3D Face Recognition

Biometric Systems (DTU 02238)

Christoph Busch

Session 13

Overview 3D Face Recognition

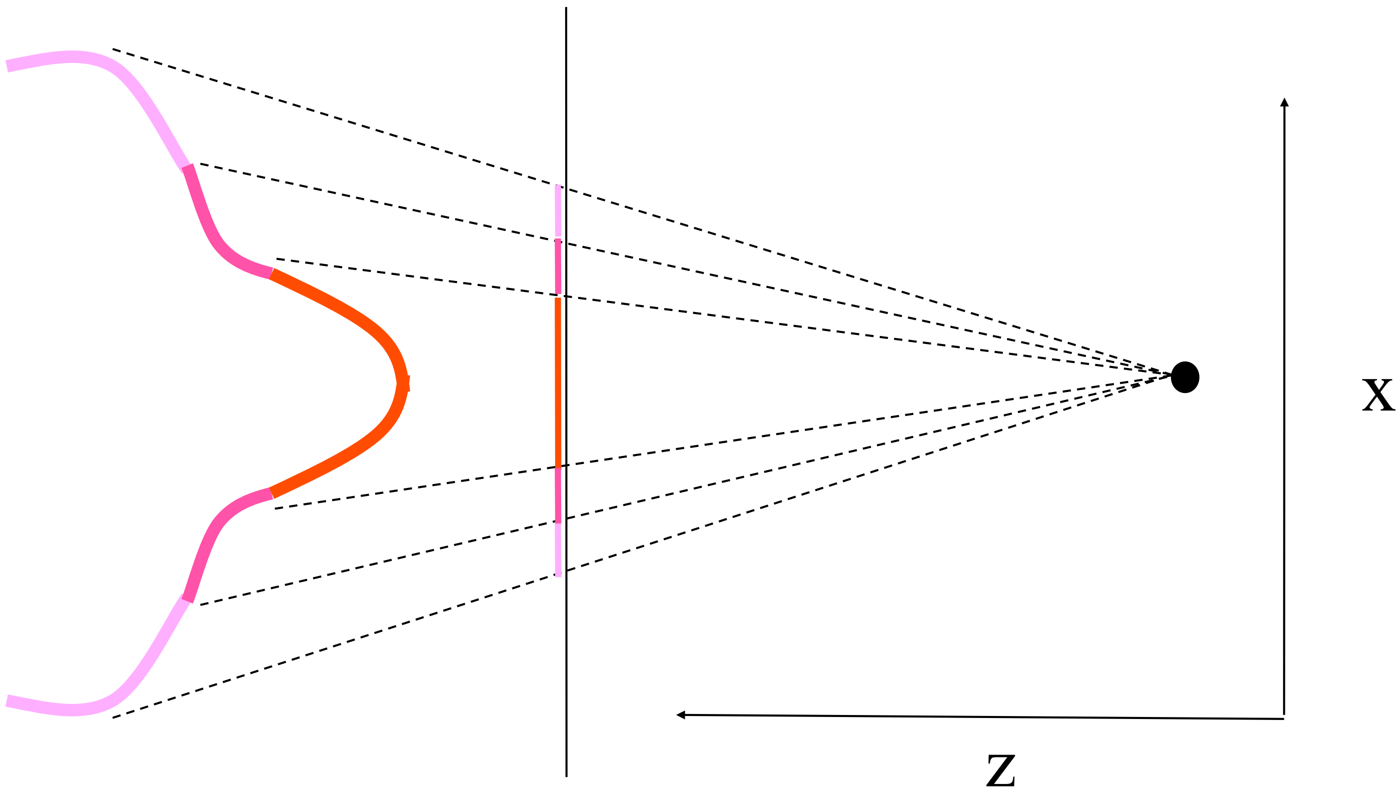
Structure of this session

- Unattended access control
- Capturing the 3D surface
- Scan normalisation - pose robustness
- Feature extraction
- Hybrid system (2D/3D) and improved biometric performance
- Usability - challenges at image acquisition

Unattended access control

2D Face Recognition

2D face recognition is recognition of a facial image

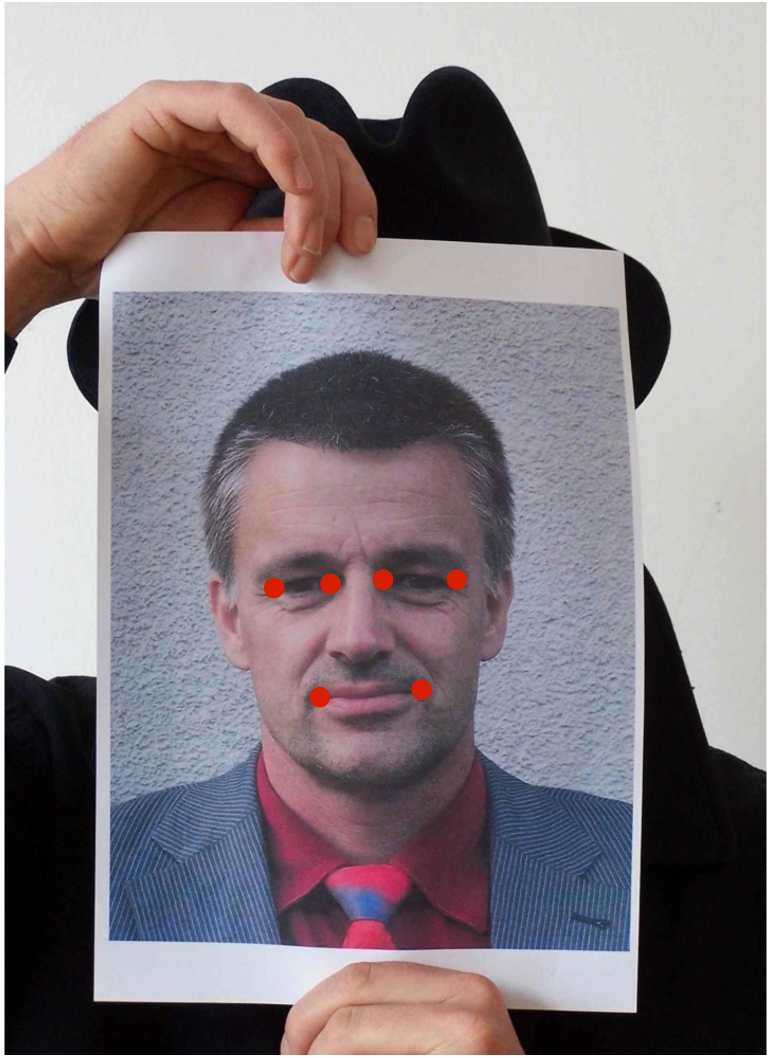


Face
(observed biometric characteristic)

Image plane

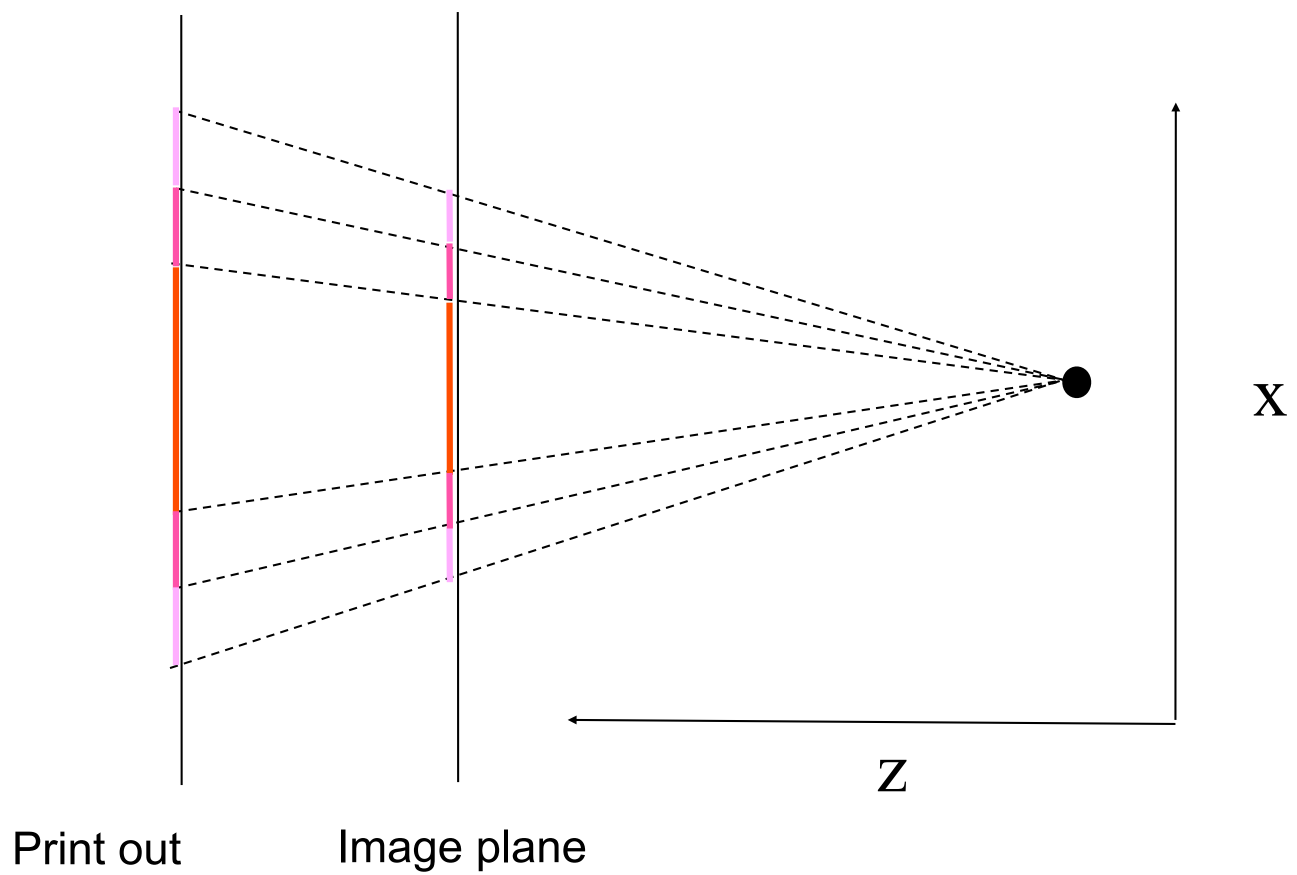
2D Face Recognition

2D face recognition is recognition of a facial image



2D Face Recognition

2D face recognition is recognition of a facial image



2D Face Recognition

PIE problems

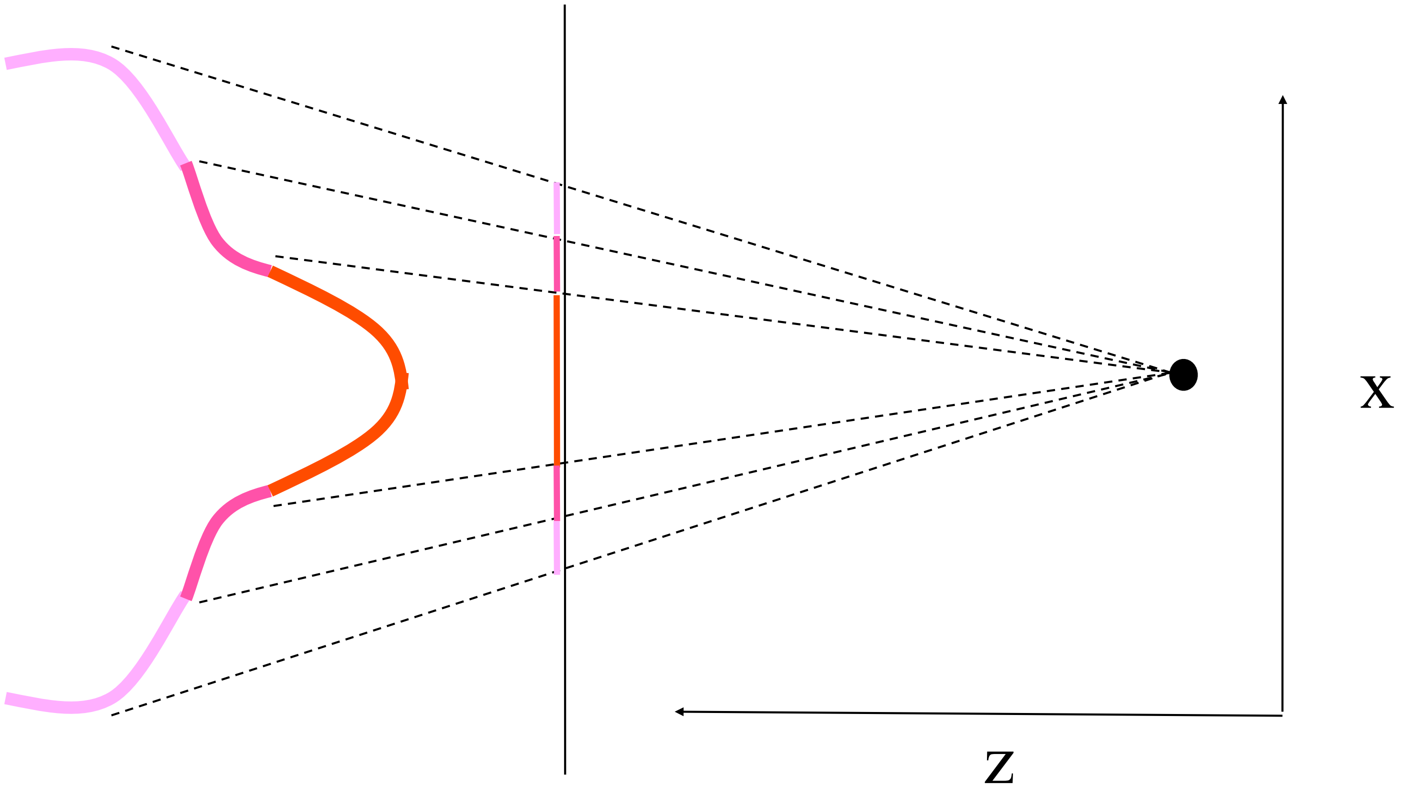
- Variations of pose


- Variations of illumination


- Variations of facial expression (mimic)

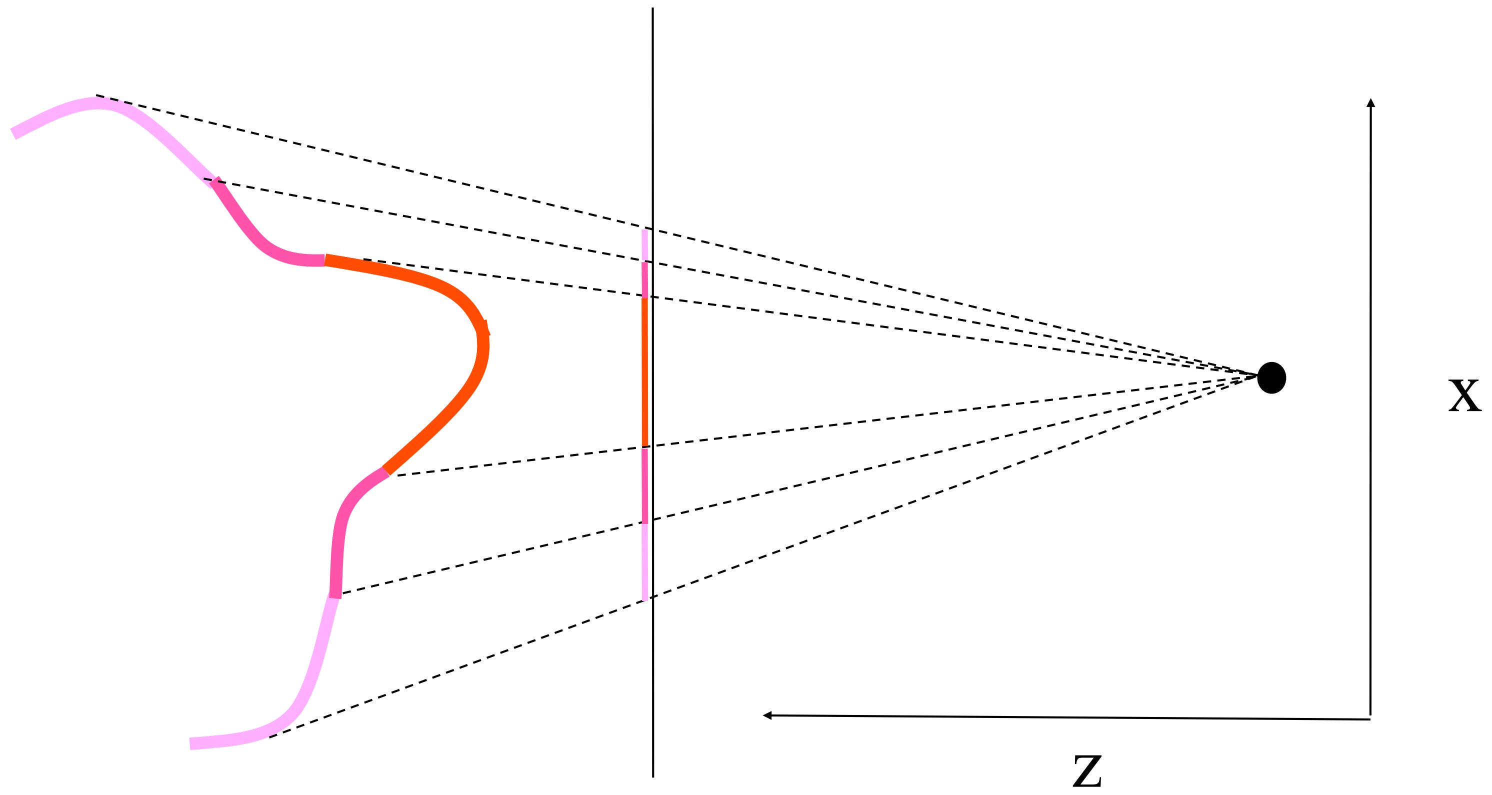

2D Face Recognition

Pose



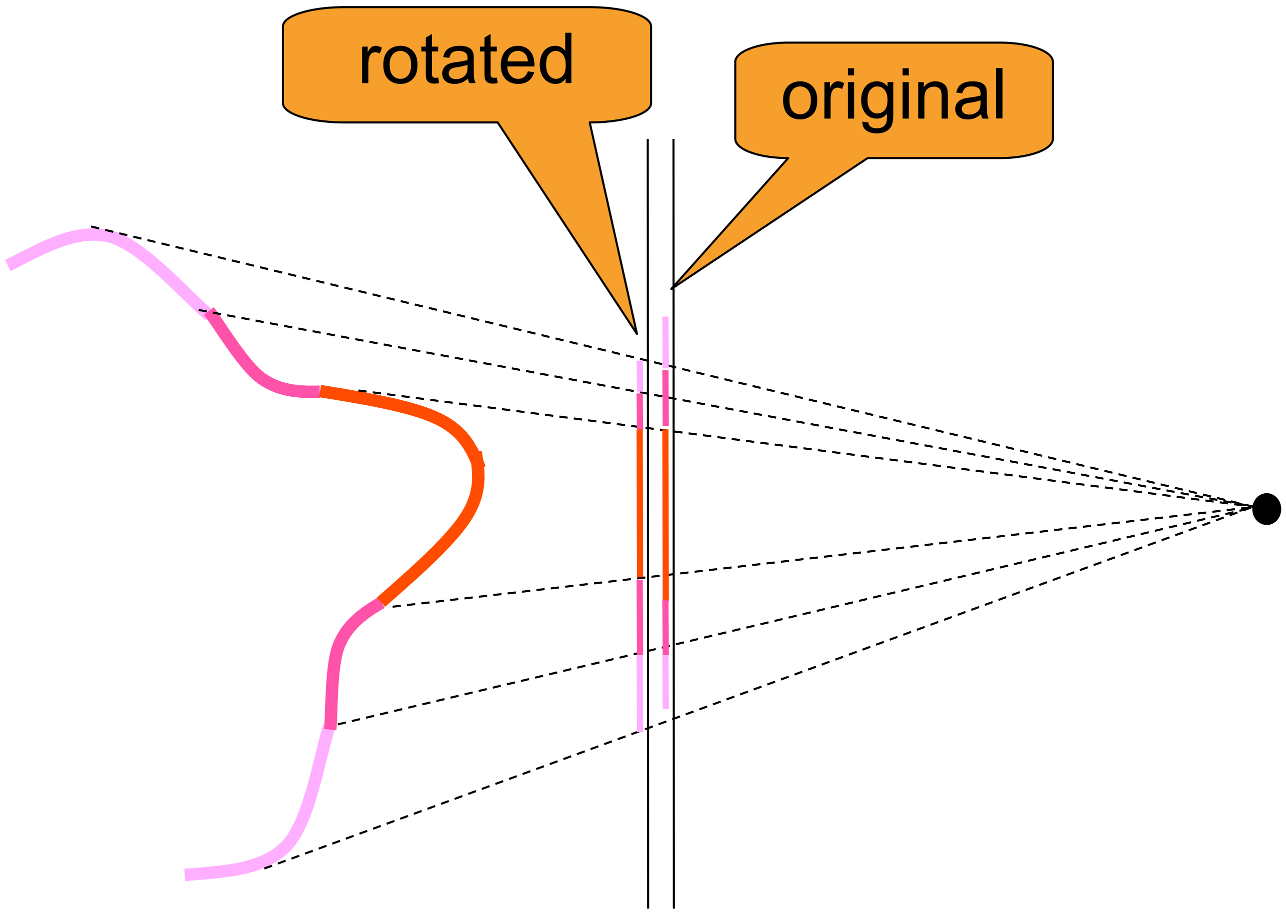
2D Face Recognition

Pose rotated



2D Face Recognition

Pose mismatch



Impact of Pose Variations

- Pose variations
  impact the biometric performance

(Y,P,R) = (0, 0, 0) (+45, 0, 0) (-45, 0, 0) (0, -45, 0) (0, +45, 0) (0, 0, -45) (0, 0, 45)

Image Source: ISO/IEC 39794-5, 2019

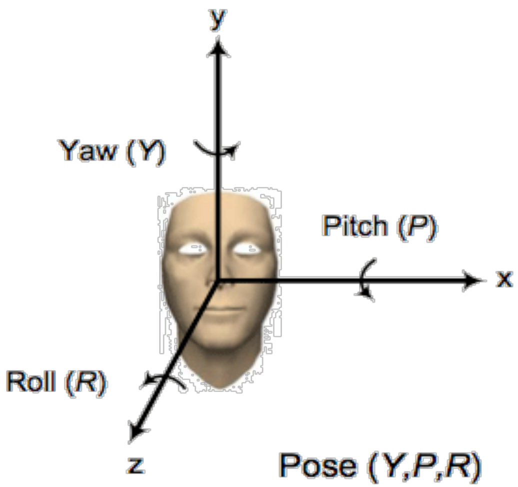
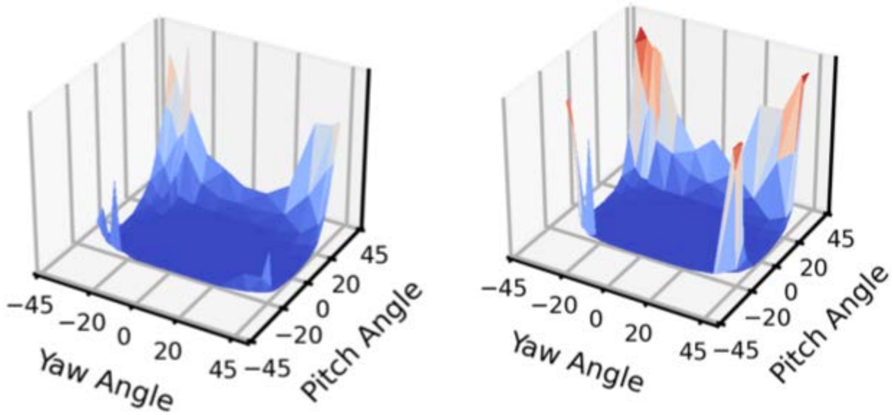
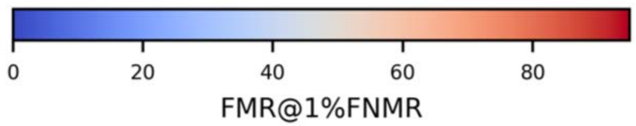


Image Source: ISO/IEC 39794-5, 2019



MagFace COTS-2022



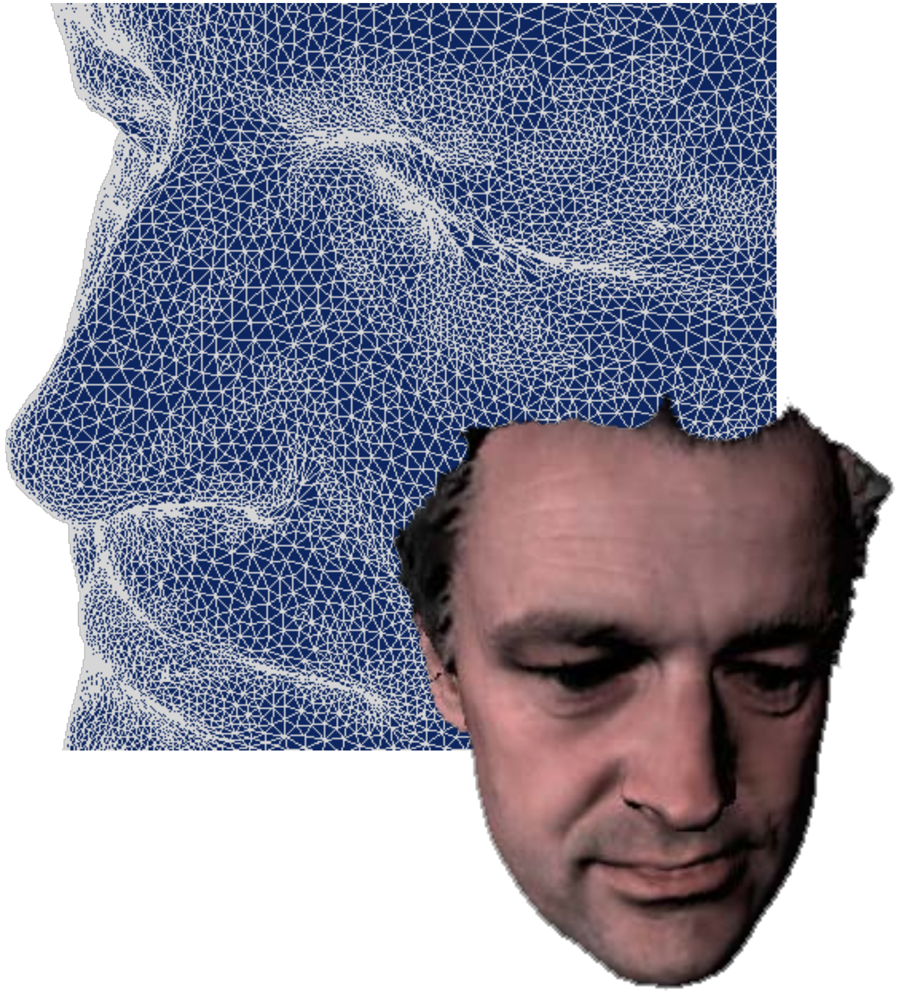
[Grimm2023] M. Grimmer, C. Rathgeb, C. Busch: "Pose Impact Estimation on Face Recognition using 3D-Aware Synthetic Data with Application to Quality Assessment“, https://arxiv.org/abs/2303.00491, (2023)

Capturing the 3D surface

3D Face Capture Device

Solution strategy

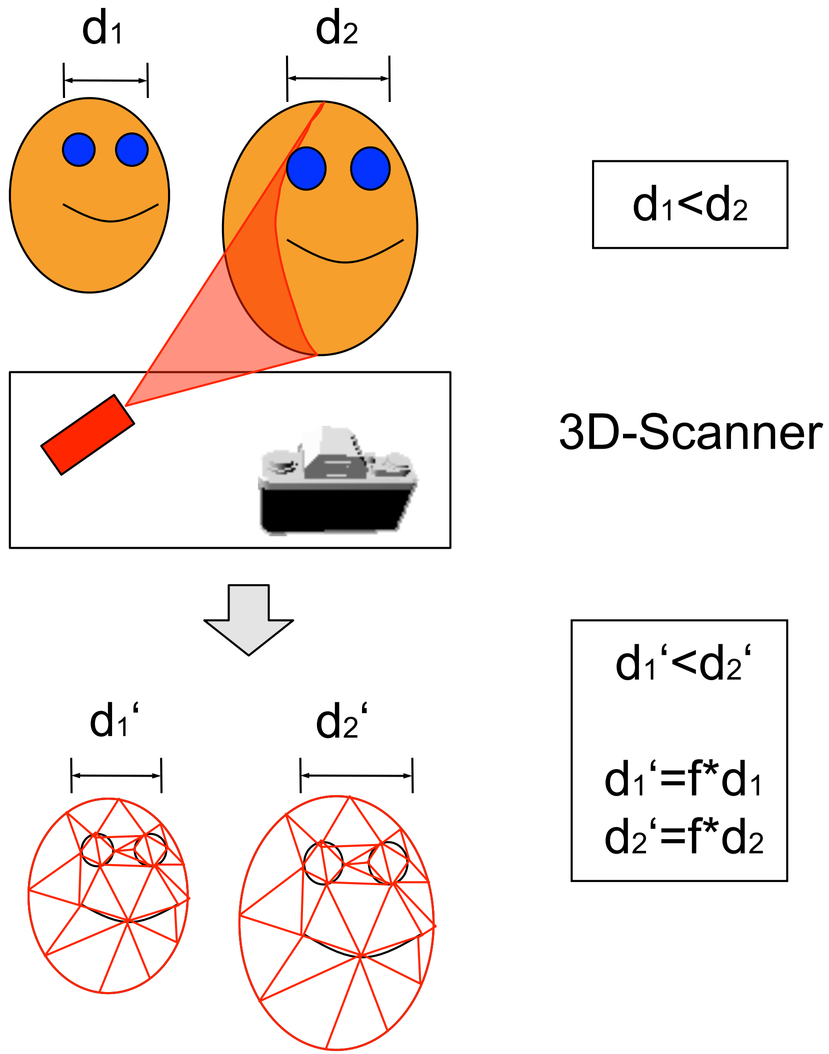
- Observation of texture and surface geometry
- Multi channel analysis



3D Face Capture Device

3D face scanning

- Calibrated scale
  - known distance
  - anatomical measure
- Pose
  - invertible rotation and translation
- Illumination
  - defined and constant illumination



3D Face Capture Device

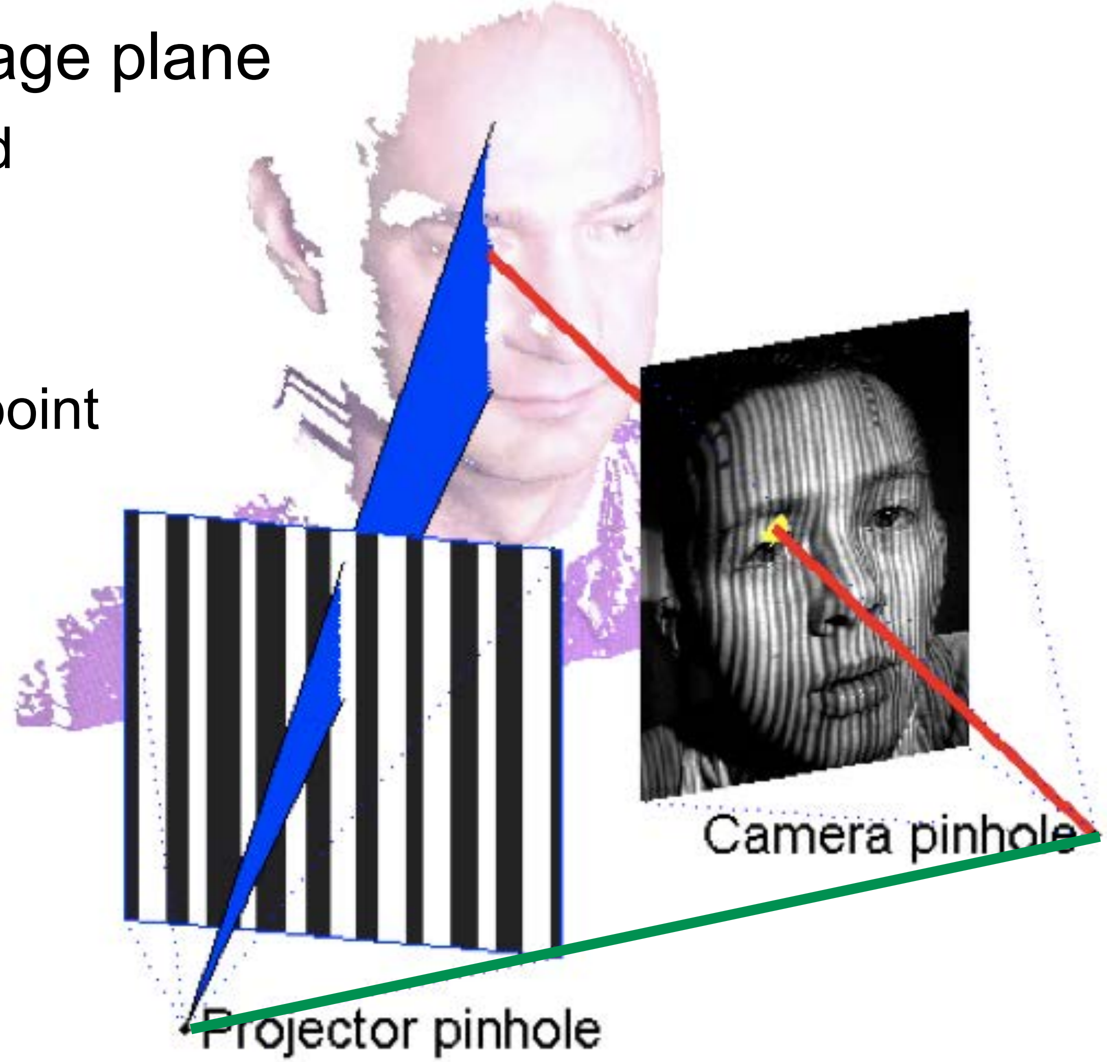
Capture technologies

- Active structured light
- Passive stereo vision
- Time of flight

3D Face Capture Device

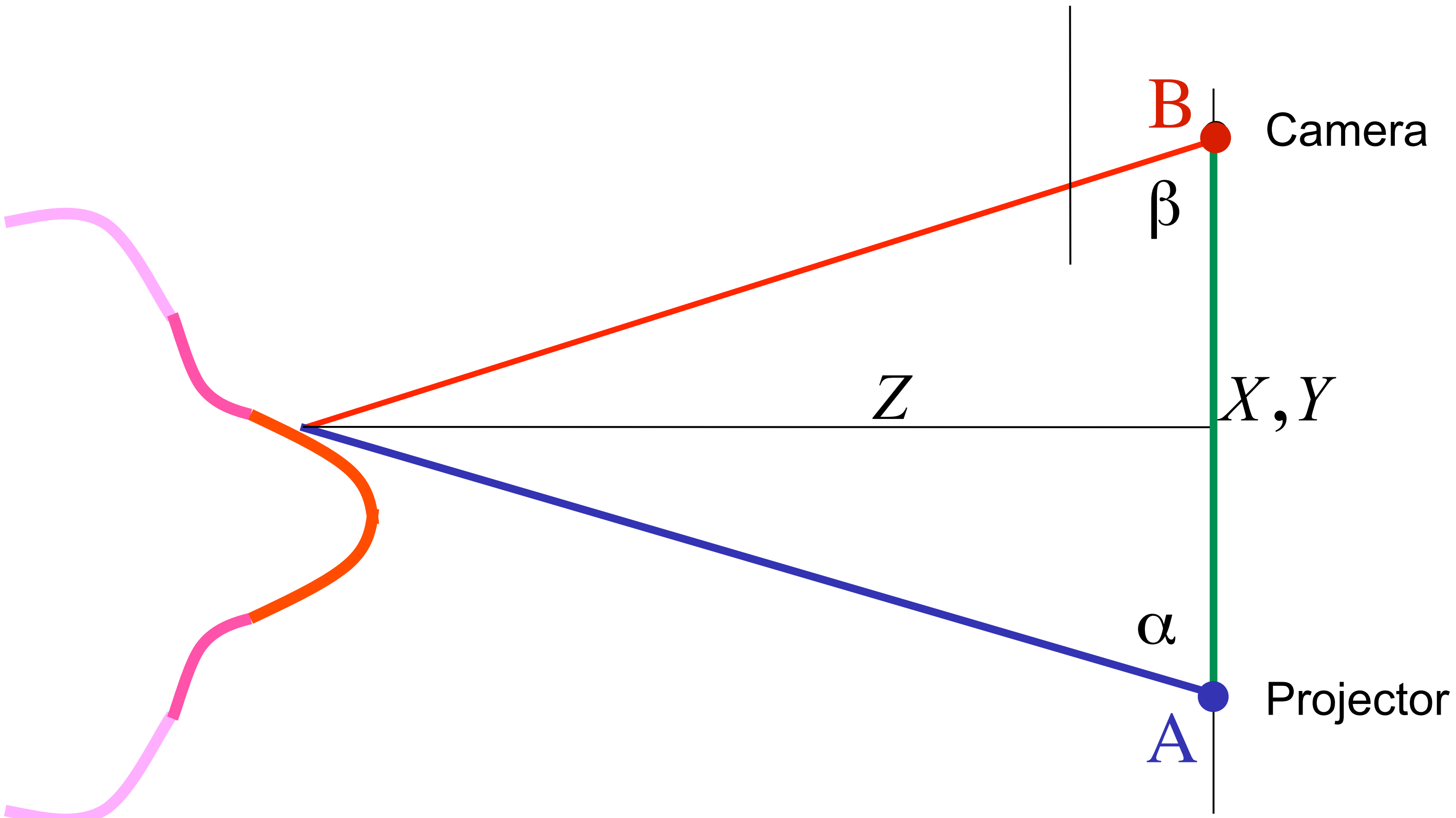
Structured light

- The projector spans a three dimensional plane i,
- For each point in the image plane
  - where a stripe is detected
  - a line is constructed from
    - the camera pinhole
    - through the detected point in the image plane
    - towards the unknown surface point



Structured light

Principle



$$(\alpha, \beta, \gamma) \rightarrow (X, Y, Z)$$

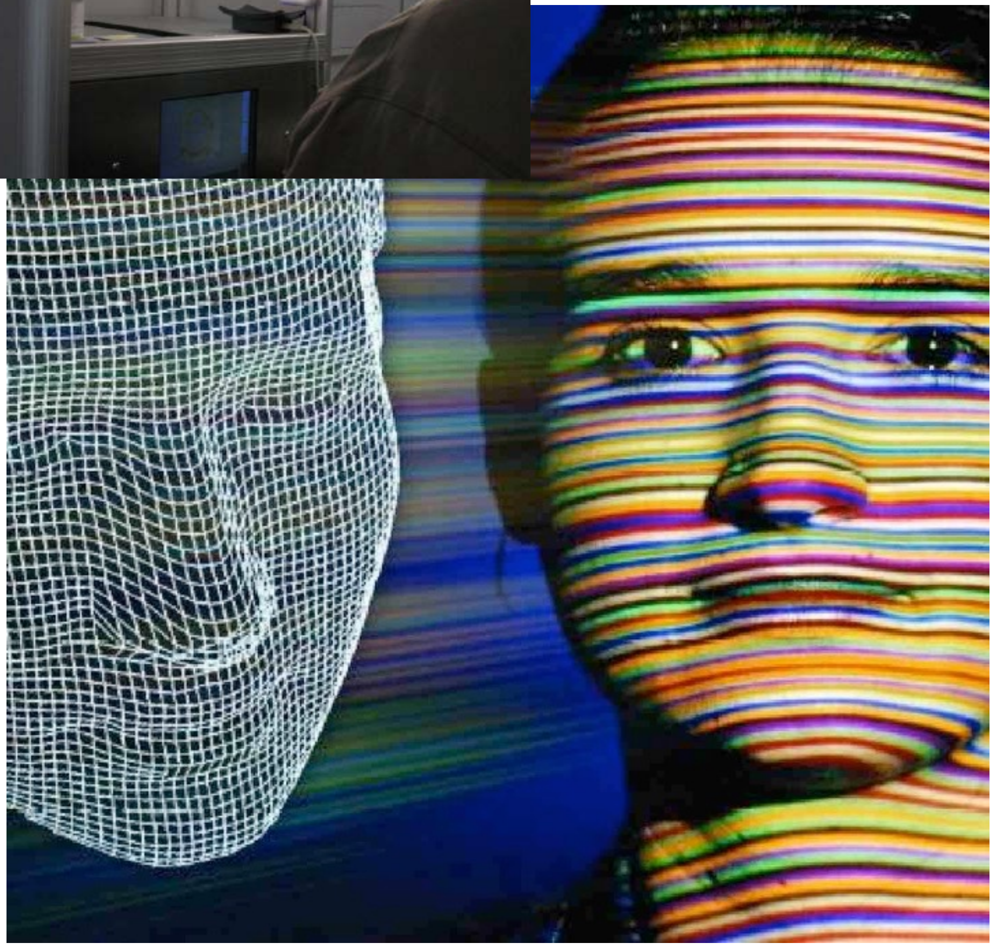
Structured Light

Stripe patterns

- Polygon
- Idemia
- Siemens

Laser scan

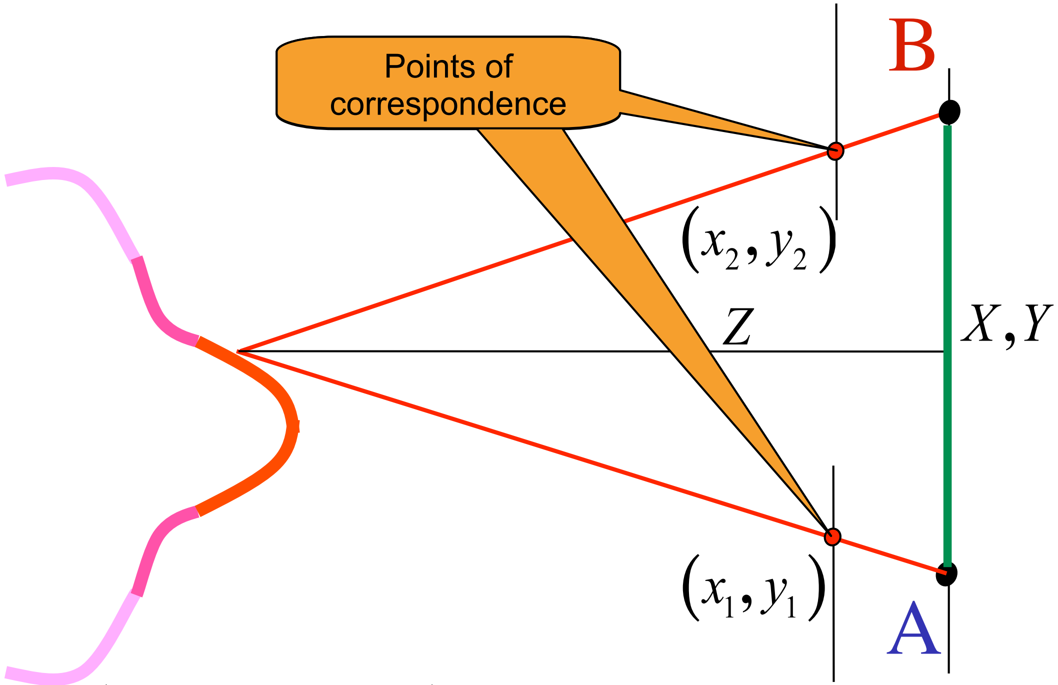
- Minolta Vivid
- Cyberware



3D Face Capture Device

Passive Stereo Vision

- Principle: stereoscopic vision



$$((x_1, y_1), (x_2, y_2)) \rightarrow (X, Y, Z)$$

Few salient points at landmarks

$$Z = \frac{baseline * focal\ length}{disparity}$$

$$disparity = ((x_1, y_1) - (x_2, y_2))$$

Passive Stereo Vision - Realsense

Intel Realsense capture device

- Stereo image sensing technology
  - active IR stereo with resolution 1280 x 720
  - field of View (FOV) is 63.4° x 40.4° or 85.2° x 58°
  - 90 Frames per Second (FPS)
  - small size (108x25x13 mm) and 55g
  - operates at 1.5 W

Image Source: https://realsense.intel.com/

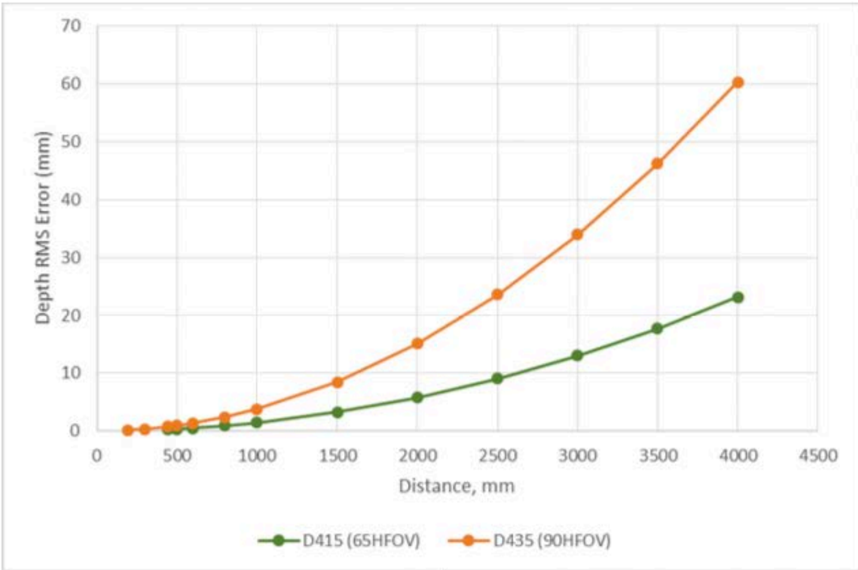


Image Source: https://dev.intelrealsense.com/docs

3D Face Capture Device

Time of flight

- Pulsed method

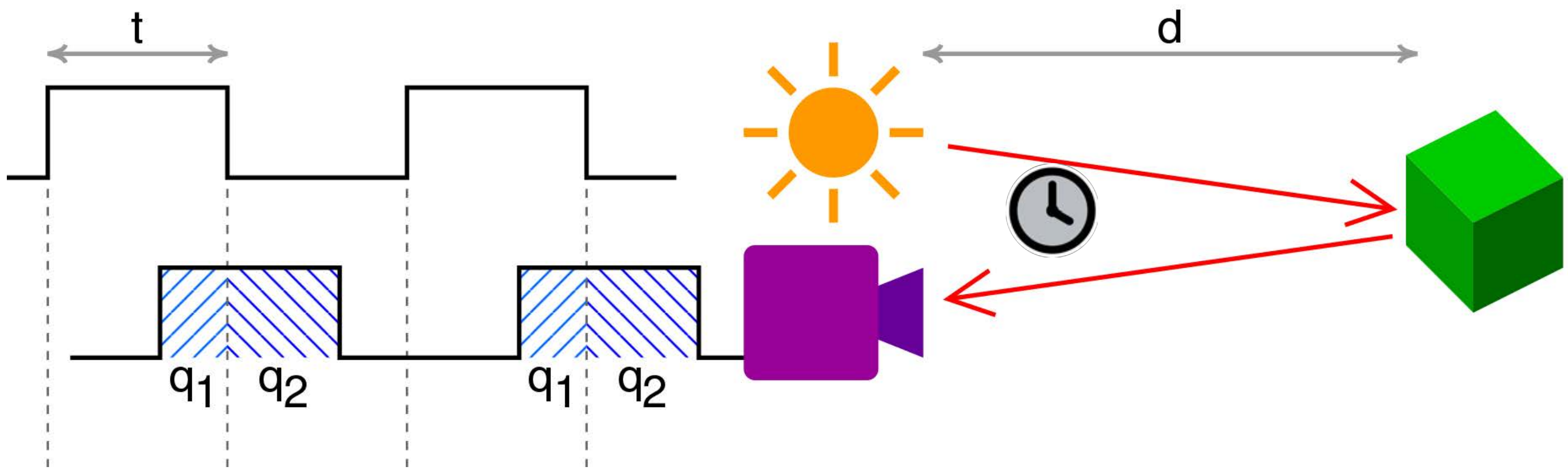


Image Source: https://en.wikipedia.org/

- c is speed of light
- t_D is the time delay
- t is the length of the pulse
- $q1$ is the accumulated charge in the pixel, when light is emitted
- $q2$ is the accumulated charge when it is not

$$t_D = \frac{2 * d}{c}$$

$$d = \frac{c * t}{2} \frac{q2}{q1 + q2}$$

Time of Flight - Kinect

Windows capture device

- RGB camera that stores data in a 1280x960 resolution
- An infrared (IR) emitter and an IR depth sensor
  - The reflected beams are converted into depth information measuring the distance between an object and the sensor
  - 30 frames per second (FPS)

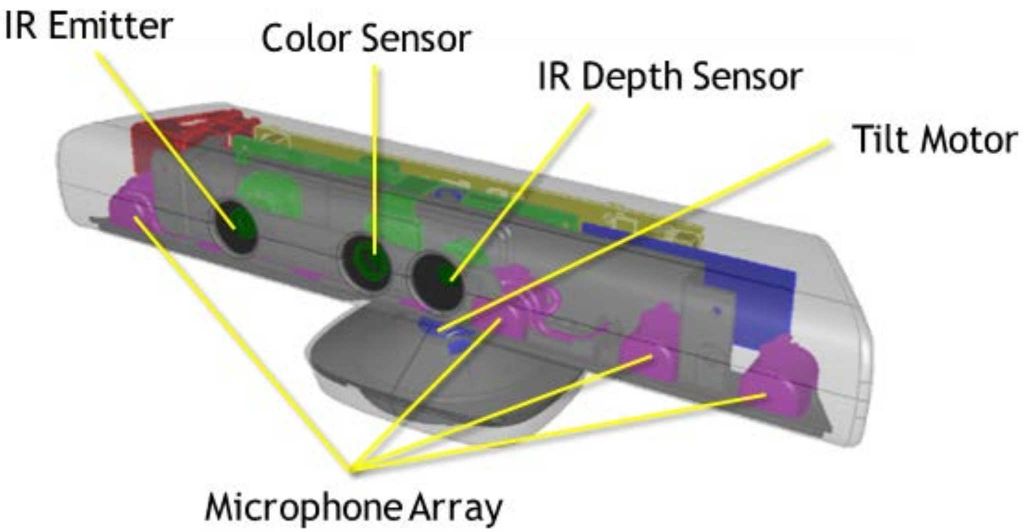


Image Source: https://msdn.microsoft.com

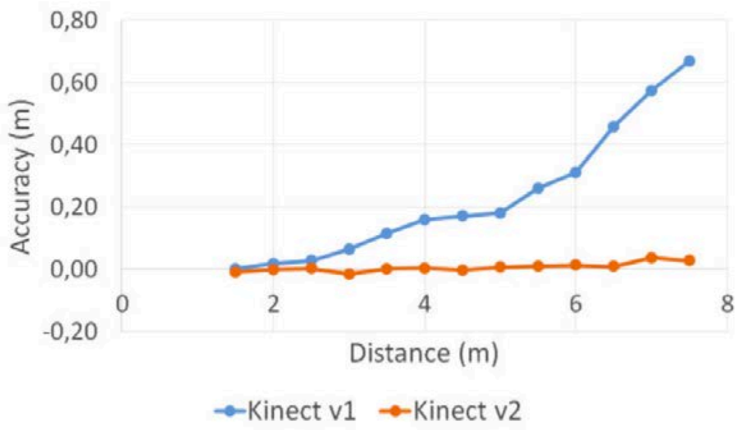


Image Source: Simone Zennaro: Performance evaluation of the 1st and 2nd generation Kinect

Time of Flight - FaceID

iPhone capture device

- Vertical-cavity surface-emitting laser (VCSEL)
  - laser beam emission
- An emitter projects more than 30.000 infrared dots
  - the pattern is projected from the laser using an Active Diffractive Optical
- Comparison in the Secure Enclave

Image Source: https://en.wikipedia.org/wiki/Face_ID

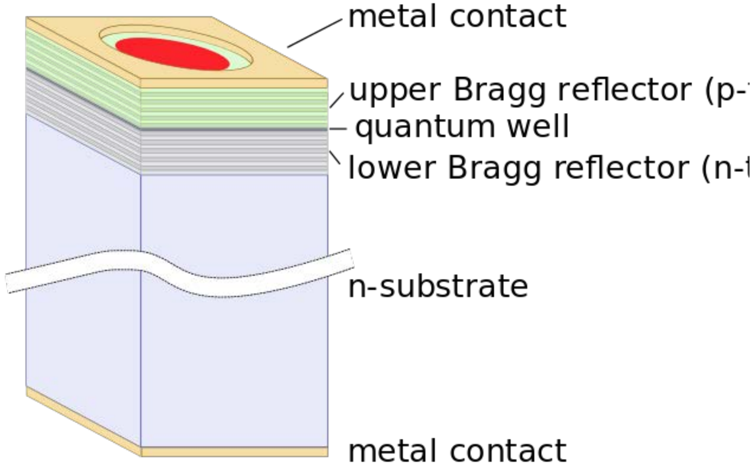


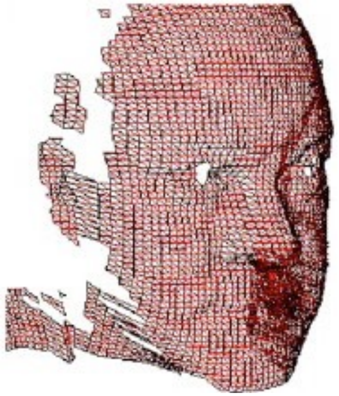
Image Source: https://en.wikipedia.org/wiki/Vertical-cavity_surface-emitting_laser

[Apple2020] https://www.biometricupdate.com/202008/apple-patent-details-new-tof-depth-camera-system-for-improved-face-biometrics

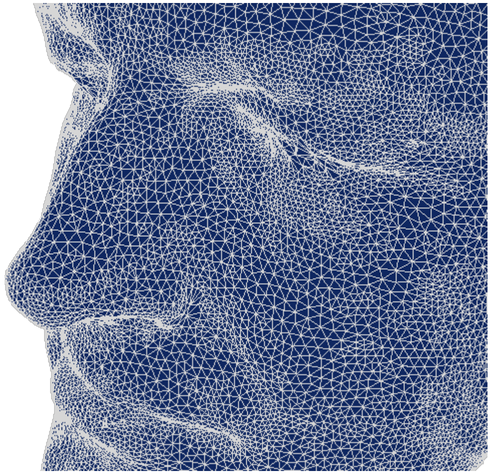
Data Structures and Representations

Point clouds

- local structures

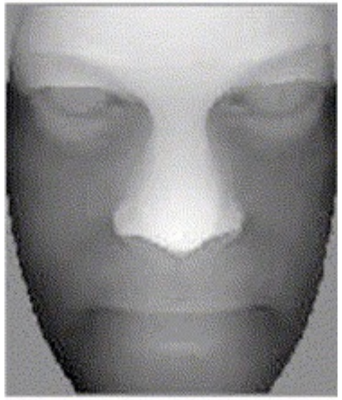


3D Mesh



Range images

- global structures
- 2D approaches



Scan normalisation - pose robustness

Pre-processing: Registration

Registration

- Rotating the scan into a frontal pose
- What is frontal?
  - the nose should be positioned towards the capture device

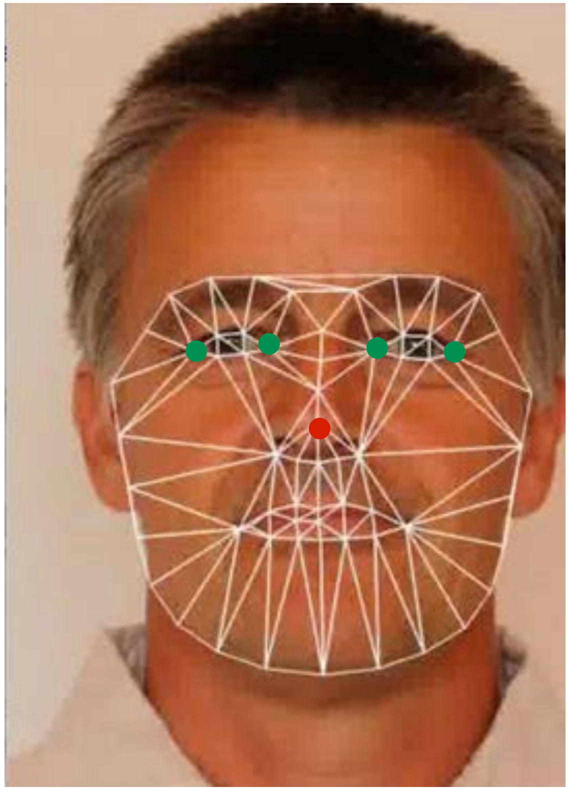
Only rotation and translation (no scaling, unlike in 2D)

Common approaches

- Extended Gaussian Image (EGI)
- Point clouds - Iterative Closest Points (ICP)
- Landmark based registration

Pre-processing: Landmark Detection

2D landmarks can be projected to the 3D surface



- Some landmarks can be detected more reliably in 2D (e.g. the corners of the eye)

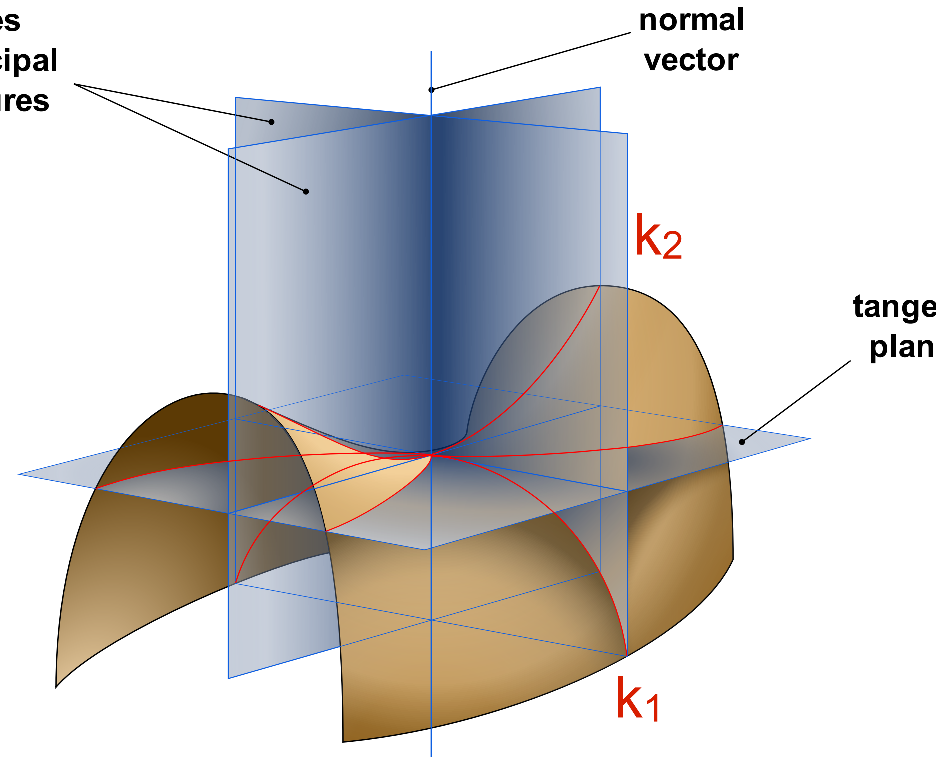
Pre-processing: Landmark Detection

3D landmark detection

- Landmark detection relies on the curvature analysis of the facial surface

Curvature

- Principal Curvatures
  - are the eigenvalues of the shape operator at the point
  - k_1 and k_2



- Mean Curvature

$$H = \frac{1}{2}(k_1 + k_2)$$

- Gaussian Curvature

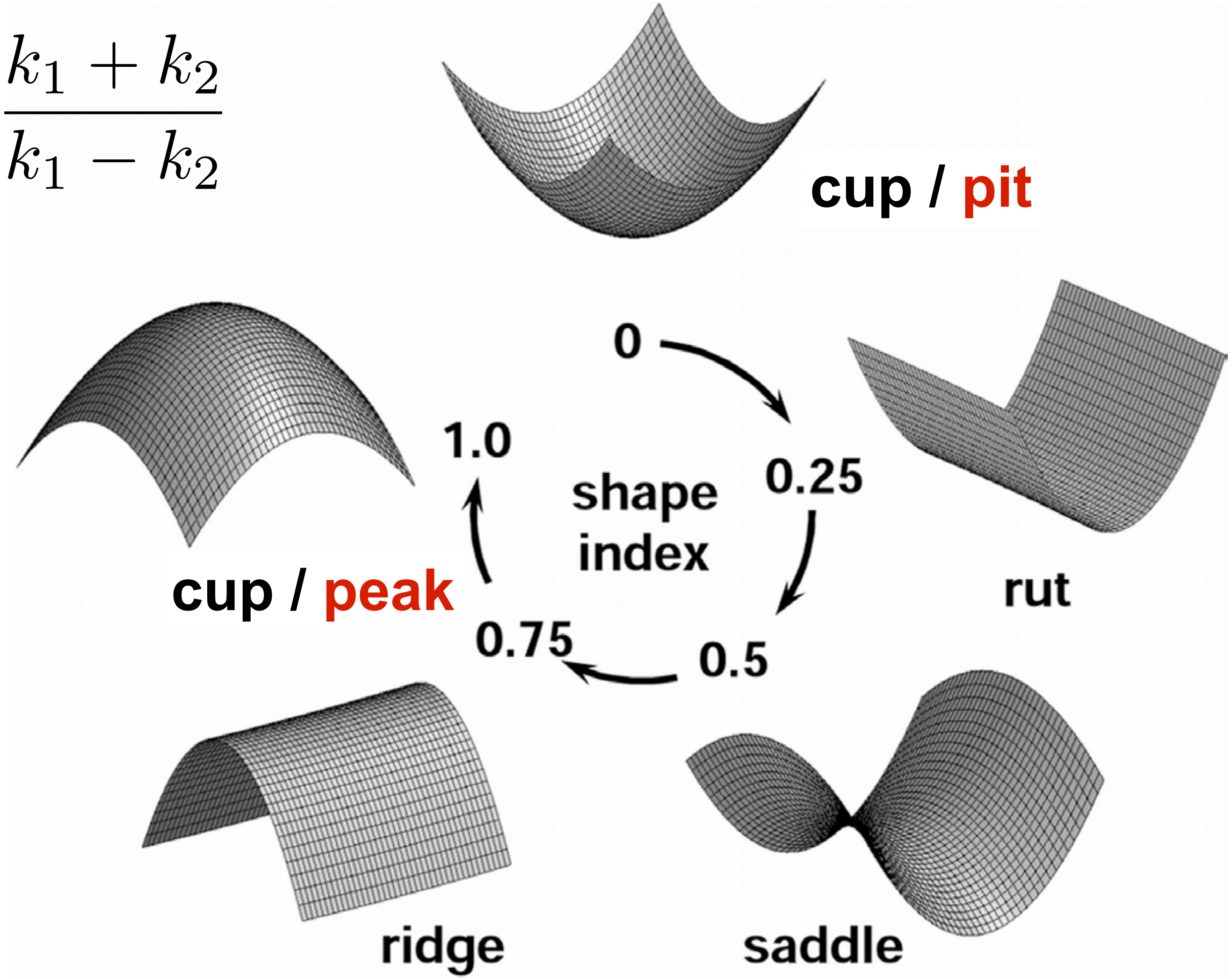
$$K = k_1 * k_2$$

Pre-processing: Landmark Detection

Shape Index

- different shapes map to different shape index values

$$S = \frac{1}{2} - \frac{1}{\pi} tan^{-1} \frac{k_1 + k_2}{k_1 - k_2}$$



Pre-processing: Landmark Detection

Categorizing the surface patch [Seg2007]

- Pits and peaks and other types of the shape can be found by analysis of the Gaussian and the Mean curvature

K / H	< 0	= 0	> 0
< 0	saddle ridge	minimal	saddle valley
= 0	ridge	flat	valley
> 0	peak	(none)	pit


Range image

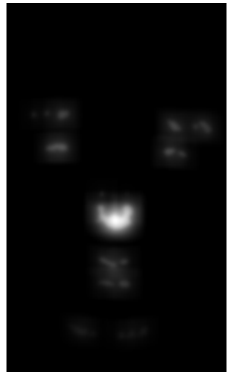

Principal curvature k_1


Principal curvature k_2


Mean curvature


Gaussian curvature


Shape index

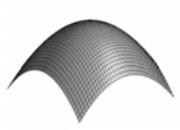

Peak density

[Seg2007] M. P. Segundo et al.: „Automatic 3d facial segmentation and landmark detection“. Proceedings of the 14th International Conference on Image Analysis and Processin (ICIAP), (2007)

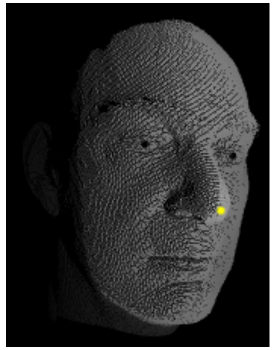
Pre-processing: Landmark Detection

Landmark Detection

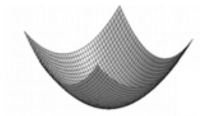
- Stephan Mracek (Masterthesis in 2010)
  - classify each surface point as peak, pit, saddle or ridge
- Create peak density map



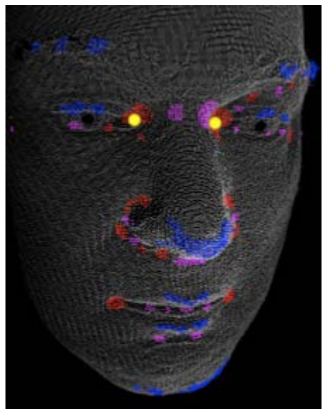
  - tip of the nose is located
    as point with highest peak density



- Create pit density map



  - the inner eye corners are located
    as points with the highest pit density

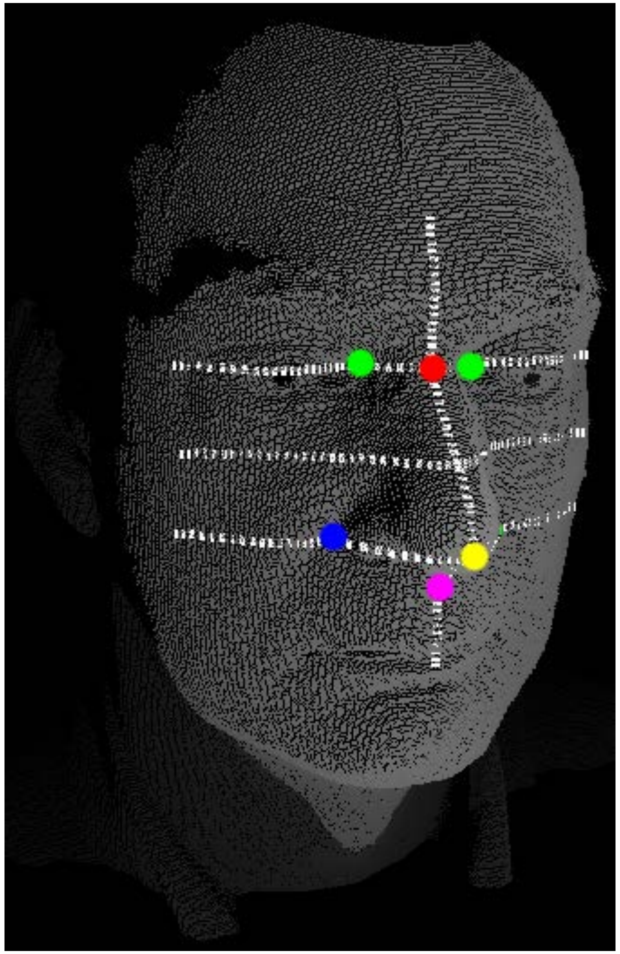


[Marc2011] S. Mracek, C. Busch, R. Dvorak, M. Drahansky: "Inspired by Bertillon – Recognition Based on Anatomical Features from 3D Face Scans", in Proceedings International Workshop on Security and Communication Networks (IWSCN), May 18-20, (2011)

Pre-processing: Landmark Detection

Full set of landmarks

- nose tip
- nasal bridge
- 2 inner eye corners
- 2 outer nose corners
- lower nose corner
- vertical profile curve
- 3 horizontal profile curves



99,6 % detection rate

Feature extraction

Feature Extraction Methods

Approaches

- Extended eigenface (Achermann 1997)
- Suface normals and intensities (Tsutsumi 1998)
- Hausdorff distance (Achermann and Bunke 2000)
- Vertical Profiles (Beunier and Acheroy 2000)
- Point signatures - on landmark points (Chua 2000)
- 3D morphable model (Blanz and Vetter 2003)
- 2D and 3D eigen decomposition of flattened textures (Bronstein 2003)
- Histogram-based depth features (Zhou 2008)
- Anatomical Bertillon features (Mracek 2010)

Forensic Anthropometry

Bertillonage -
included eleven measurements:

1. Height
2. Stretch: Length of body from left shoulder to right middle finger when arms raised
3. Brust: length of torso from head to seat, taken when seated
4. Length of head: Crown to forehead
5. Width of head: Temple to temple
6. Length or right ear
7. Length of left foot
8. Length of left middle finger
9. Length of left cubit: Elbow to tip of middle finger
10. Width of cheeks
11. Length of left little finger

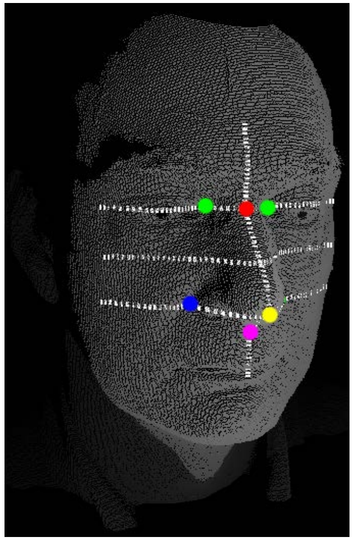
RELEVÉ
DU
SIGNALEMENT ANTHROPOMÉTRIQUE

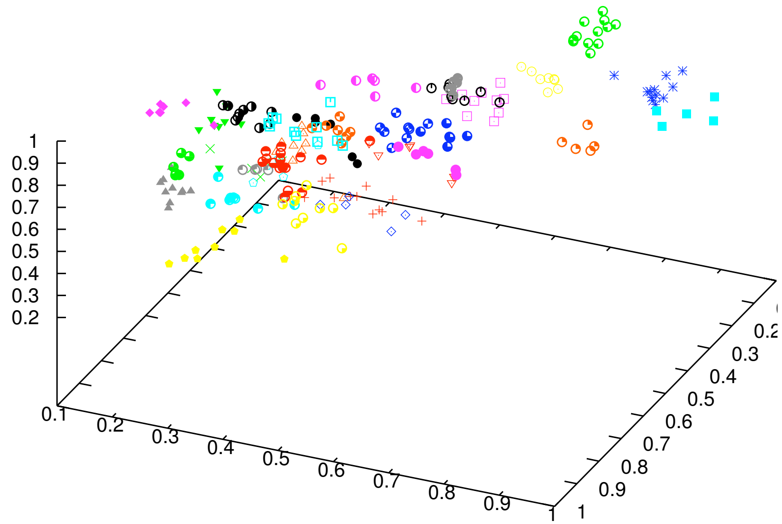
1. Taille. — 2. Envergure. — 3. Buste. —
4. Longueur de la tête. — 5. Largeur de la tête. — 6. Oreille droite. —
7. Pied gauche. — 8. Médius gauche. — 9. Coudée gauche.

3D Feature Extraction

Bertillonage

- Still in use today - but in a different context!





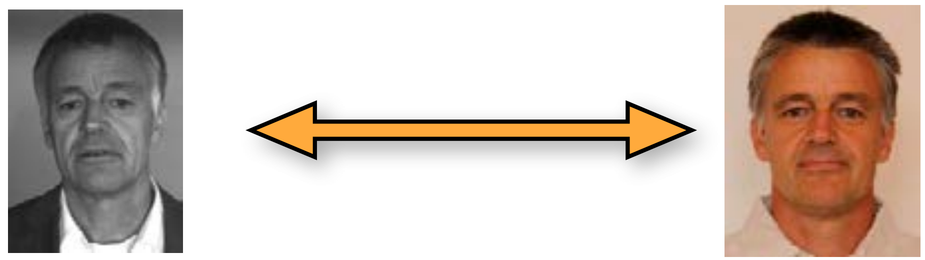
- Anatomical features
  - distance between nasal bridge and lower nose corner
  - distance between eye corners
  - correlation of profile curve to mean face
  - curvature at specific points

Hybrid system (2D/3D)

Benefit with existing 2D ePassports

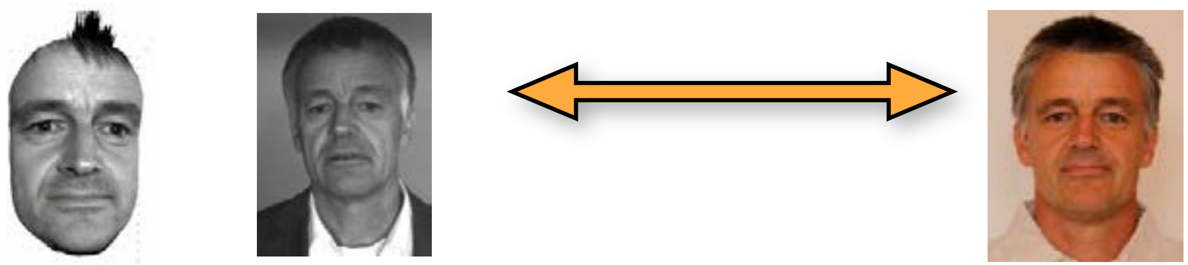
The following three combinations are benchmarked:

- **Today's passport setup:** plain 2D comparison



- **3D Face project setup**: Both (2D + 3D) stored as reference

- **Hybrid setup:** 3D probe compared with 2D reference



Benefit with existing 2D ePassports

Hybrid setup:

Test the efficacy of 3D when applied only to the probe sample

- Capture types
  - probe sample: 2D texture image + 3D shape data

  - reference sample: 2D texture image

- Sample characteristics
  - probe sample: taken from challenging data contains pose and variable expression
  - reference sample is a frontal image with neutral expression, as used for passport images (but with higher resolution)

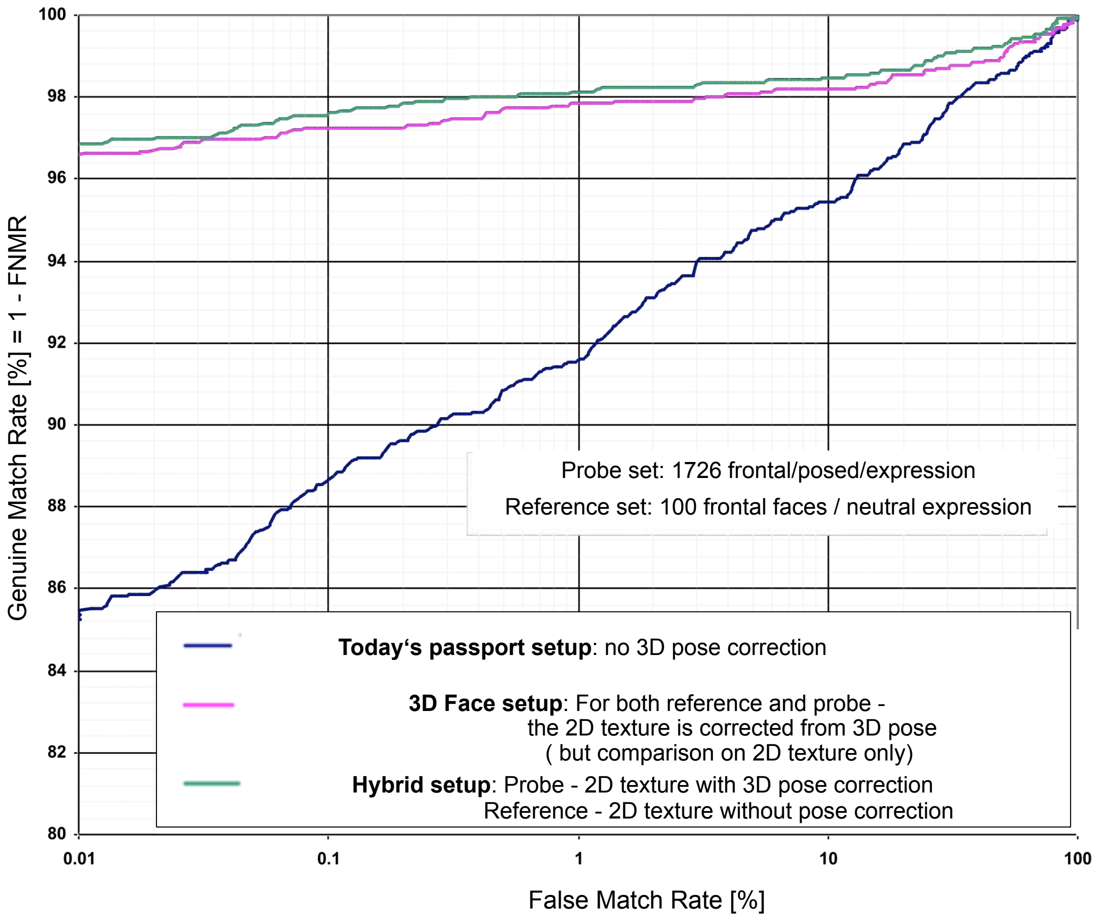
- Method: correct the probe image employing the 3D data for image normalization

Benefit with existing 2D ePassports

Conclusion:

- **pose correction** improves texture recognition considerably when benchmarked against plain **2D recognition**
- 3D for pose correction is useful whenever there is something to correct.

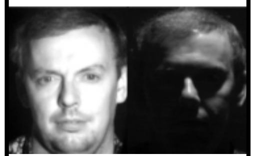
ROC



Conclusion

Conclusion

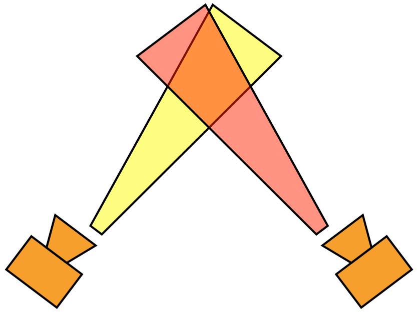
Benchmarking the challenges in 2D vs 3D

Challenge		3D Face recognition problem	2D Face recognition problem
Head rotation		solved	unsolved
unkown **Distance** from the capture device		solved	unsolved
Lighting conditions		depends	unsolved
Facial expression		unsolved	unsolved

Conclusion

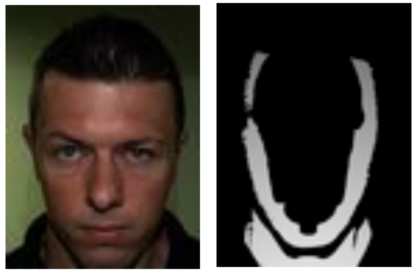
3D face recognition is a promising approach

- Can solve pose and illumination problem
- Achieves better biometric performance than 2D FR
  - 2D vs. high res 2D
- Stronger robustness against attacks
  - PAD for unsupervised Automated Access Control



Structural disadvantages

- Complexity
- Limited scan volume - constrained distance of data subject to capture device



References

Web

- 3D Capture devices
  http://www.simple3d.com
- 3D Face Project
  https://christoph-busch.de/3dface/

Complementary reading

- M. Boersma: „Biometric Recognition based on 3D Face Geometry“, M.Sc. Thesis, 2005
- Bowyer et al.: „A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition”, Computer Vision and Image Understanding, 2006
- S. Mracek, C. Busch, R. Dvorak, M. Drahansky: „Inspired by Bertillon – Recognition Based on Anatomical Features from 3D Face Scans", International Workshop on Security and Communication Networks (IWSCN), 2011
- C. Busch et al.: „Towards a more Secure Border Control with 3D Face Recognition", in Proceedings of the 5th Norsk Informasjons Sikkerhets Konferanse (NISK), 2012