Co-funded by the
Erasmus+ Programme
of the European Union

TrustID - Face Recognition Framework

Multiplier Event @ ISR Coimbra

Pedro Martins

Institute of Systems and Robotics (ISR)

University of Coimbra, Portugal

https://www.isr.uc.pt/~pedromartins

pedromartins@isr.uc.pt

22/05/2023

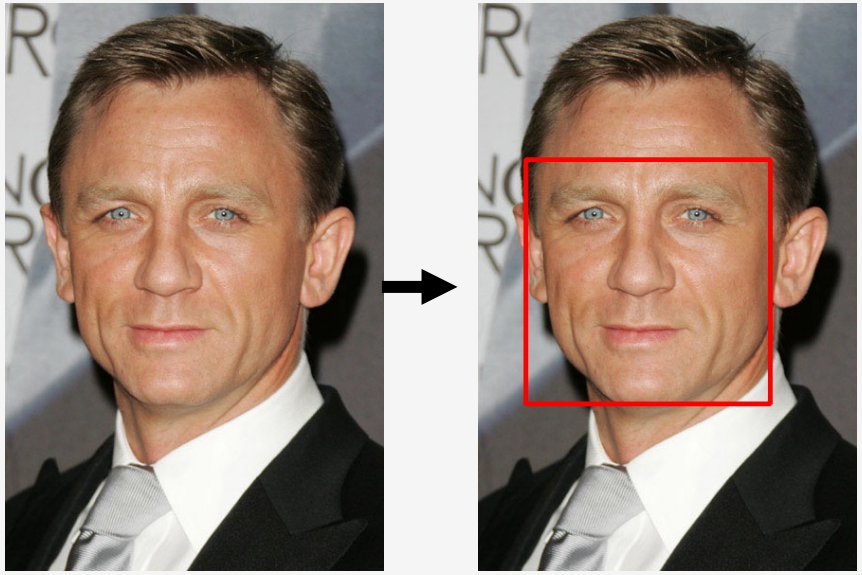
TrustID Face Based Authentication

Co-funded by the
Erasmus+ Programme
of the European Union

- Image/video based Biometric system.
- **Goal:** Identify users from pre-recorded image/video examples.
- System constrains:
  - Low quality data (consumer grade webcams, low resolution, noise, unconstrained environments, etc).
  - Computational performance limitations (needs run smoothly in 'unknown' target PCs).
  - Needs to be as accurate and reliable as possible.
- Overview of the Computer Vision and Machine Learning system architecture:
  - Face Detection.
  - TrustID - Face Recognition System v1.0.
  - TrustID - Face Recognition/Verification System v2.0.
  - Face Alignment (localization of facial features).
  - 3D Head Pose Estimation.
- Privacy and Computational Concerns.
- Face Recognition vs Face Verification.
- Face Verification Approach.
- Future Work.

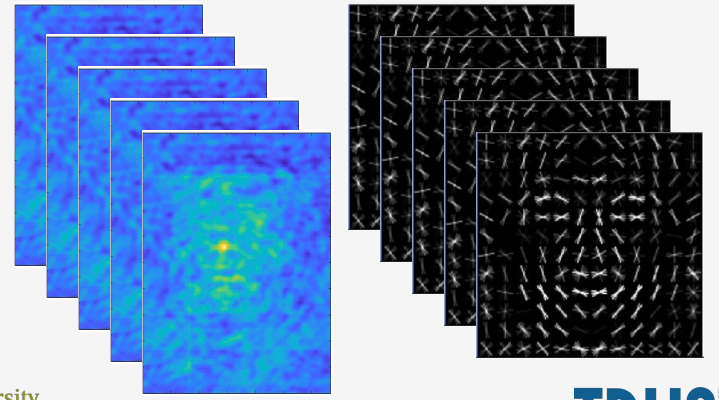
Face Detection

Co-funded by the
Erasmus+ Programme
of the European Union



Input image

Face detection

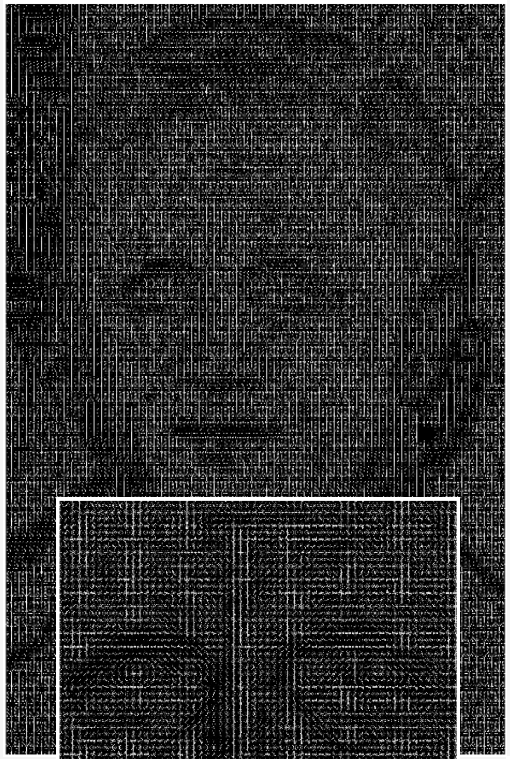


- DLIB HoG+SVM detector
  - Discriminative template matching.
  - Histogram of oriented Gradients (HoG) - features
  - Linear Support Vector Machine (SVM).
- Bank of 5 HOG filters.
  - Cell size 8x8.
- Trained in subset of LFW database.
- Advantages:
  - Lightweight model.
  - Fast operation on CPU.
  - Works well for frontal (and near non-frontal) faces
  - Allows small occlusions.
- Disadvantages:
  - Minimum face size of 80×80 px.
  - Unable to deal w/ extreme poses.

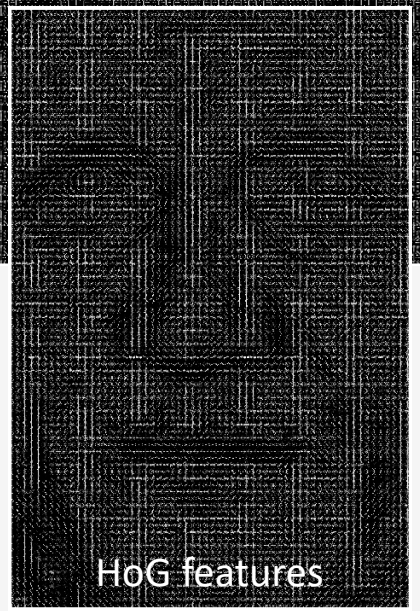
Face Detection - HoG Filters

Co-funded by the
Erasmus+ Programme
of the European Union

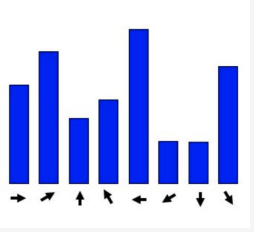
RGB Input image



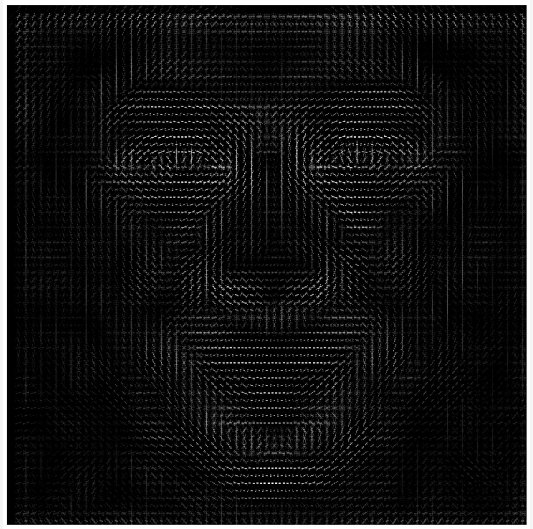
zoom



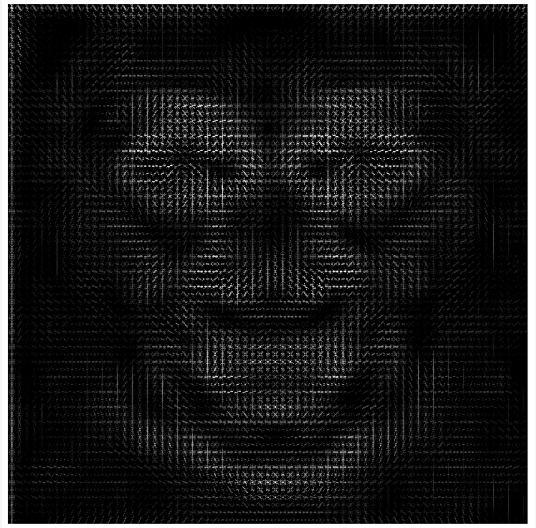
Distribution (histogram)
of gradient orientations



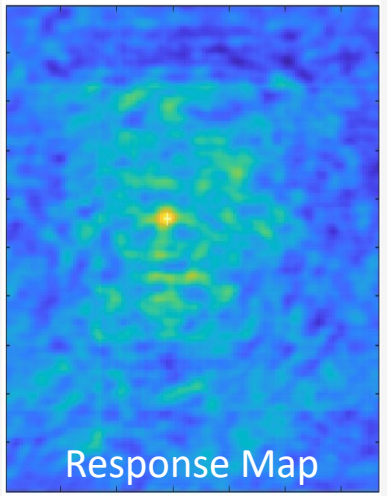
Histogram of Orientations



Filter: positive weights



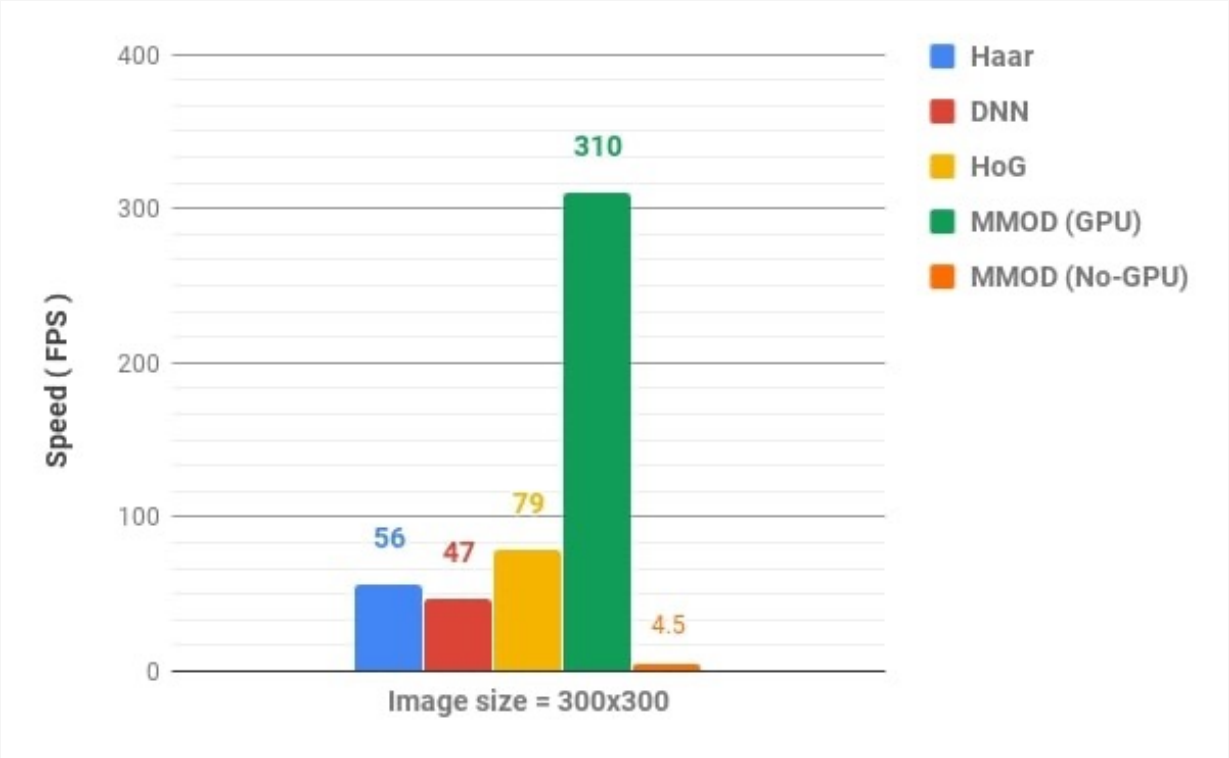
Filter: negative weights



Face Detection - Computational Performance

Co-funded by the
Erasmus+ Programme
of the European Union

- Computational Performance Evaluation of several methods:
  - Haar Features + AdaBoost (Viola Jones) - OpenCV
  - Single-Shot-Multibox detector - DNN (ResNET10) - OpenCV
  - HoG + SVM (DLIB)
  - Maximum-Margin Object Detector (MMOG) (with with CNN features) (GPU) (DLIB).
  - MMOG (CPU) (DLIB).
- Testing protocol:
  - 300x300 test image.
  - Each method run 10000 times.
  - Record average execution time.
- Hardware:
  - CPU: Intel Core i7 6850K (6 Core)
  - 32 GB RAM
  - NVIDIA GTX 1080 Ti (11 GB)
  - OS : Ubuntu 16.04 LTS
  - Programming Language : Python

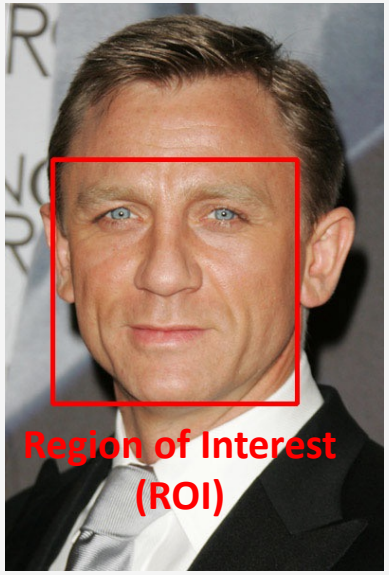


Courtesy of Vikas Gupta

TrustID - Face Recognition System v1.0

Co-funded by the
Erasmus+ Programme
of the European Union

Input image



Face detection

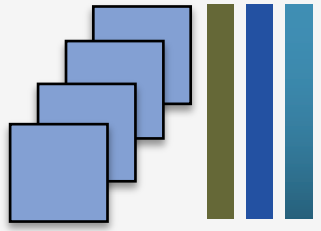
Discriminative template matching (HoG + linear SVM)

Image warp

Normalized Image

Canonical reference frame (224x224x3)

zero mean & unit std



Convolutional Neural Network (CNN)

ResNet-18 network architecture
Cross-Entropy loss function

"Daniel Craig"

Predicted identity

Training stage:

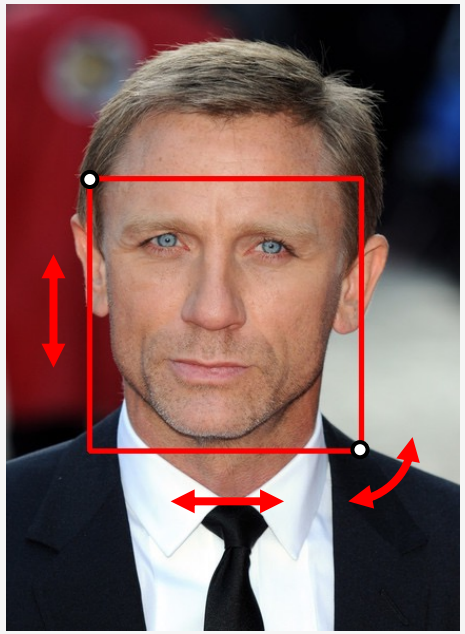
- Assemble user's image database.
- Face detection (locate faces in all images).
- Data augmentation (add "virtual" variation to database images).
- Learn multiclass classifier from corresponding image/users examples.

Testing stage:

- Face detection.
- Predict user identity, using the pretrained classifier model.

Image Augmentation (similarity transformations)

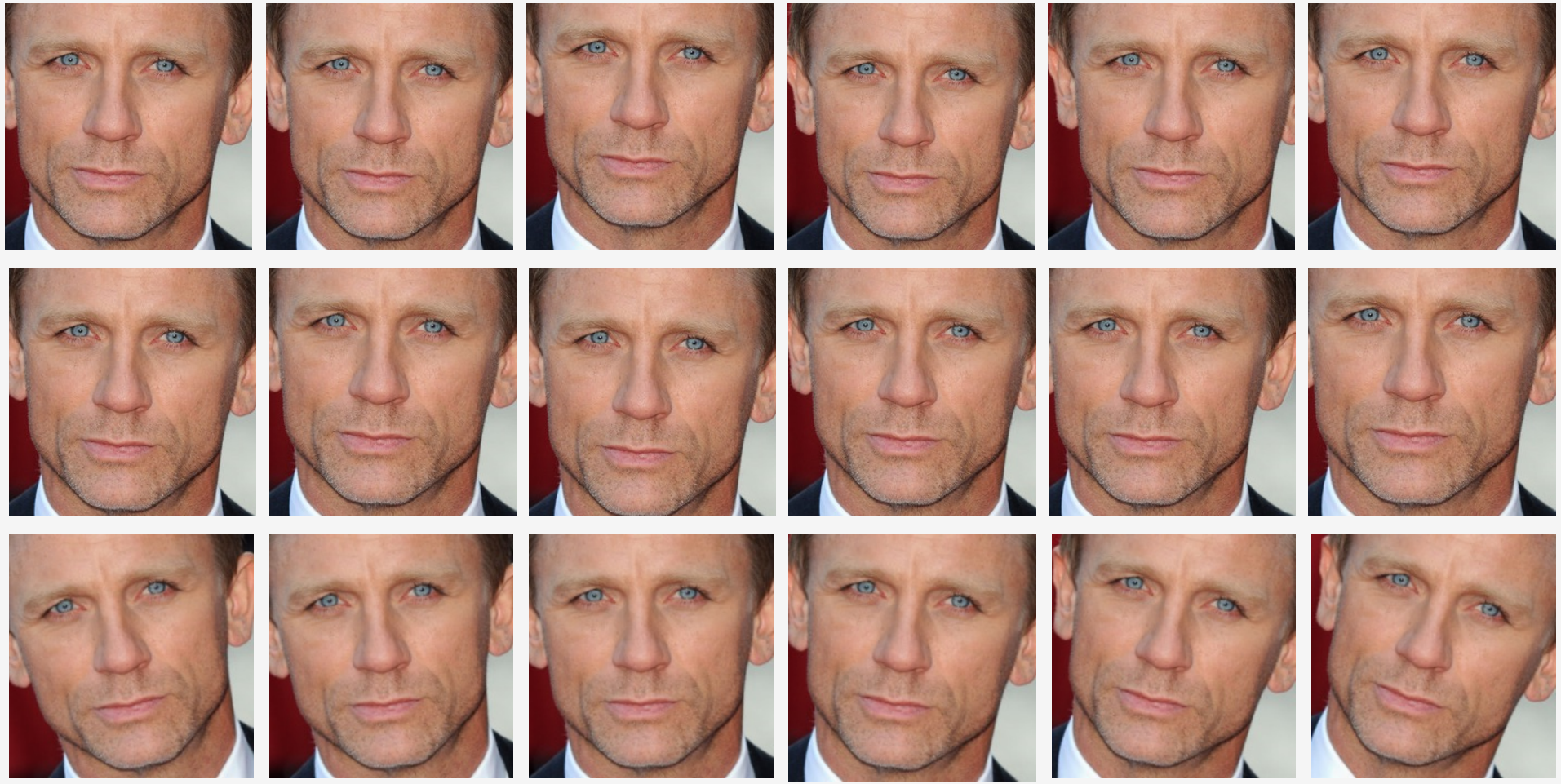
Co-funded by the
Erasmus+ Programme
of the European Union



Face Detection

x19

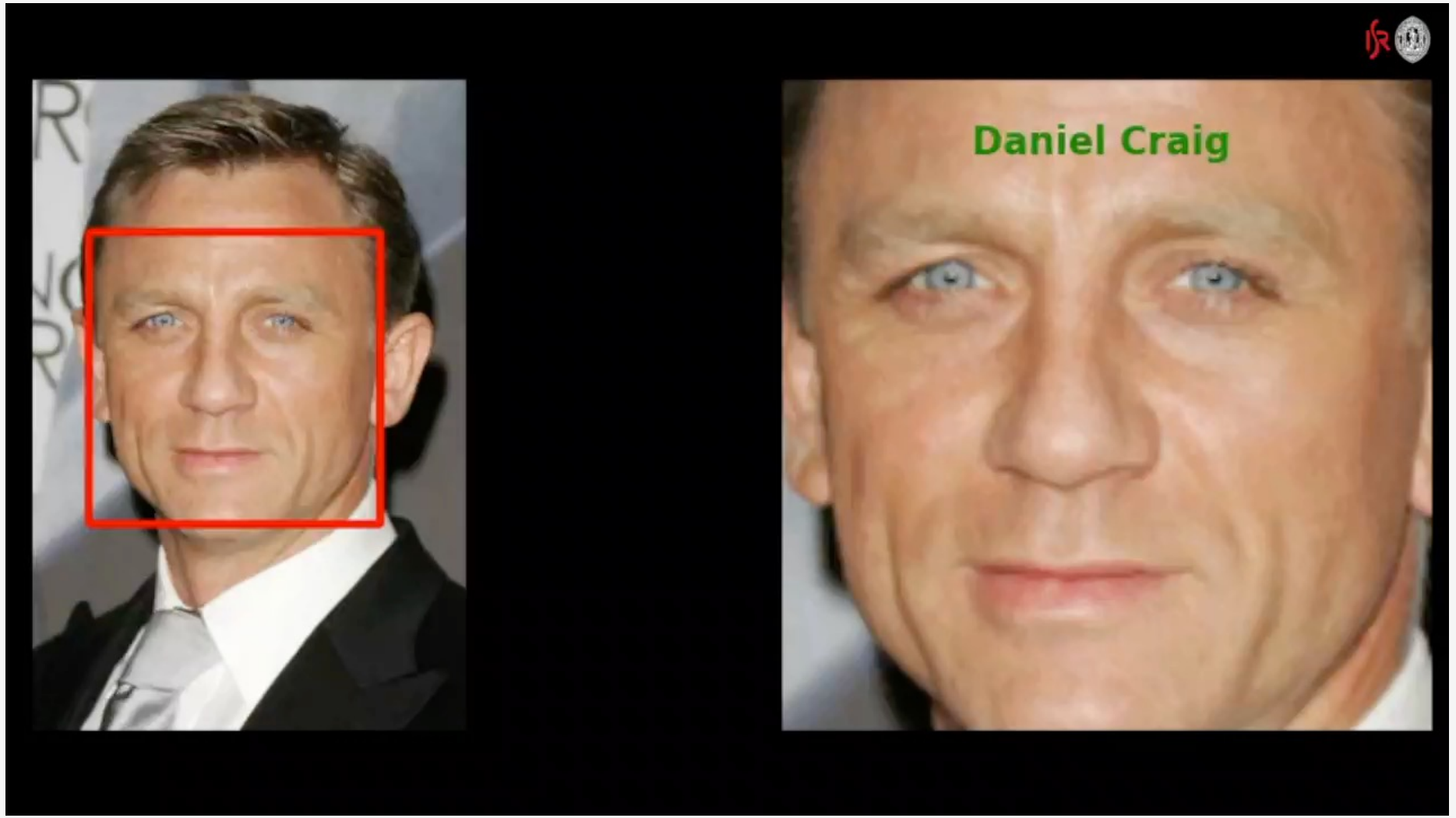
Region of Interest (ROI)



Augmented Images

Demo video (single image inference)

Co-funded by the
Erasmus+ Programme
of the European Union



Demo video (inference frame by frame)

Co-funded by the
Erasmus+ Programme
of the European Union

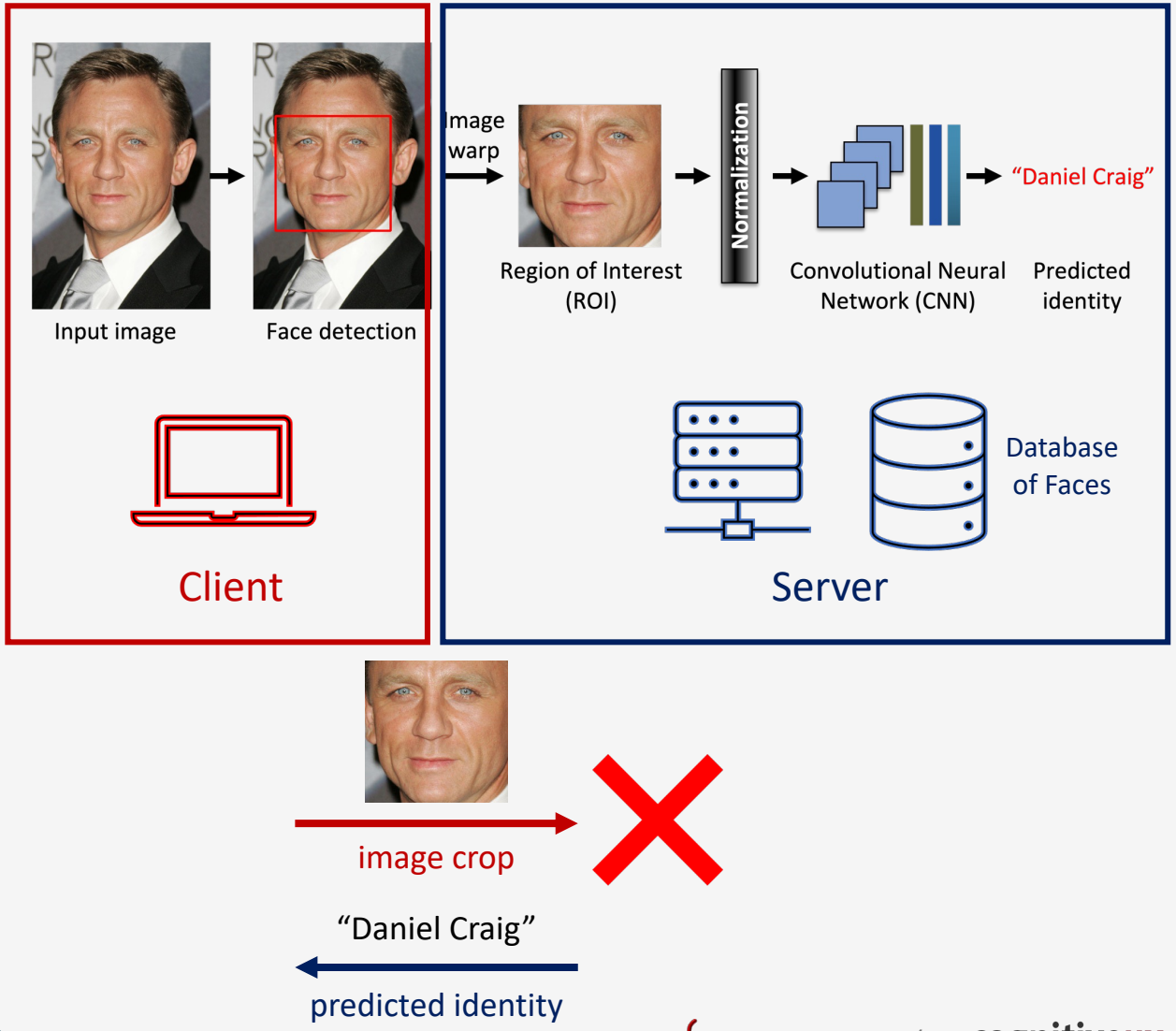
Privacy & Computational Concerns

Co-funded by the
Erasmus+ Programme
of the European Union

- Face Recognition System v1.0:

- Server:
  - User account / login.
  - Application management.
  - Image Warp / ROI.
  - **Convolutional Neural Network inference.**
  - Identity prediction.

- Client:
  - Image acquisition.
  - Face detection.



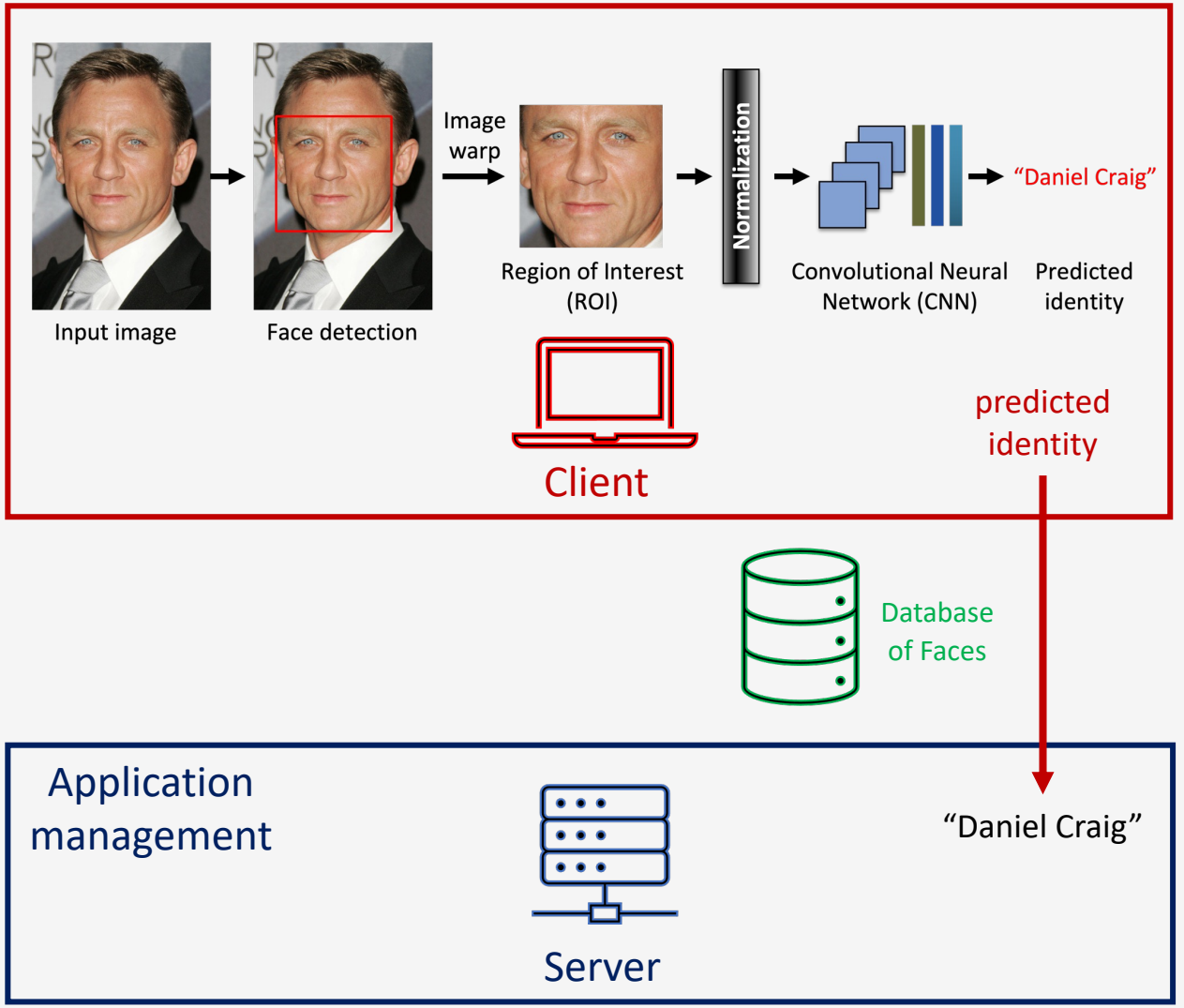
Privacy & Computational Concerns

Co-funded by the Erasmus+ Programme of the European Union

- Possible Solution?:

- Server:
  - User account / login.
  - Application management.

- Client:
  - Image acquisition.
  - Face detection.
  - Image Warp / ROI.
  - **Convolutional Neural Network inference.**
  - Identity prediction.



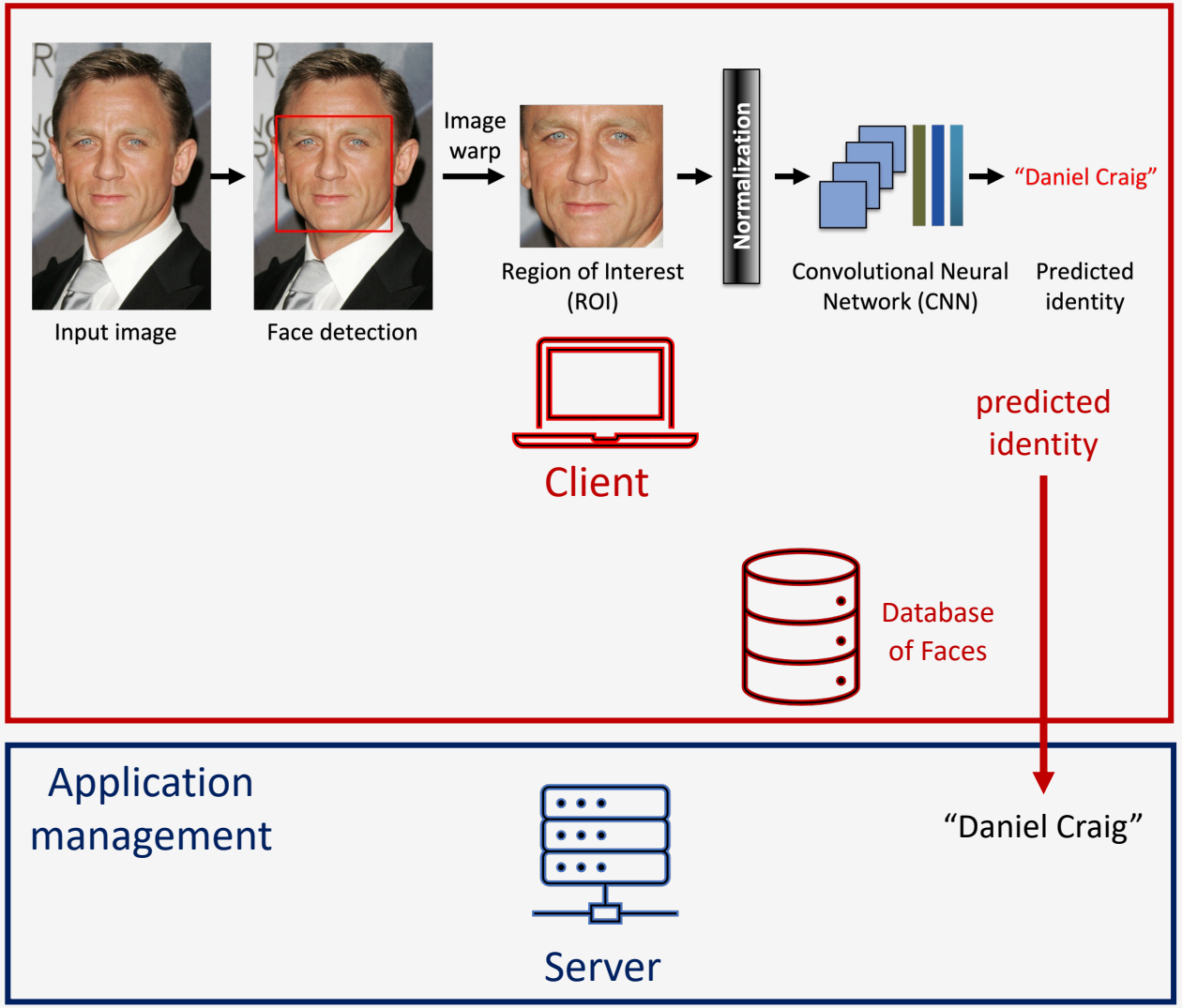
Privacy & Computational Concerns

Co-funded by the Erasmus+ Programme of the European Union

- Possible Solution?:

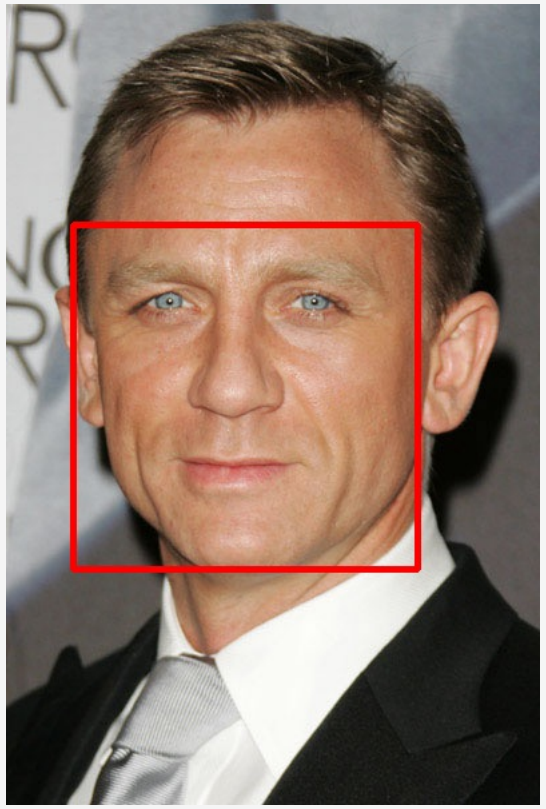
- Server:
  - User account / login.
  - Application management.

- Client:
  - Image acquisition.
  - Face detection.
  - Image Warp / ROI.
  - **Convolutional Neural Network inference.**
  - Identity prediction.



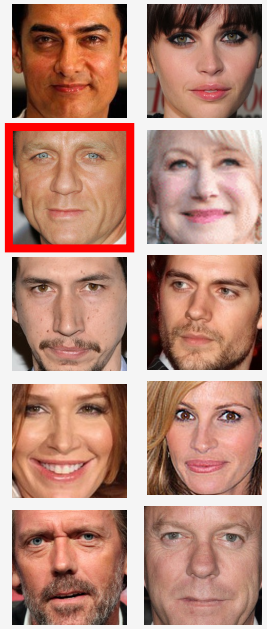
Face Recognition vs Face Verification

Co-funded by the
Erasmus+ Programme
of the European Union

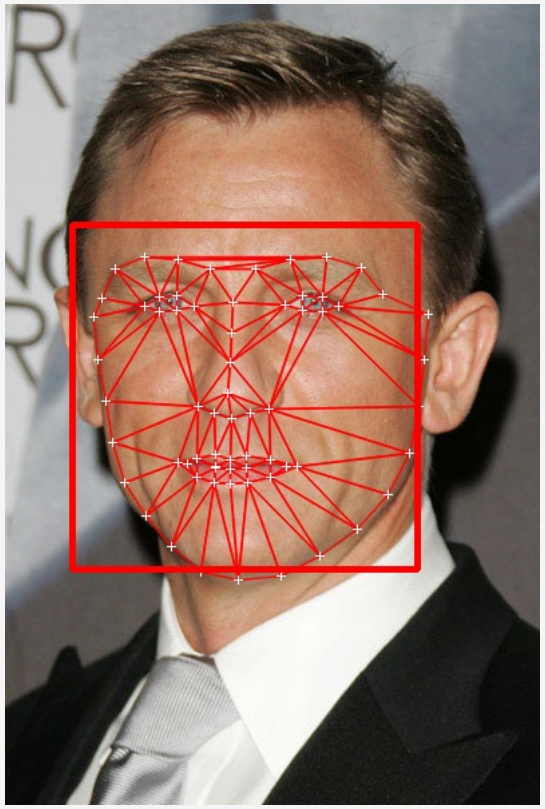


Face Recognition

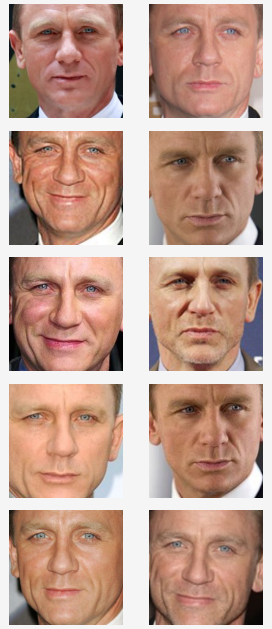
Query a database
of face images



Data from multiple
individuals



Face Verification



Data from a single
individual

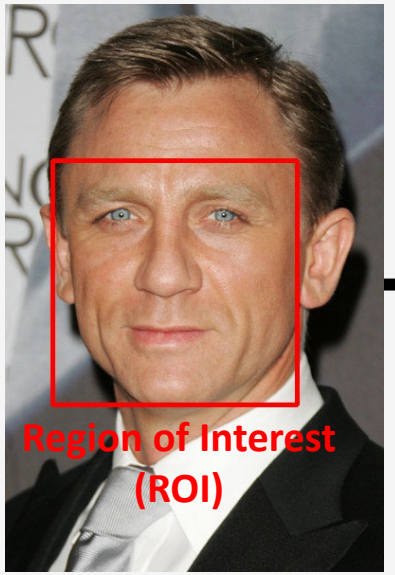
Who is this person? → Is “Daniel Craig”

Is this person, Daniel Craig? → “Yes”

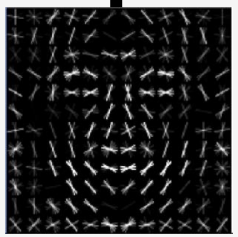
TrustID - Face Verification System v2.0

Co-funded by the
Erasmus+ Programme
of the European Union

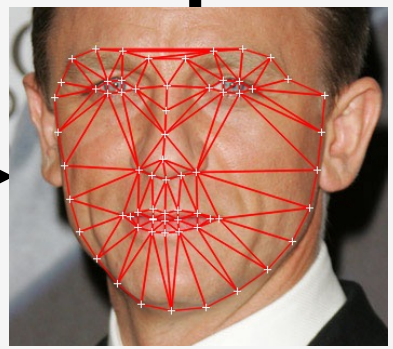
Input image



Face detection



Discriminative template matching (HoG + linear SVM)



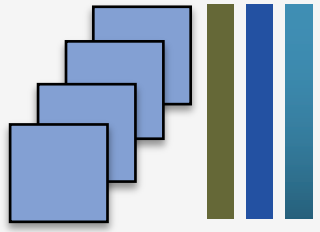
Facial Alignment

Warped Image

Normalization

zero mean
& unit std

ResNet-34 network architecture
Triplet loss function



Convolutional Neural Network (CNN)

Face Embedding

Query Embedding Statistics

$\mathbf{x} \in \mathcal{R}^{128} < \mu, \Sigma >$

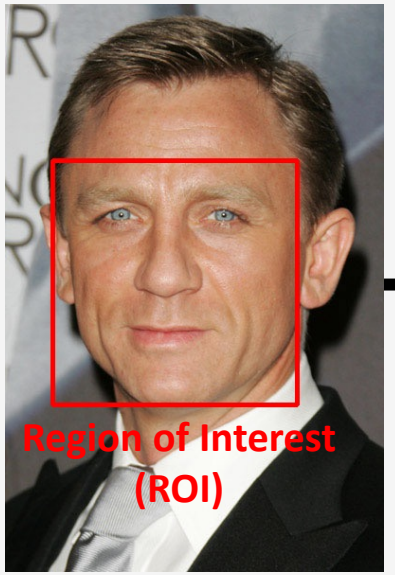
Yes is
"Daniel Craig"

Validate identity

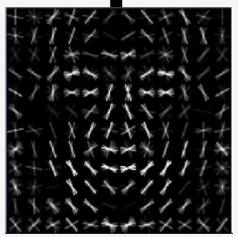
TrustID - Face Verification System v2.0

Co-funded by the
Erasmus+ Programme
of the European Union

Input image

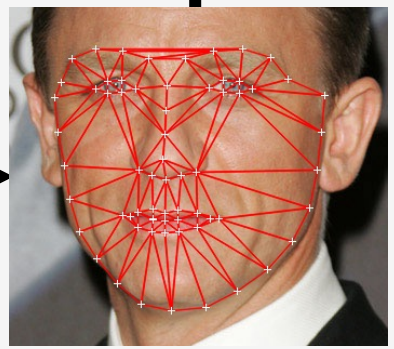


Face detection



Discriminative template matching (HoG + linear SVM)

Warped Image

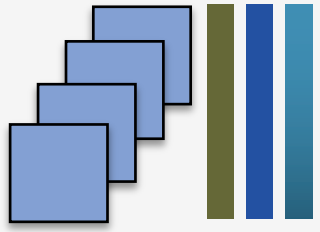


Facial Alignment

Normalization

zero mean
& unit std

ResNet-34 network architecture
Triplet loss function



Face Embedding

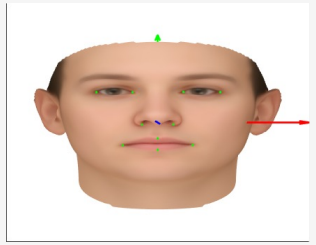
Query Embedding Statistics

Convolutional Neural Network (CNN)

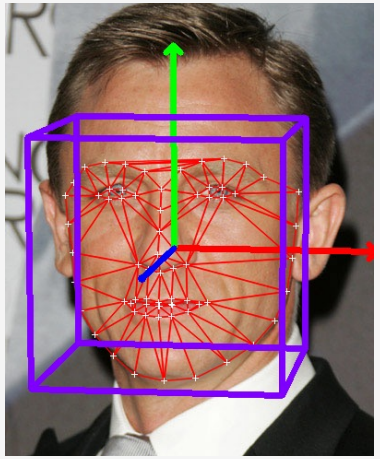
$\mathbf{x} \in \mathcal{R}^{128} < \mu, \Sigma >$

Yes is
"Daniel Craig"

Validate identity



Rigid 3D Head Model



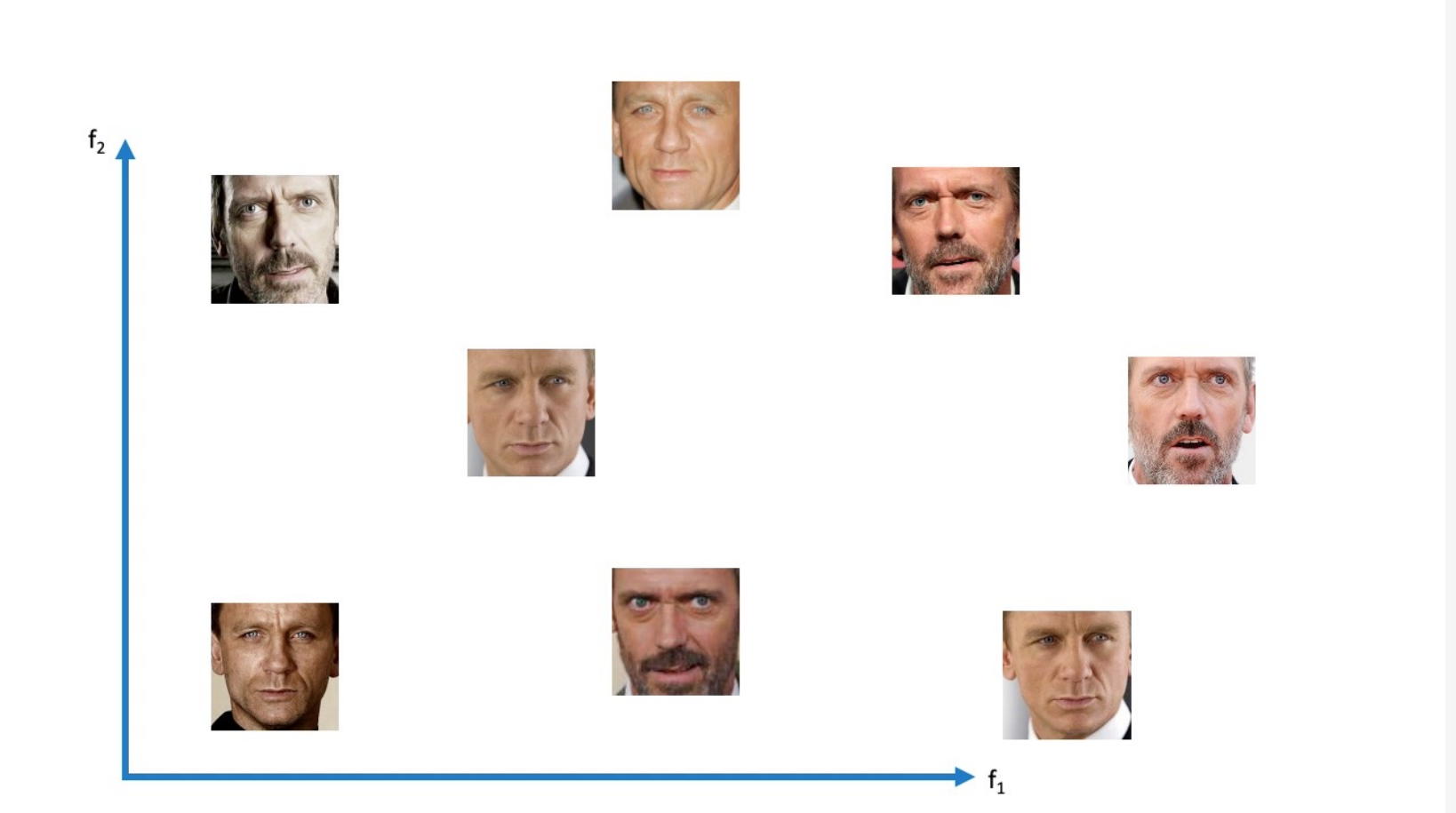
3D Head Pose Estimation

$< \mathbf{t}, \alpha, \beta, \gamma >$

3D Pose

Metric Learning w/ Triplet Loss

Co-funded by the
Erasmus+ Programme
of the European Union

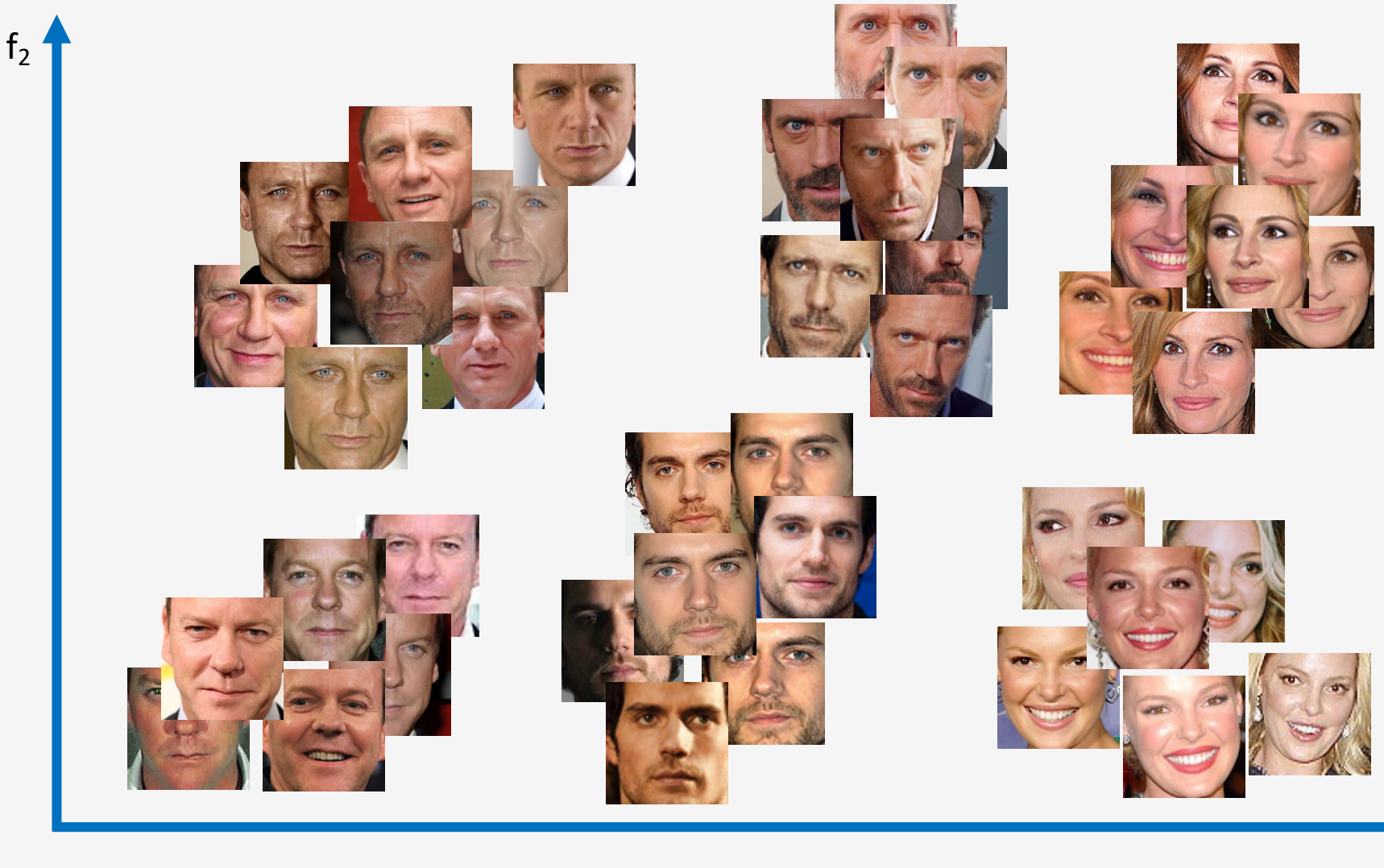


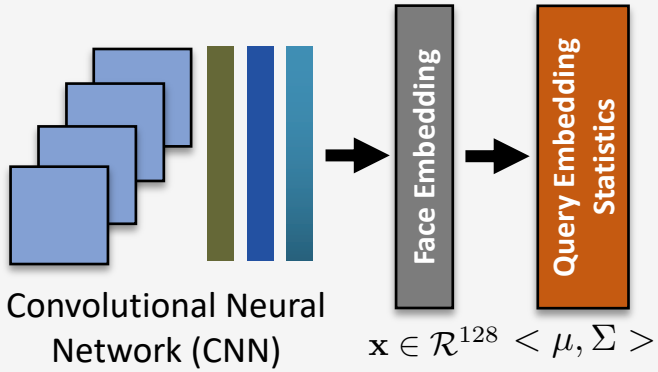
$$L(a, p, n) = \max(\|a - p\|^2 - \|a - n\|^2) + \Delta, 0)$$

anchor positive anchor negative margin

Embedding Space

Co-funded by the
Erasmus+ Programme
of the European Union

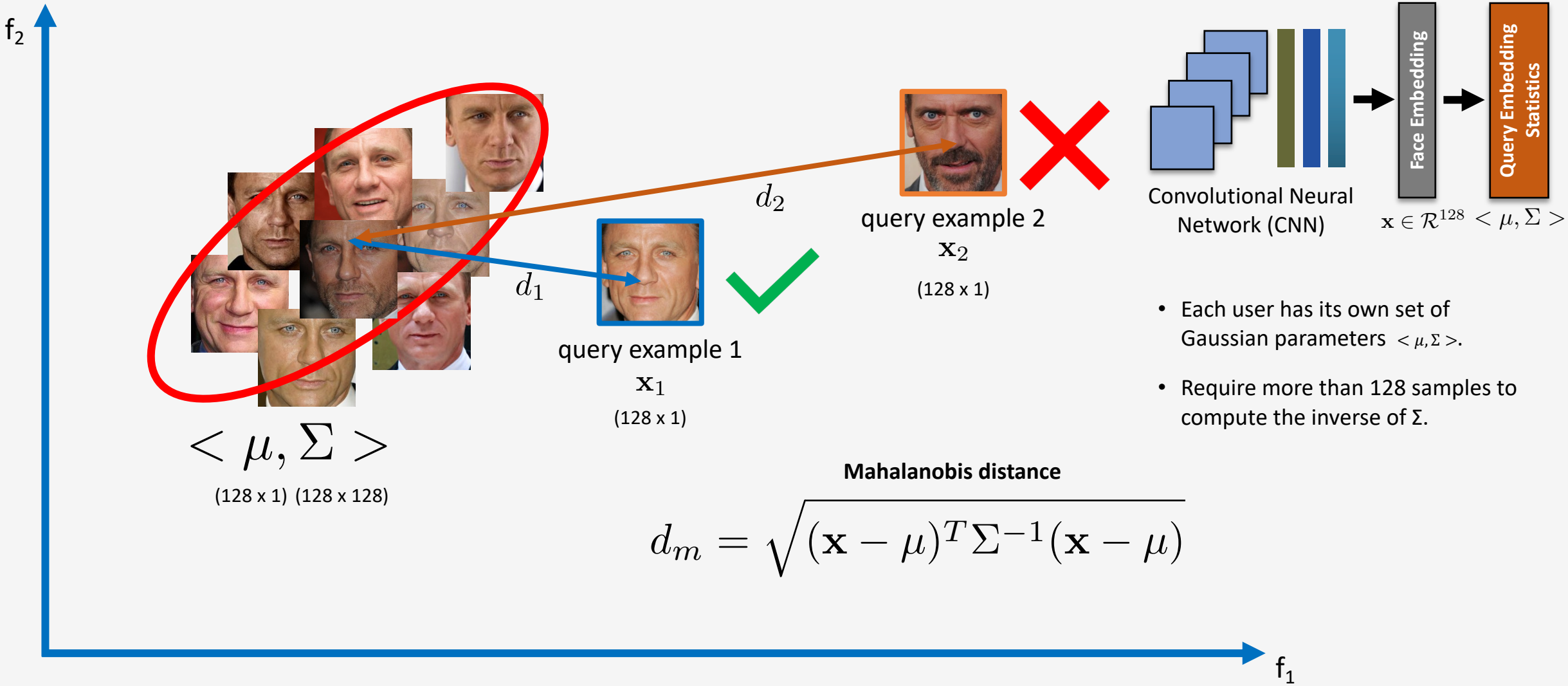




- Each normalized face is represented by a 128 dimensional vector.
- It is assumed that the training data for a single user, spans across a multidimensional Gaussian distribution.

Identity Validation Strategy

Co-funded by the Erasmus+ Programme of the European Union



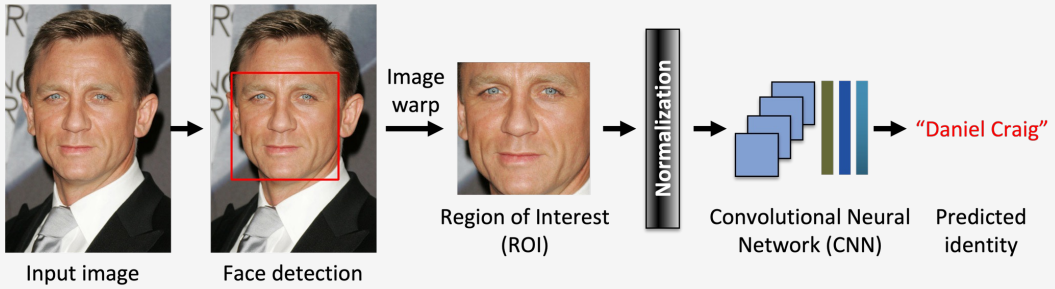
- Each user has its own set of Gaussian parameters $< \mu, \Sigma >$.
- Require more than 128 samples to compute the inverse of Σ.

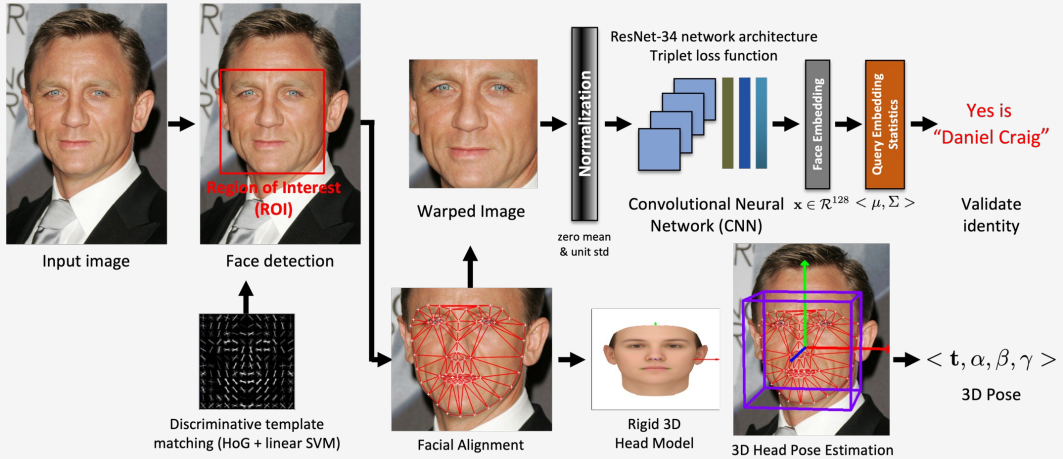
Mahalanobis distance

$$d_m = \sqrt{(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)}$$

Privacy & Computational Concerns v2.0

Co-funded by the
Erasmus+ Programme
of the European Union



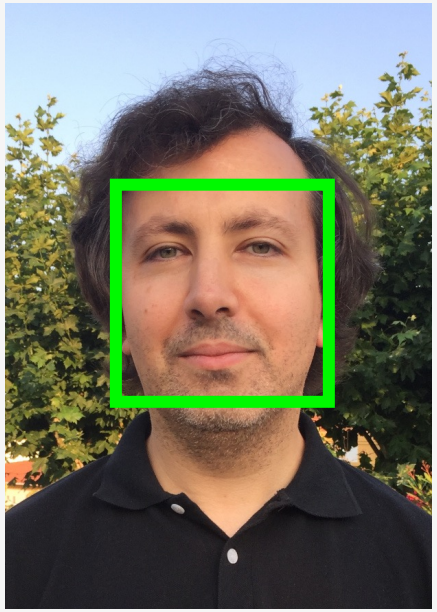


- Face Recognition System v1.0:
- Server:
  - User account / login.
  - Application management.
  - Image Warp / ROI.
  - Convolutional Neural Network inference.
  - Identity Prediction
- Client:
  - Image acquisition.
  - Face detection.

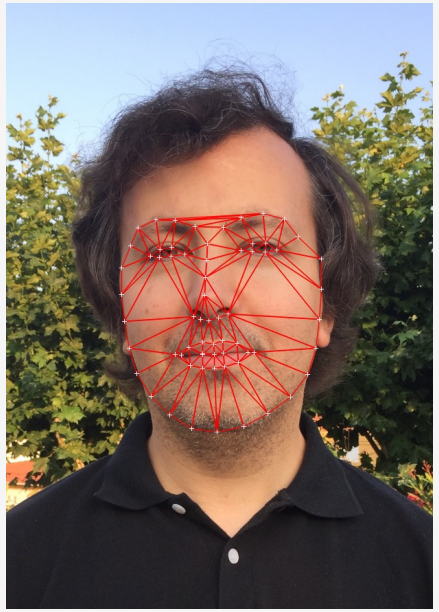
- Face Recognition System v2.0:
- Server:
  - User account / login.
  - Application management.
- Client:
  - Image acquisition.
  - Face detection.
  - Face Alignment + Image Warp / ROI.
  - Head Pose Estimation.
  - Convolutional Neural Network inference.
  - Identity validation.

Auxiliary Face Modules

Co-funded by the
Erasmus+ Programme
of the European Union

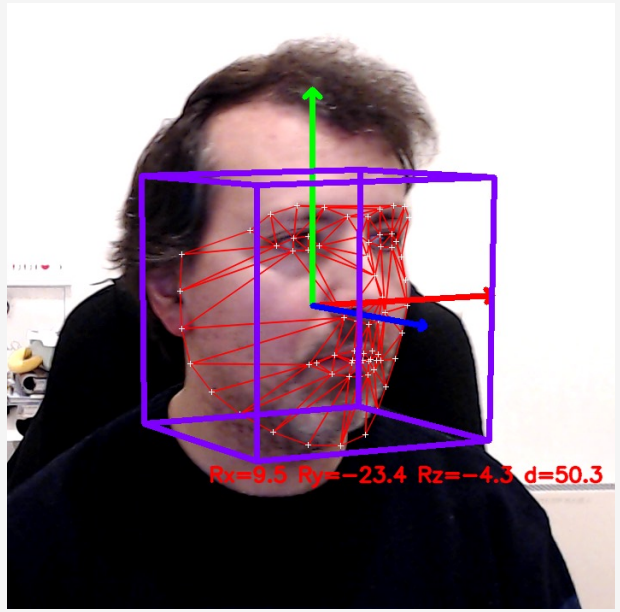


Face Detection



Face Alignment

(facial landmaks localization)



3D Head Pose Estimation

Face Alignment (Cascaded Regression)

Co-funded by the
Erasmus+ Programme
of the European Union

$$\mathbf{s}^k = \mathbf{s}^{k-1} + \mathbf{R}^{k-1}\mathcal{F}(\mathbf{I}, \mathbf{s}^{k-1})$$

k **- cascade level**

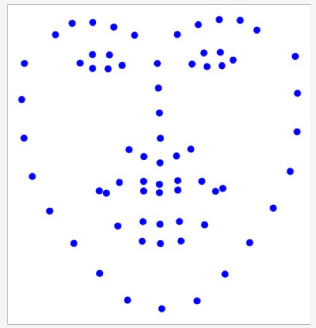
Updated shape vector

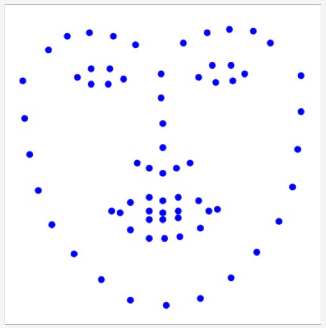
Previous shape vector

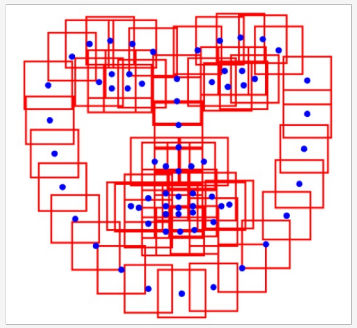
Regression Matrix

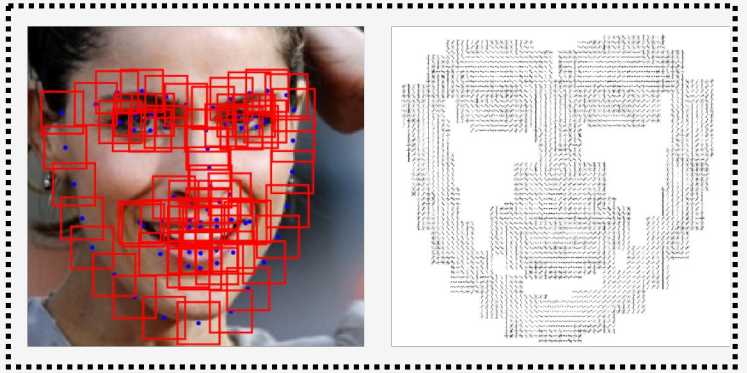
Feature Extraction

$$\mathbf{s} = \begin{pmatrix} x_0 \\ \vdots \\ x_v \\ y_0 \\ \vdots \\ y_v \end{pmatrix}$$









v **- landmarks**

$2v \times 1$

$2v \times 1$

$2v \times d$

$d \times 1$

RGB

HoG

Face Alignment (Cascaded Regression)

Co-funded by the
Erasmus+ Programme
of the European Union

Estimate $\mathbf{R}^k$ under Multiple Initializations

$$\arg\min_{\mathbf{R}^k} \sum_{i=1}^{N} \sum_{j=1}^{M} \left\| \Delta \mathbf{s}_j^k - \mathbf{R}^k \mathcal{F}(\mathbf{I}_i, \mathbf{s}_j^k) \right\|^2$$

k - cascade level
i - training image
j - virtual sample

Estimate noise

$$\Sigma^k = \mathrm{cov}(\mathbf{s}_* - \mathbf{s}_j^k)$$

Deviation from Ground Truth

Regression Labels

$$\Delta \mathbf{s}_j^k = \mathbf{s}_* - \mathbf{s}_j^k$$

Data Matrix (all features)

$$\mathbf{F} = [\ \ |\ \ |\ \ |\ \ ]$$

N images x M virtual samples

Least Squares Solution

$$\mathbf{R}^k = \Delta \mathbf{S} \left(\mathbf{F}^T \mathbf{F} \right)^{-1} \mathbf{F}^T$$

Linear Regression

Data Collection (F matrix)

$$\mathbf{s}_j^k \sim \mathcal{N}(\mu^k, \Sigma^k)$$

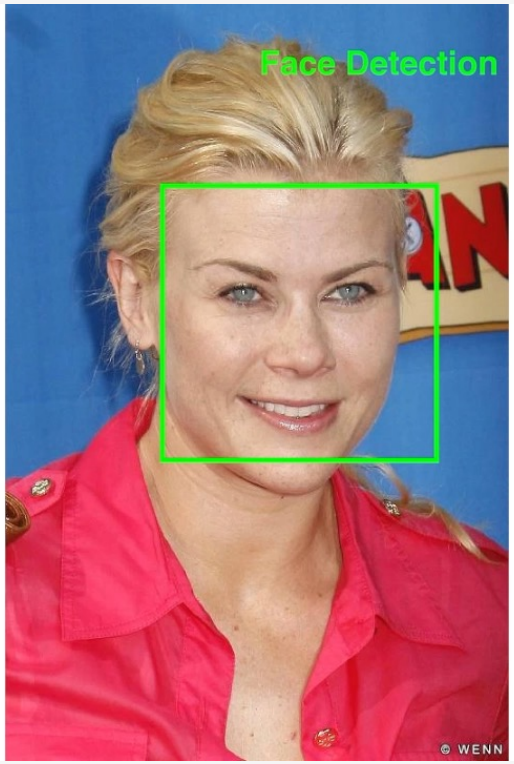
virtual sample

Face Alignment (Cascaded Regression)

Co-funded by the
Erasmus+ Programme
of the European Union

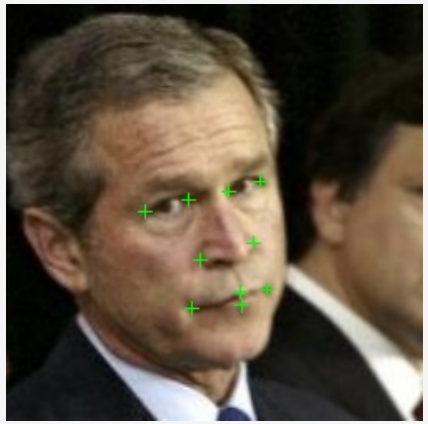
$$\mathbf{s}^k = \mathbf{s}^{k-1} + \mathbf{R}^{k-1}\mathcal{F}(\mathbf{I}, \mathbf{s}^{k-1})$$

k **- cascade level**

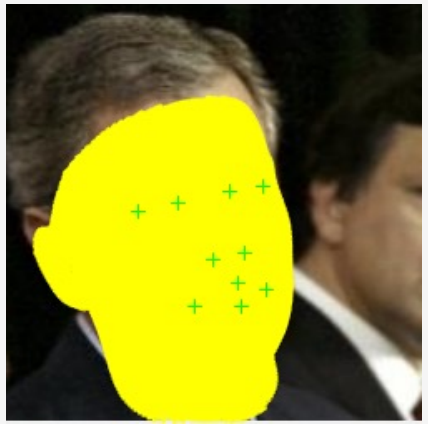


3D Head Pose Estimation (orientation + translation)

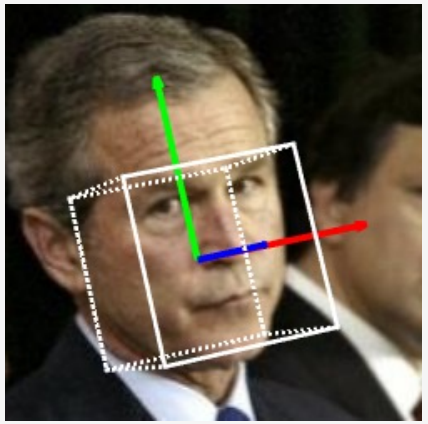
Co-funded by the
Erasmus+ Programme
of the European Union



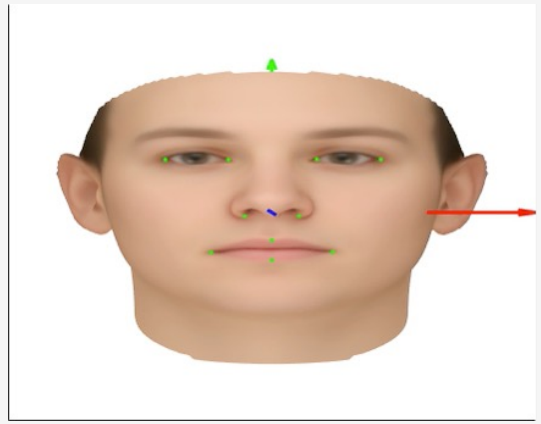
2D landmarks



3D model projection



3D pose representation



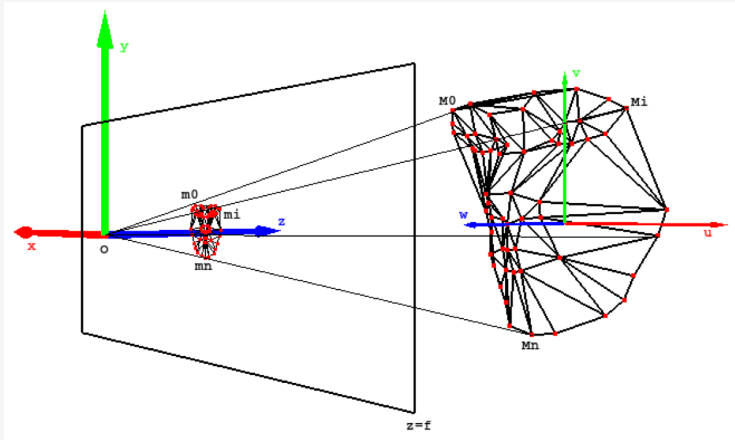
3D model

$$\arg\min_{\alpha,\beta,\gamma,\mathbf{t}} = \left(\mathbf{x}_{2D} - \mathbf{K}\left[\mathbf{R}_z(\alpha)\mathbf{R}_y(\beta)\mathbf{R}_x(\gamma), \mathbf{t}\right]\mathbf{X}_{3D}\right)^2$$

2D Projections — Camera — Euler Z-Y-X angles + 3D Translations — 3D Model

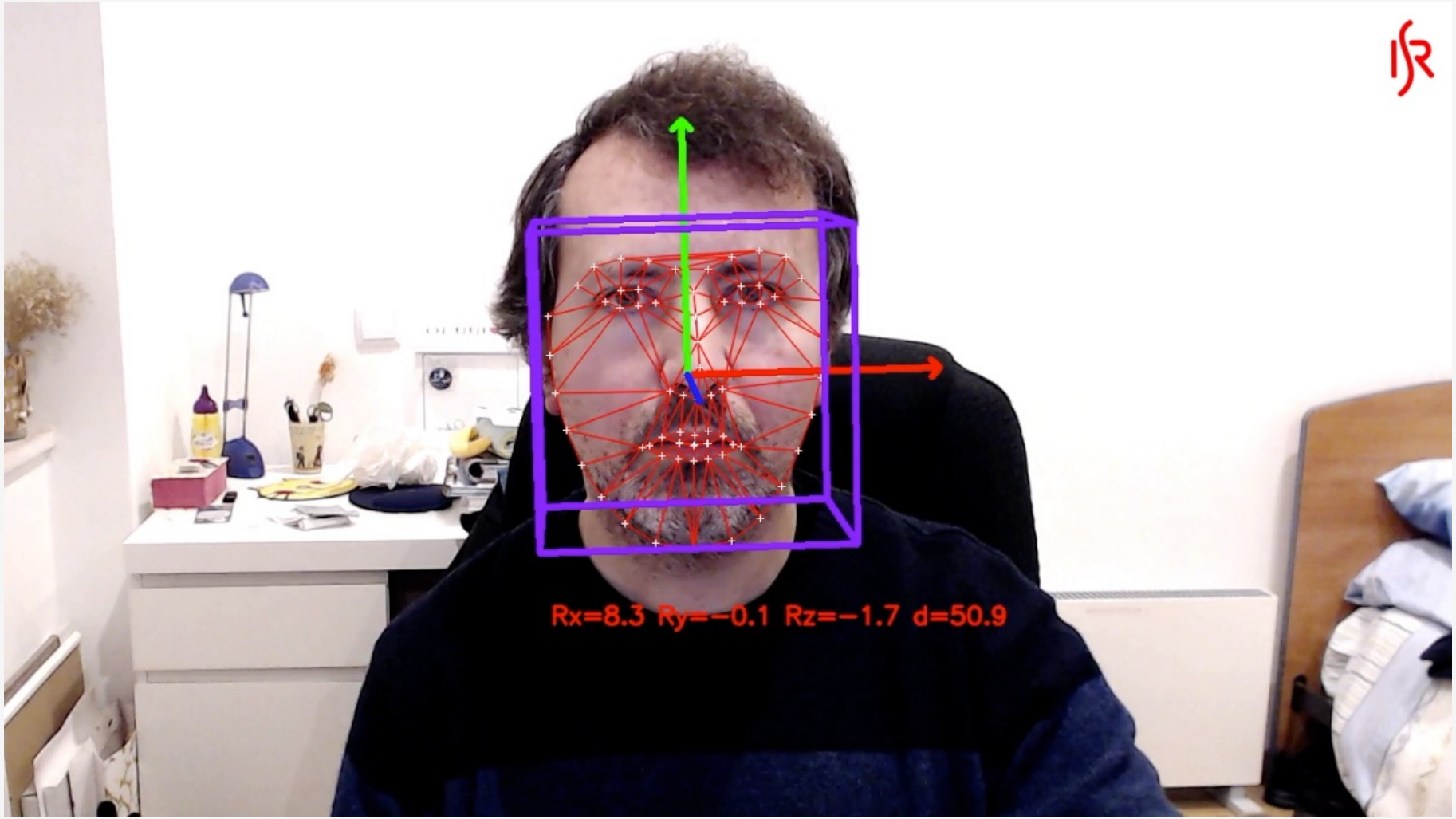
- Solve w/ Gauss-Newton method:

$$\begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \mathbf{t} \end{bmatrix}^{(i+1)} \leftarrow \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \mathbf{t} \end{bmatrix}^{(i)} - (\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T\mathbf{r}^{(i)}$$



3D Head Pose Estimation (video)

Co-funded by the
Erasmus+ Programme
of the European Union



Future Work

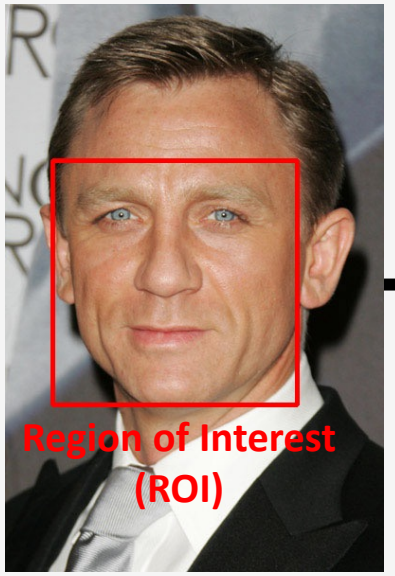
Co-funded by the
Erasmus+ Programme
of the European Union

- Face detection.
  - Operate on a larger range of head Poses.
  - Deal with Occlusion.
- Face Verification.
  - Improved face ROI normalization (3D Pose “frontalization”).
  - Global comparison loss (p.e. ArcFace).
- User Attention Metrics.
  - Eye gaze estimation.
  - Predominant head orientation, facial expression / emotion recognition.
- Liveness detection.
  - Basic motion estimation (blink detection, eye movement, mouth motion).
  - Texture analysis (p.e. LPBs, Fourier analysis).
  - Variable focusing analysis (variation of pixel values between two consecutive frames).
  - Use 3D facial shape information.
- Prevent Morphing Attacks.

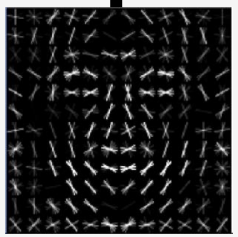
Future Work - Face Verification System v3.0

Co-funded by the
Erasmus+ Programme
of the European Union

Input image



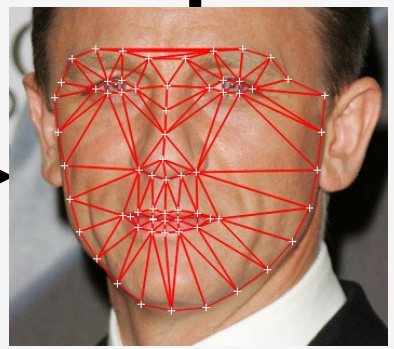
Face detection



Discriminative template matching (HoG + linear SVM)

PieceWise Affine Warp

Warped Image

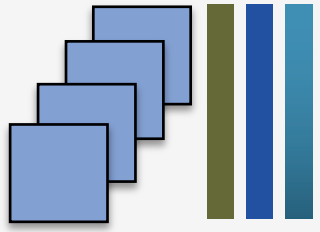


Facial Alignment

Normalization

zero mean
& unit std

ResNet-34 network architecture
Triplet loss function



Face Embedding

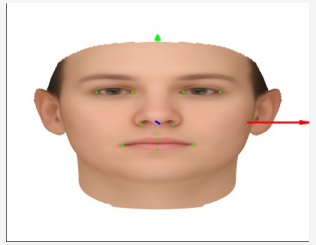
Query Embedding Statistics

Convolutional Neural Network (CNN)

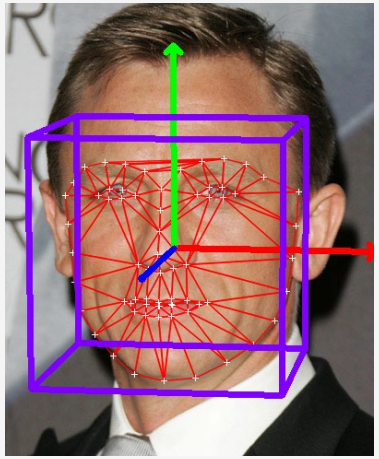
$\mathbf{x} \in \mathcal{R}^{128} < \mu, \Sigma >$

Yes is
"Daniel Craig"

Validate identity



Rigid 3D Head Model



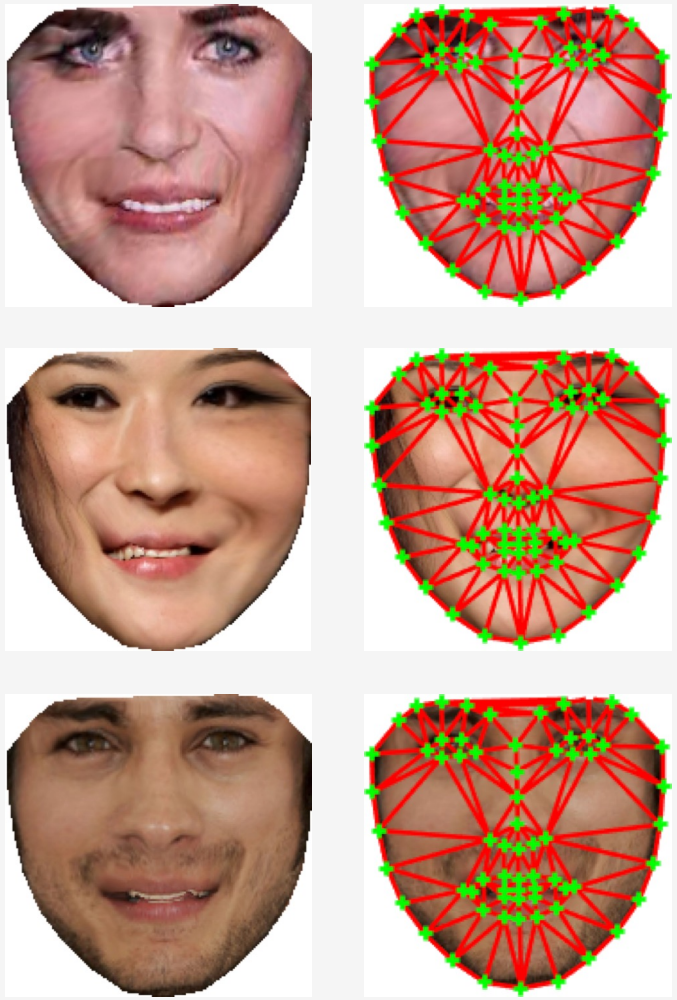
3D Head Pose Estimation

$< \mathbf{t}, \alpha, \beta, \gamma >$

3D Pose

”Pose Normalization” - Piecewise Affine Warp

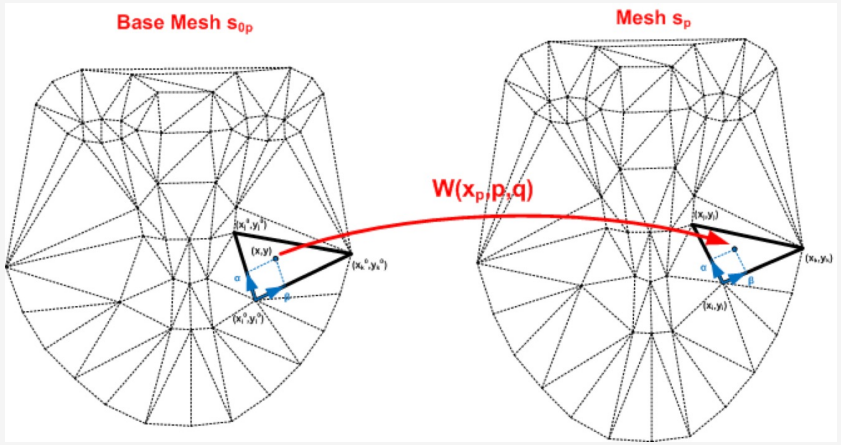
Co-funded by the
Erasmus+ Programme
of the European Union



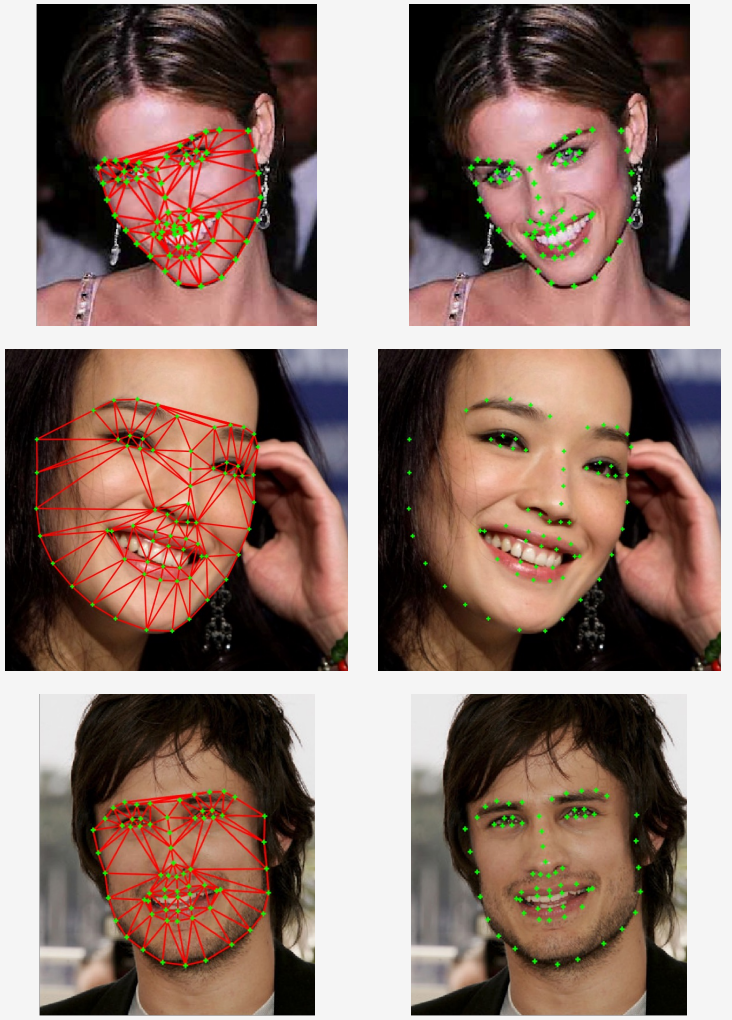
Warped Example

Base Mesh

Piecewise Affine Warp



$$\mathbf{W}(\mathbf{x}, \mathbf{p}) = \mathbf{x}_i + \alpha\left(\mathbf{x}_j - \mathbf{x}_i\right) + \beta\left(\mathbf{x}_k - \mathbf{x}_i\right)$$



Delaunay Triangulation

Landmarks

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ UNIVERSITY OF PATRAS
University of Cyprus
TRUSTID :: https://trustid-project.eu
INSTITUTO DE SISTEMAS E ROBÓTICA UNIVERSIDADE DE COIMBRA
cognitiveux

Co-funded by the
Erasmus+ Programme
of the European Union

Thank you.

Project webpage:

https://trustid-project.eu/