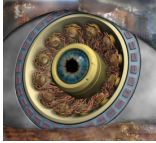


Computer Vision



Emile A. Hendriks
Information and Communication Theory group
Faculty of EEMCS
TU Delft

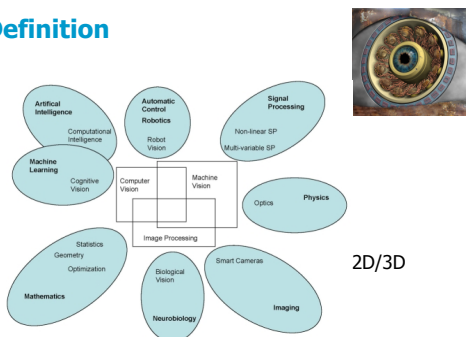


Content

- What is Computer Vision, Applications
- Strong and weak points
- Basic Modules of a Computer Vision framework
- Computer Vision Example



Definition

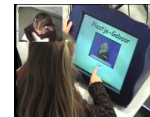


2D/3D



Applications

- Human/Animal behavior observation
- Manufacturing
- Video surveillance
- Domotics
- Multimedia semantic annotation and indexing
- Human-computer interfaces
- Affective computing
- Learning
-

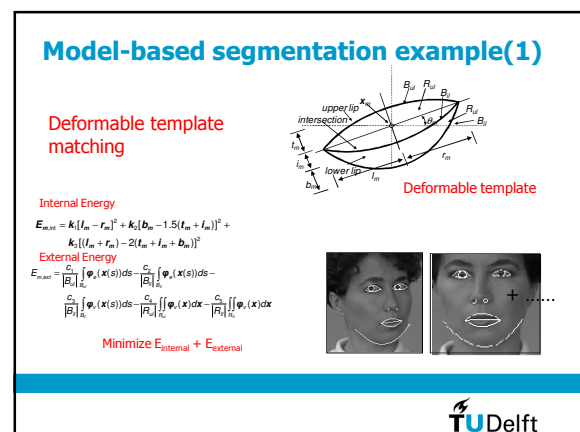
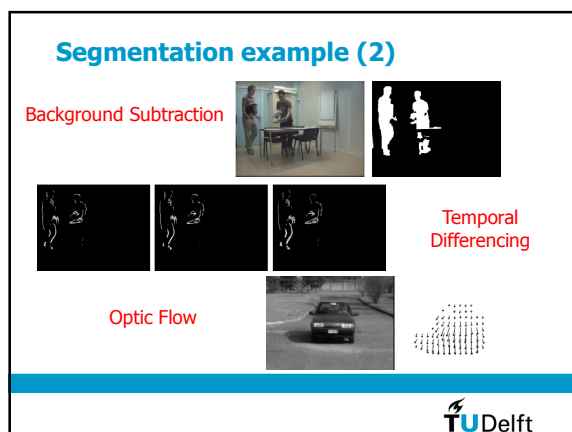
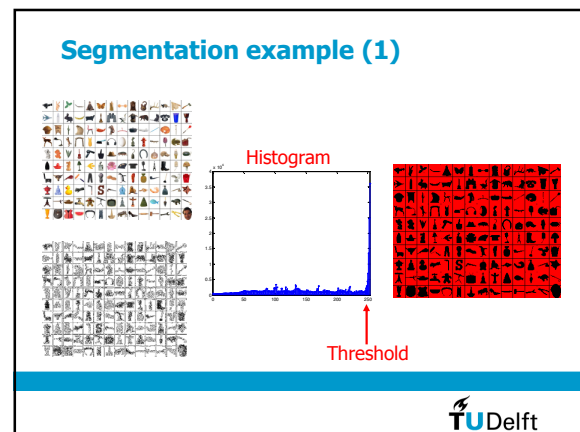
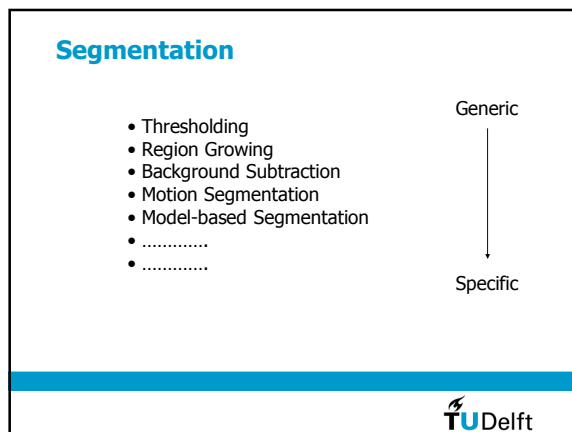
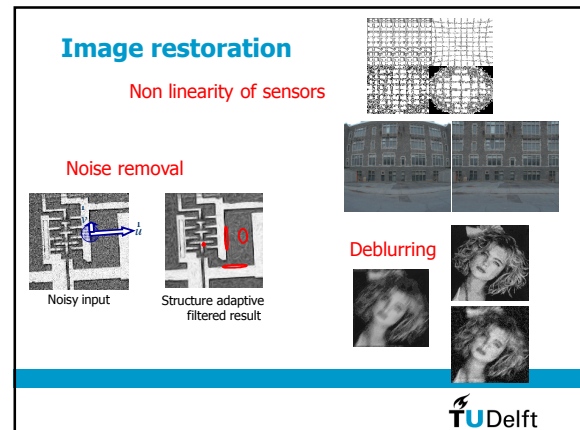
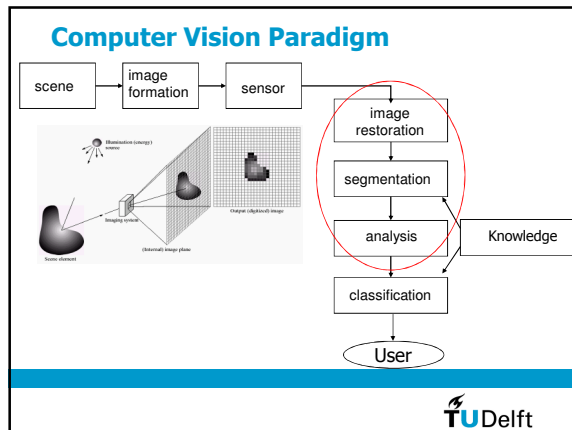


How many Faces??



Size of the objects??

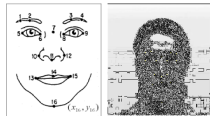




Model-based Segmentation example(2)

Active Shape Models

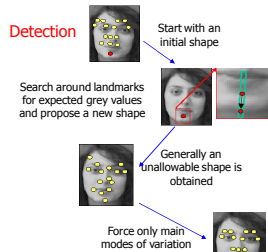
Shapes are represented by landmarks



Training



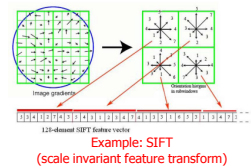
Use training set to find allowable modes of variations



Analysis

Feature detection:

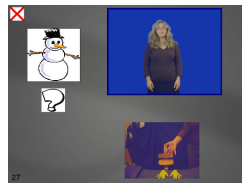
- interest point detection
- feature description



Example: Harris detector

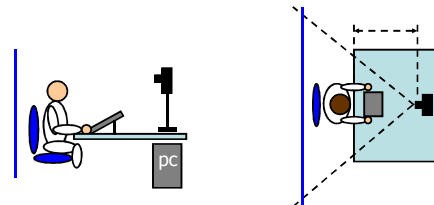
Computer Vision Application: Dutch Sign Language Recognition

- Deaf children age 3-5 years
- Lack of exposure to (sign) language at home
- Exposure can be increased with an electronic learning environment
- To practice sign production the computer must be able to give feedback
- Automatic sign recognition to facilitate learning

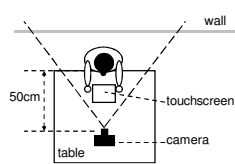


ELO: Electronic Learning environment

- Chair
- Table
- Touch screen
- Camera at eye-level
- Personal computer



Limitations of Recognition System



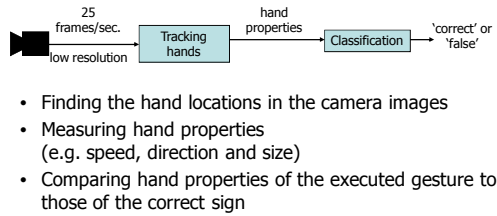
General limitations

- Facial expression is excluded
- Hands on the table before and after a sign
- No other persons can be nearby
- Only frontal view

Limitations of Recognition System

- Uniform wall behind the child
- Long sleeves only
- No loose hair visible
- Hand shape is not evaluated
- Hands are 'lost' near the face, neck and each other
- Limited set of signs (50)

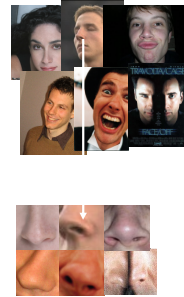
Description of Recognition System



- Finding the hand locations in the camera images
- Measuring hand properties (e.g. speed, direction and size)
- Comparing hand properties of the executed gesture to those of the correct sign

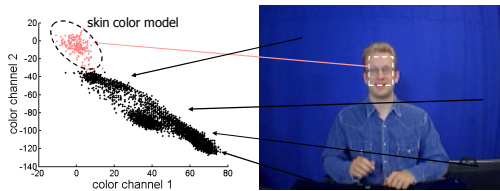
Finding Hands

- Hands are found by detection of skin color
- Skin color appearance in video is never the same
- Skin color is modeled from face color
- The face is found by its distinctive structure



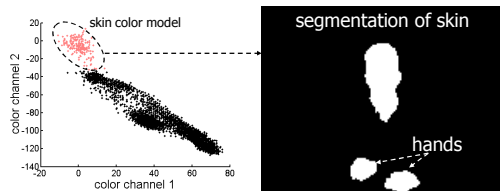
Finding Hands

- The face is found by its distinctive structure
- Skin color is modeled from face color



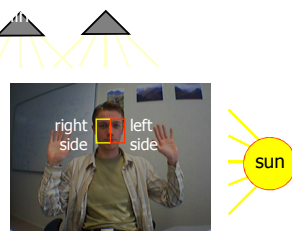
Finding Hands

- The face is found by its distinctive structure
- Skin color is modeled from face color



Finding Hands

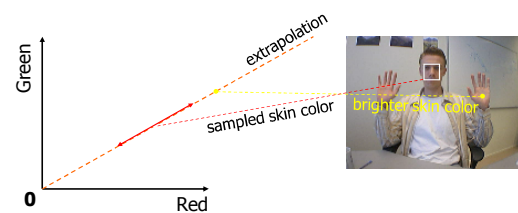
- Multiple light sources of different color
- Left and right side modeled separately



Shortcomings of Current Methods

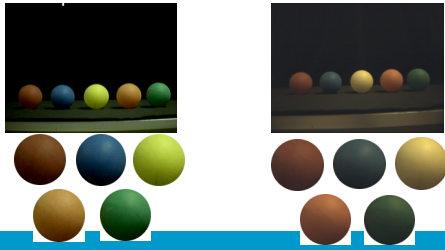
- Models for extrapolating color
- Hue Saturation Value (HSV)
 - normalized RGB

- Main assumption
- Black is represented by $[R,G,B]=0$



Shortcomings of Current Methods

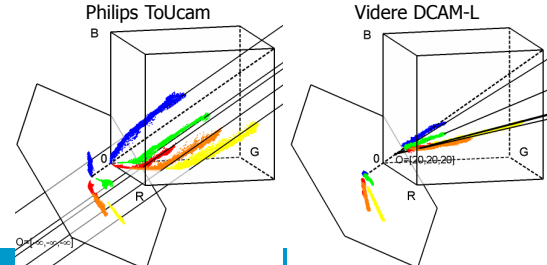
Example:



TU Delft

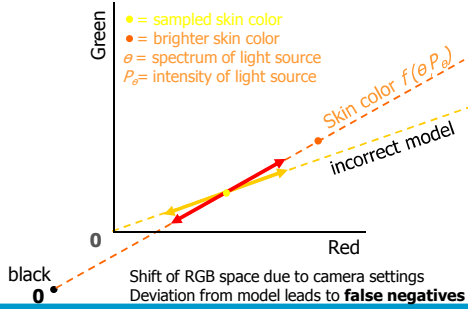
Shortcomings of Current Methods

Example:



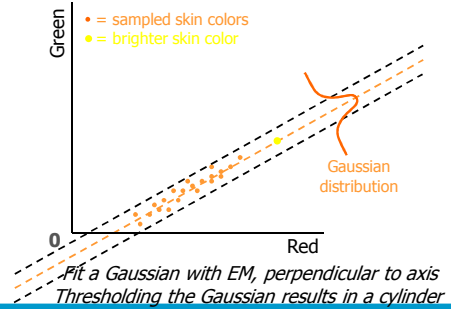
TU Delft

Shortcomings of Current Methods



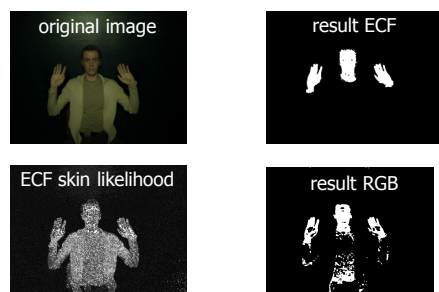
TU Delft

Solution: Elliptic Cylinder Fitting



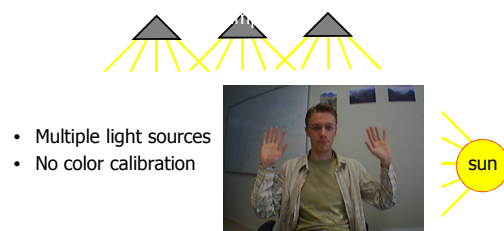
TU Delft

Results



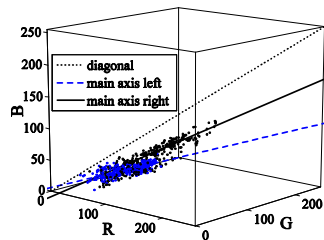
TU Delft

Results



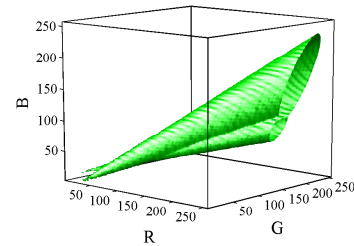
TU Delft

Results



Sampled face pixels and main axes in RGB space

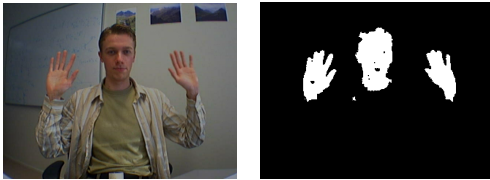
Results



Twin Cylindrical skin model, compensated by prior

Results

Skin Segmentation using ECF and hysteresis thresholding



Measured Properties of Hands

- Properties of left- and right hand during their movements:
 - Position
 - Velocity
 - Direction
 - Acceleration
 - Size change



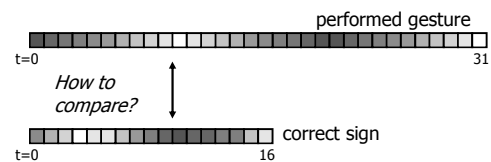
Classification



- Each sign is gestured by 10 different persons
- A correct gesture must be similar enough to that of 1 of these 10 persons
- A score around '1' indicates how similar a gesture is to a sign (the higher the better)

Classification

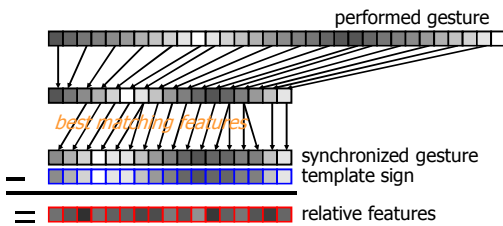
- Matching time sequences of features:



- Matching not trivial due to different gesture speed
- Warping & synchronization necessary
- Classifier for each gesture

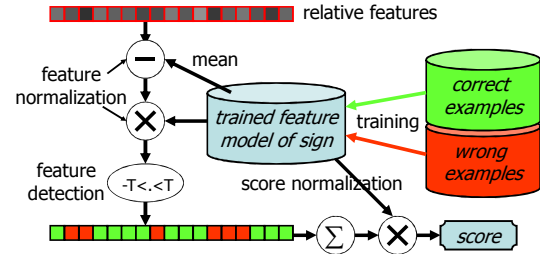
Classification

3) Feature difference with template: relative features

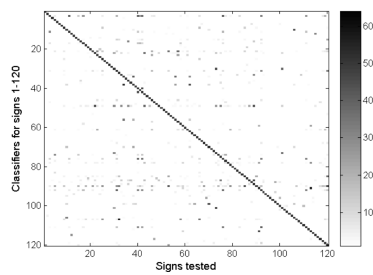


Classification

4) Classification of relative features



Results (confusion matrix)



Concluding Remarks

- Very Challenging Topic
- Powerful in restricted environments
- Strong in objective measurements
- Weak in subjective measurements
- Supervision/human interaction usually required