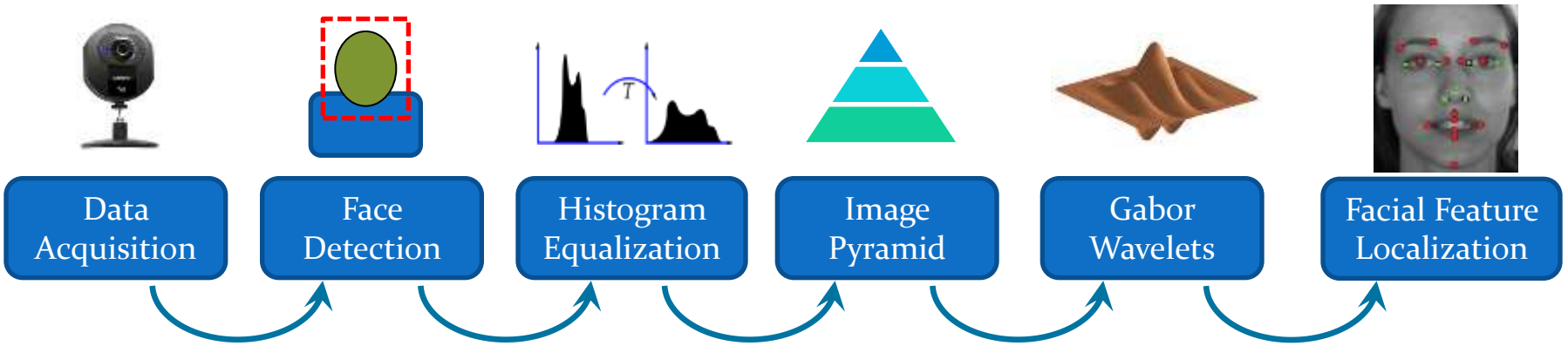


Fear



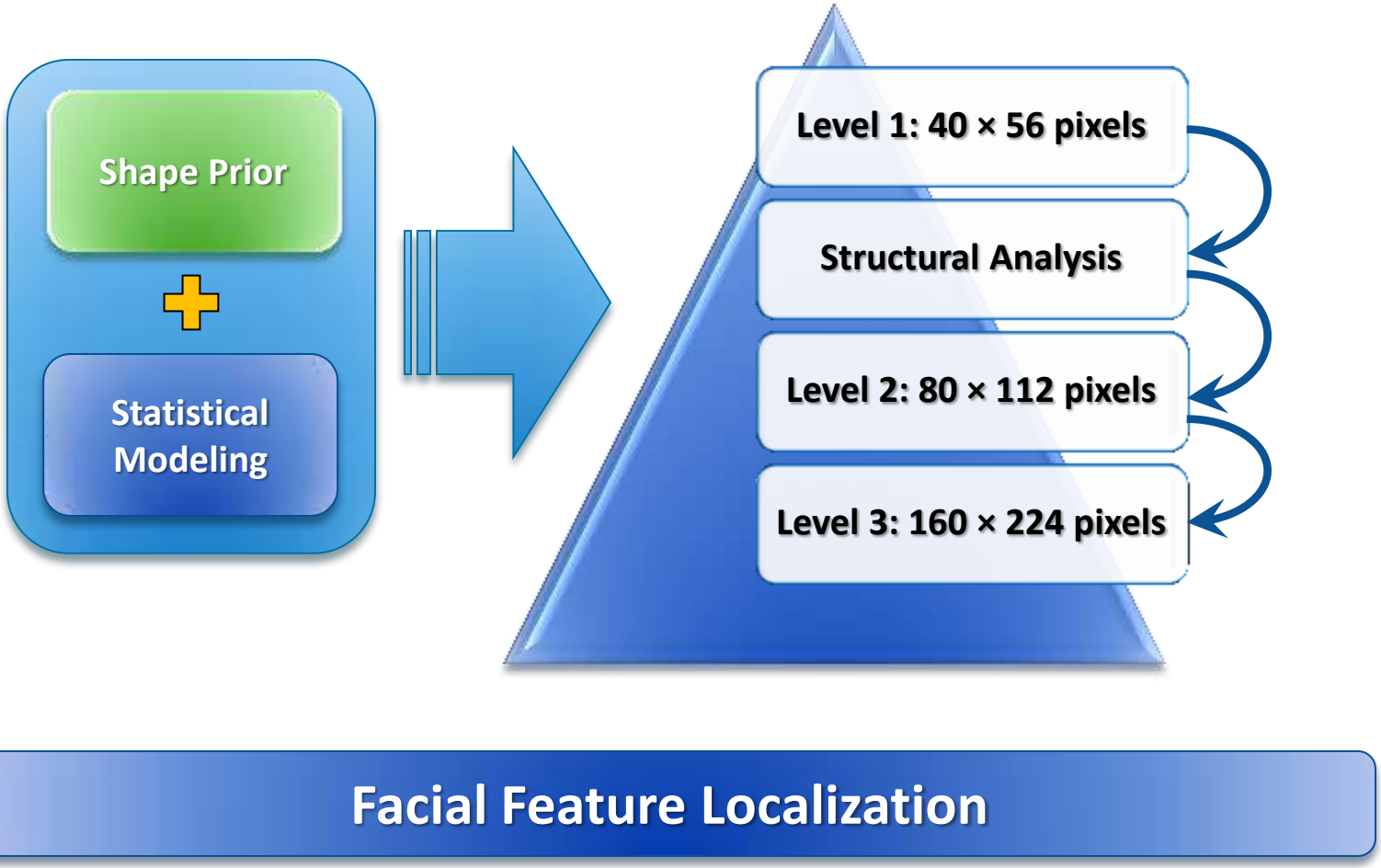
Sadness

Facial Feature Localization

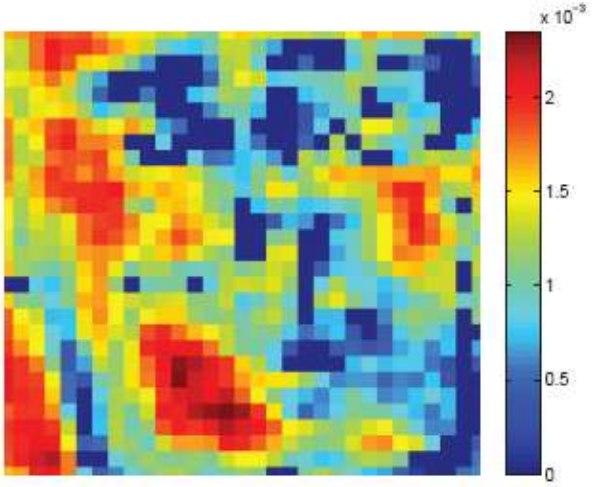


* H. Dibeklioğlu, A.A. Salah, T. Gevers.
A Statistical Method for Facial Landmarking. *IEEE Trans. Image Processing*, 2012.

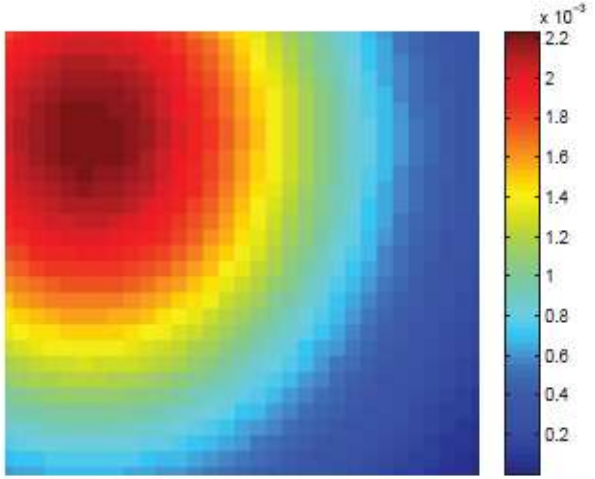
Facial Feature Localization



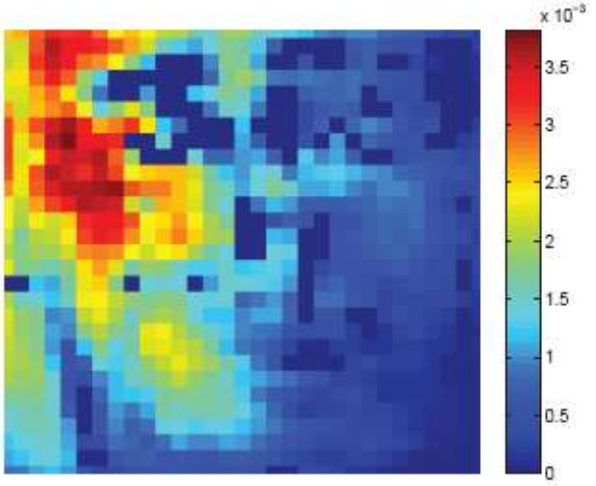
Facial Feature Localization



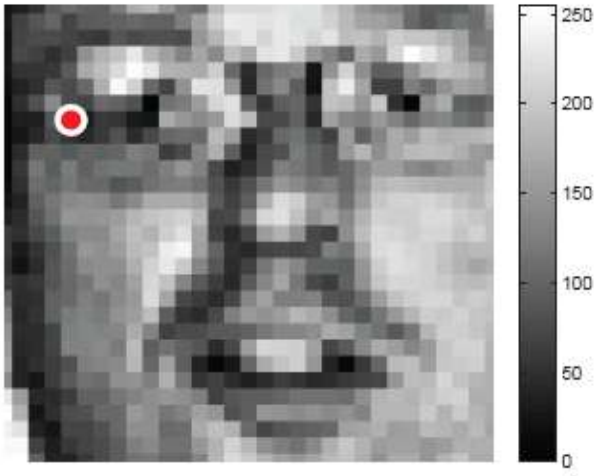
Likelihood-based Probability Map



Shape Prior

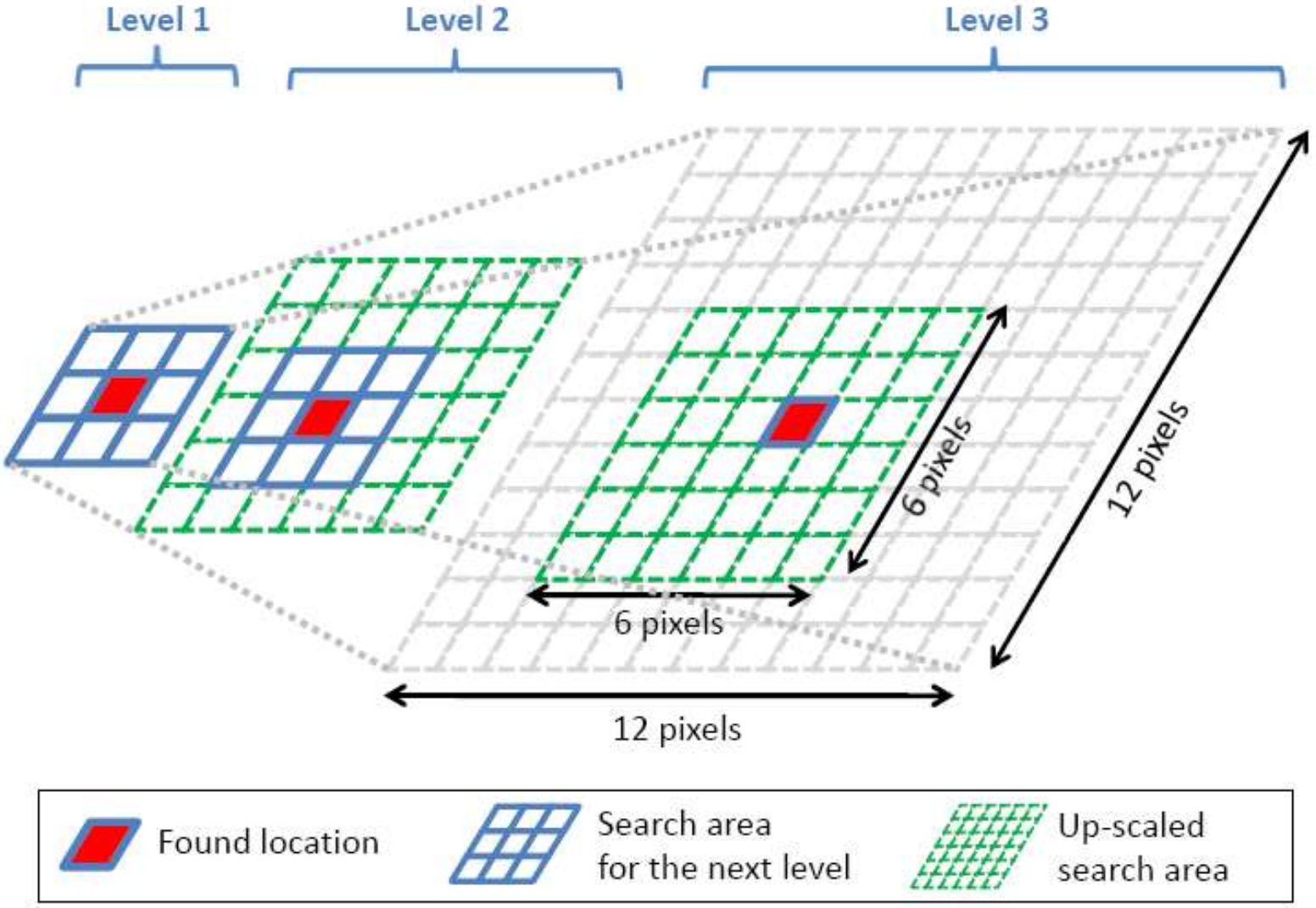


Combination Map



Coarse Level Localization

Facial Feature Localization

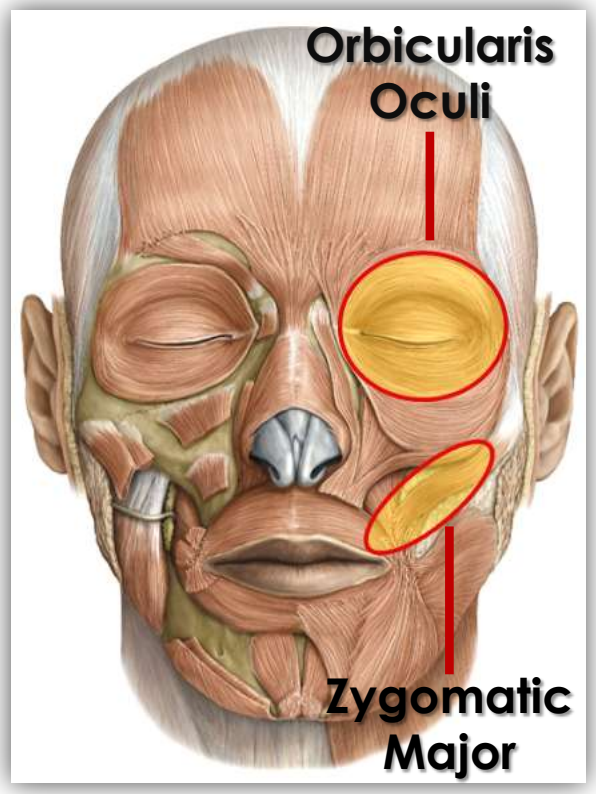


Spontaneous versus Posed Smiles

Enjoyment: Spontaneous or Posed?

Motivation

- It is said that eyes are the mirror of the soul.
- We believe that the state of mind of one person can be seen through his eyes.



- During a smile, *Zygomatic Major* raises the corners of the mouth.*
- In spontaneous smiles, eyelids are lowered and wrinkles (*Duchenne Marker*) are formed around eyes by *Orbicularis Oculi* muscle.*
- Duchenne Markers are spontaneous signs of felt enjoyment that cannot be feigned.**

*Ekman & Friesen, 1982.
**Ekman, 1985, 1989, 1993.

Motivation

- Some empirical studies show that *Orbicularis Oculi* can be active or inactive under both spontaneous and deliberate conditions*.

- However, there is also convincing evidence that the Duchenne Marker is consistently used by untrained people to *recognize* spontaneous and posed enjoyment smiles.**

Duchenne Marker

*Krumhuber & Manstead, 2009; Schmidt & Cohn, 2001; Schmidt et al., 2006, 2009.
**Frank et al., 1993; Williams et al., 2001; Miles & Johnston, 2006, 2007; Thibault et al., 2009.

Motivation

- Temporal parameters such as the relative durations and amplitudes of onset, apex, and offset phases of a smile are discriminative for classifying different spontaneous/posed smiles.

*Cohn & Schmidt, 2004; Schmidt et al., 2006, 2009.

Literature

Cohn & Schmidt *[Int. Journal of Wavelets, Multiresolution and Information Processing '04]*	• Duration, amplitude, and $\frac{\text{duration}}{\text{amplitute}}$ measures of smile onsets.
Valstar & Pantic [ICMI '07]	• Fusion of shoulder, head and inner facial movements.
Dibeklioğlu et al. [ACM Multimedia'10]	• Only eyelid movements.

Smiles

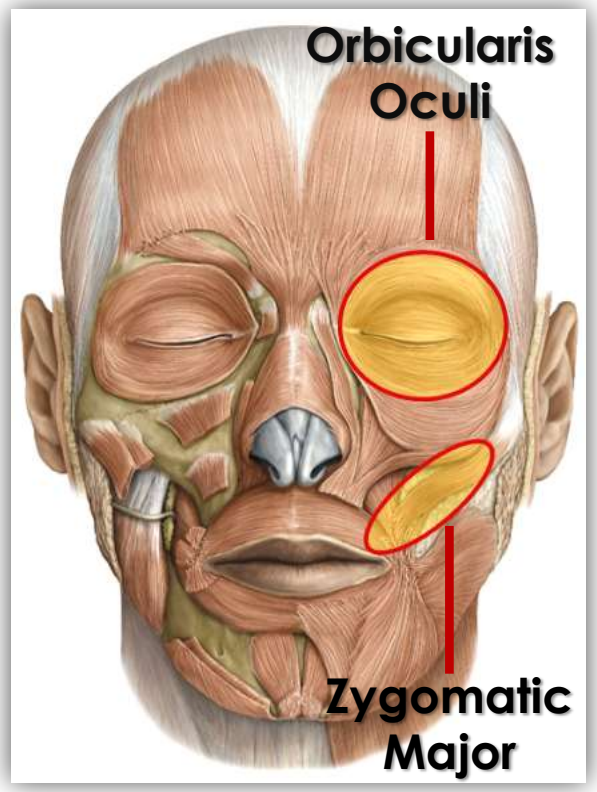
Duchenne and his patient

- In the mid-nineteenth century, Guillaume Duchenne proposed that smiles resulting from true happiness not only utilize the muscles of the mouth but also those of the eyes.
- Such “genuine" smiles are known as Duchenne smiles in his honor.
- Researchers analysed Duchenne’s observations and find a strong correlation with felt enjoyment smiles after 120 years.*
- Ekman described 18 types of different felt smiles such as enjoyment, fear, miserable, embarrassment, *listener response smiles*.**

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

** P. Ekman. Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage. W.W. Norton, New York, 1985.

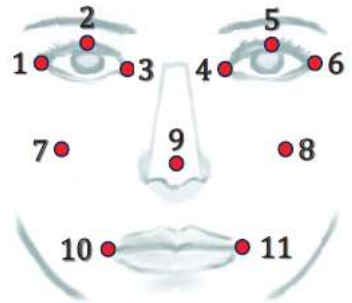
Facial Muscles



- Zygomatic major raises the corners of the mouth.
- Orbicularis oculi muscle raises the cheeks and forms crows-feet around the eyes.
- Activation of Orbicularis oculi also lowers the eyelids.

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

Features

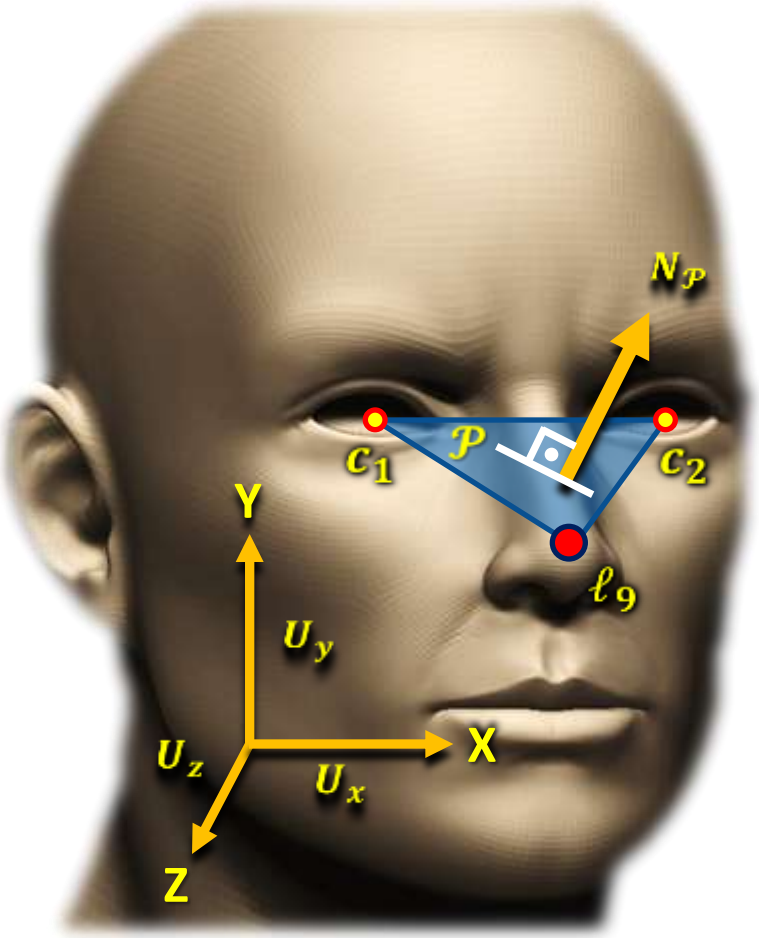


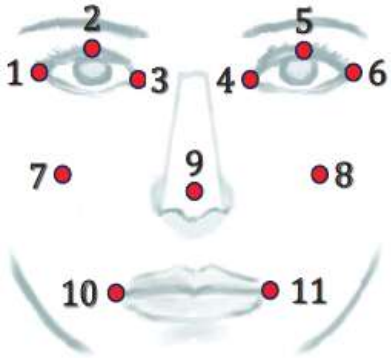
- 11 points are detected* automatically and tracked on three regions:
  - Eye region: eyelids and eye corners,
  - Cheek region,
  - Mouth region: lip corners.
- Piecewise Bézier Volume Deformation tracker is used to trace the 3d motion of the facial features.**

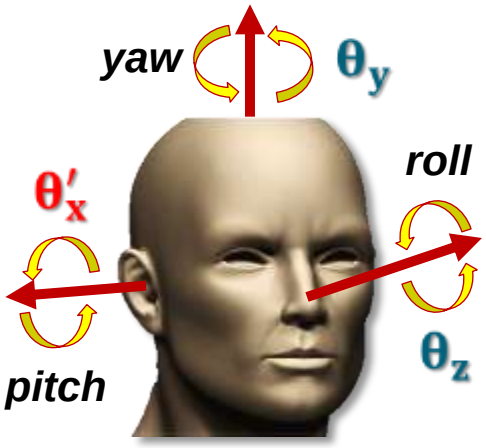
*H. Dibeklioglu, A.A. Salah, and T. Gevers. A statistical method for 2-d facial landmarking. *IEEE Trans. Image Processing*, 21(2):844–858, 2012.

**H. Tao and T. Huang. Connected vibrations: A modal analysis approach for non-rigid motion tracking. In Proc. CVPR, 735–740, 1998.

Face Alignment and Normalization







$$\theta = \cos^{-1}\frac{U.N_{\mathcal{P}}}{\|U\|\|N_{\mathcal{P}}\|}, N_{\mathcal{P}} = \overrightarrow{\ell_9 c_2} \times \overrightarrow{\ell_9 c_1}$$

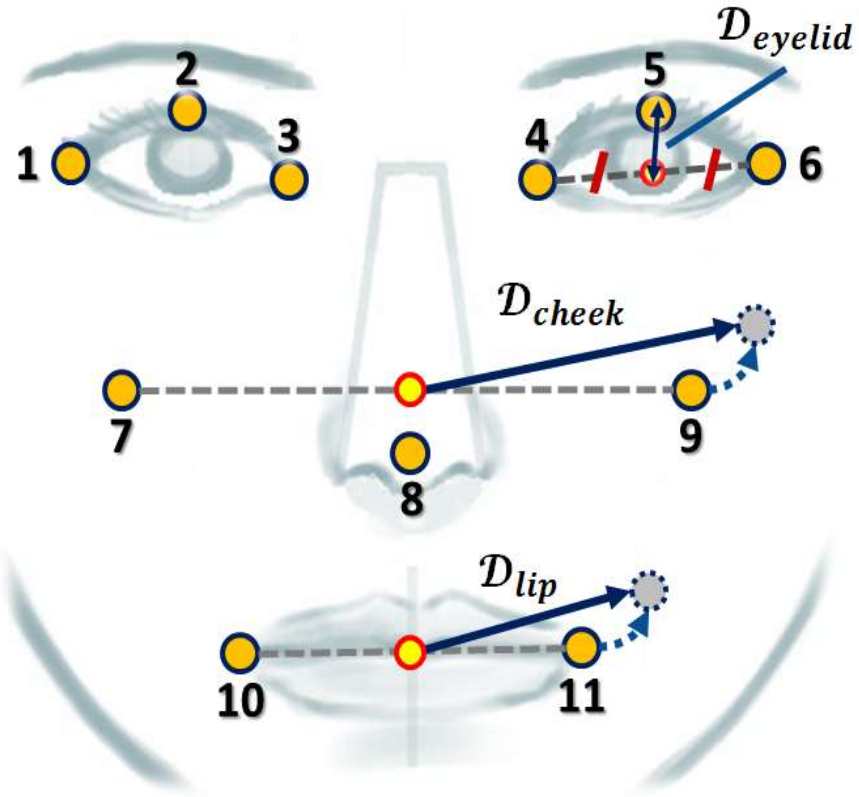
$$c_1 = \frac{\ell_1 + \ell_3}{2},\ c_2 = \frac{\ell_4 + \ell_6}{2}$$

$$\ell'_{\mathrm{i}} = \left[\ell_{\mathrm{i}} - \frac{c_1 + c_2}{2}\right] R_x(-\theta'_x) R_y(-\theta_y)\ R_z(-\theta_z) - \frac{100}{\rho(c_1, c_2)}$$

$$\theta'_x = \theta_x - \theta_x^{t=1}$$

Dynamic Features

- Using the tracked points, amplitude ($\mathcal{D}$), speed ($\mathcal{V}$), and acceleration ($\mathcal{A}$) signals are extracted.



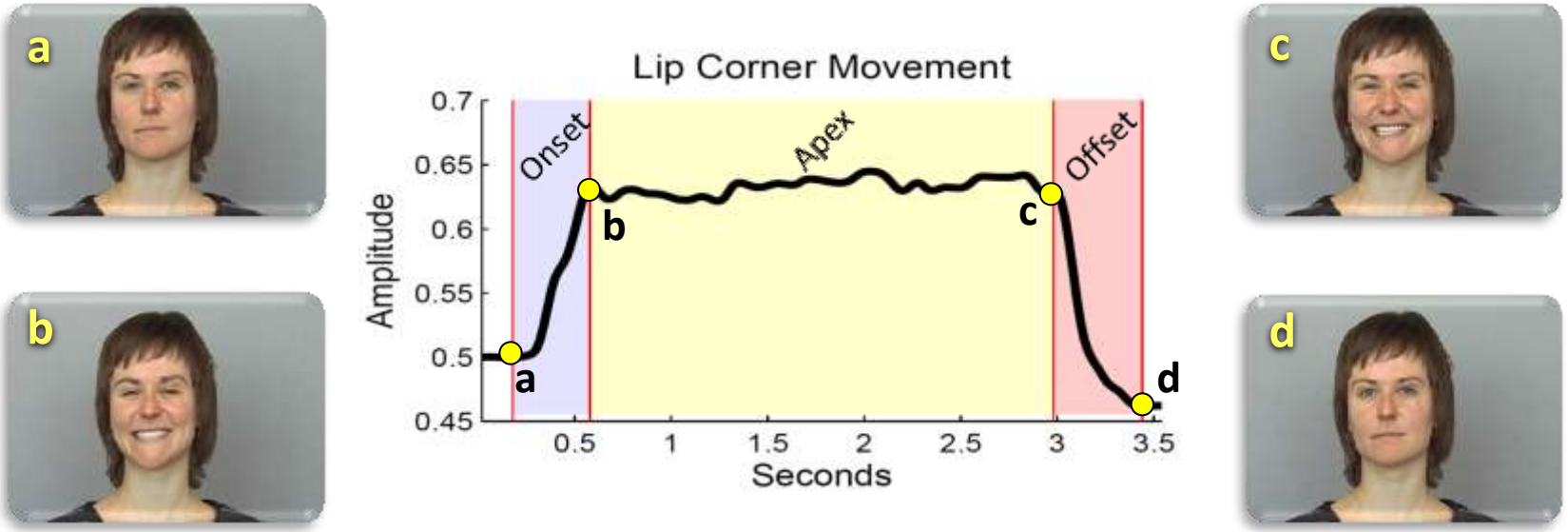
*H. Dibeklioglu, A.A. Salah, and T. Gevers. Are you really smiling at me? Spontaneous versus posed smiles. In Proc. ECCV 2012 (3):525-538, 2012.

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

Temporal Phase Segmentation

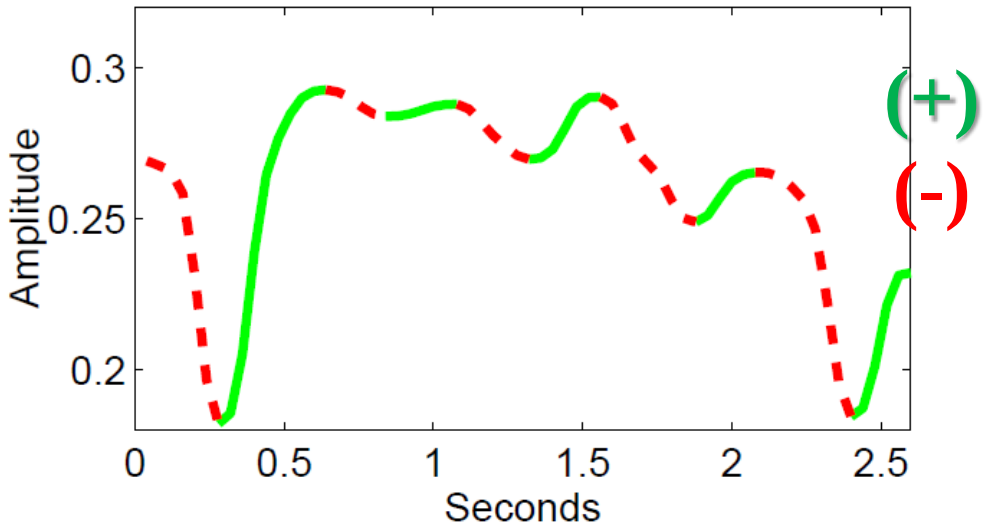
- Onset, apex, and offset segmentation*:
  - Longest continuous increase is selected as onset.
  - Longest continuous decrease is selected as offset.



*J. F. Cohn and K. Schmidt. *Int. Journal of Wavelets, Multiresolution and Information Processing*, 2004.
*K. Schmidt et al., *Biol. Psychology,* 2003.

Positive/Negative Sub-segments

- For more detailed analysis, we propose to use sub-segments of the signals:
  - Positive sub-segments: Increasing durations (+)
  - Negative sub-segments: Decreasing durations (-)



*H. Dibeklioglu, A.A. Salah, and T. Gevers. Are you really smiling at me? Spontaneous versus posed smiles. In Proc. ECCV 2012 (3):525-538, 2012.

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

- $\mathcal{D}$: 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.