Towards an Evolution in the Characterization of the Risk of Re-identification of Medical Images

Antoine Boutet, Amine Dahmouni, Carole Frindel & Mohamed Maouche

Plan

- Context
- How to evaluate privacy?
- Face Recognition methods
- Experiments
- Ongoing Work
- Conclusion and perspectives

Context & Motivation: Hunger for data

- Imaging data increasingly shared for research purposes

Context & Motivation: Hunger for data

- Imaging data increasingly shared for research purposes

biobank uk

Imaging data For 50,000 (2020) … Exome sequencing For 470,000 (2022) | Genome sequencing For 500,000 (2023)

HCL HOSPICES CIVILS DE LYON #team HCL

TDM	IRM	RX standard	RX inter-ventionnelle	Écho	Mammo	Médecine nucléaire
110 000 examens/an	55 000 examens/an	300 000 examens/an	20 000 examens/an	80 000 examens/an	+ 4 500 examens/an	20 000 examens/an (8 000 TEP/TDM)

COLYBRI Plateform

Context & Motivation: Security Breaches

- Hospitals face cybersecurity attack leading to data breaches

Context & Motivation: Security Breaches

- Hospitals face cybersecurity attack leading to data breaches

French hospital suspends operations after cyber attacks

A hospital in Versailles, near Paris had to cancel operations and transfer some patients after being hit by a cyberattack over the weekend, France's health ministry said Sunday. Issued on: 05/12/2022 - 01:41

https://www.france24.com/en/france/20221205-french-hospital-suspends-operations-after-cyber-attacks

Context & Motivation: Security Breaches

- Hospitals face cybersecurity attack leading to data breaches

FRANCE 24

French hospital suspends operations after cyber attacks

A hospital in Versailles, near Paris had to cancel operations and transfer some patients after being hit by a cyberattack over the weekend, France's health ministry said Sunday. Issued on: 05/12/2022 - 01:41

https://www.france24.com/en/france/20221205-french-hospital-suspends-operations-after-cyber-attacks

Institution	TRICARE	CHS Community Health Systems, Inc.	UCLA		l'Assurance Maladie Agir ensemble, protéger chacun
Date	09/2011	06/2014	07/2015	…	03/2022
N° Patients impacted	5 millions	4.5 millions	4.5 millions		510,000

https://www.upguard.com/blog/biggest-data-breaches-in-healthcare

Context & Motivation : Easy Facial Recognition

- An increase in the potential of facial recognition software
- Easily available through commercial software
- Needs less qualifications to be used

Clearview.ai

Motivation: Mandatory Risk Assesment

Legal frameworks **mandate** the **quantification** of privacy risks

These laws require the consideration evolving practices, **available tools**, and **adversaries' capacities.**

In GDPR, there is a strong emphasis on considering **contextual factors and all potential identification methods**, especially in light of **technological advancements** and **increased computing power.**

Motivation: Better Evaluation Protocols

- Weak privacy protocol: lackluster methods to evaluate privacy risks of imaging data

Motivation: Better Evaluation Protocols

- Weak privacy protocol: lackluster methods to evaluate privacy risks of imaging data

- We can leverage important findings from other fields such as the voice privacy challenge
  - e.g., protocol, attackers, privacy metrics…

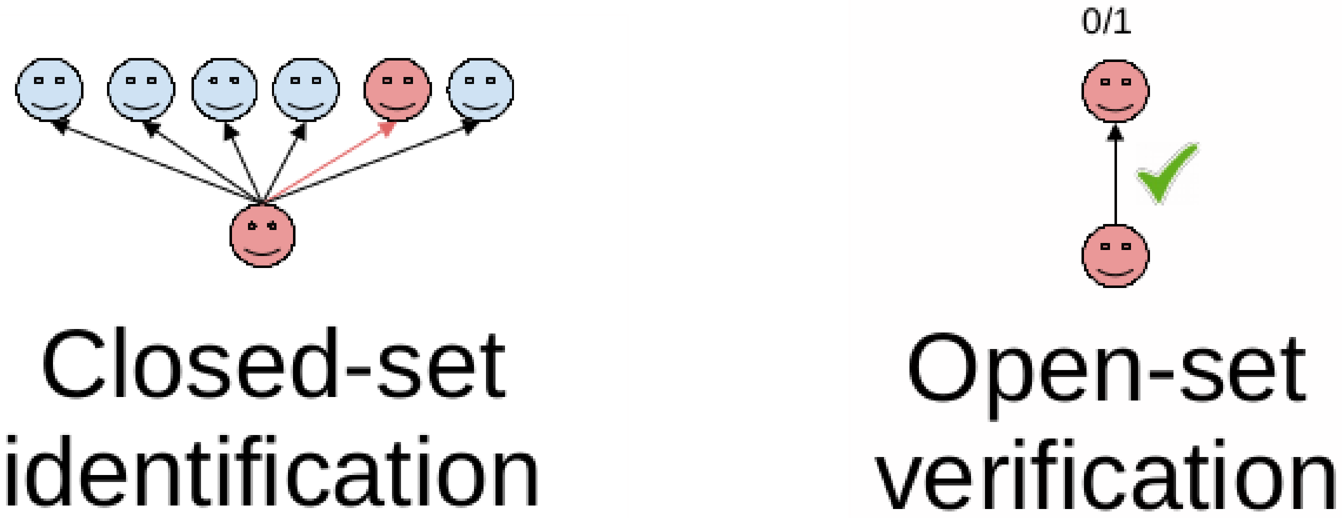
https://www.voiceprivacychallenge.org/

Privacy Evaluation Protocol

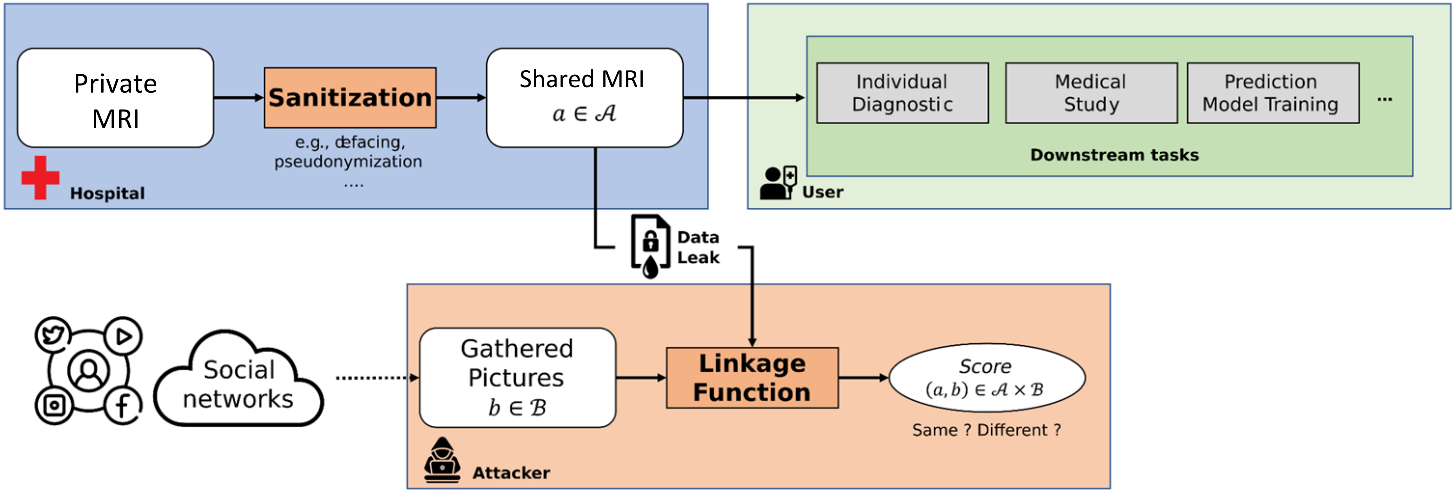
Identification vs Verification

Identification: What is the identity tied to this data?

Verification: Are those two data from the same user?



Evaluation of Privacy with Patient Verification

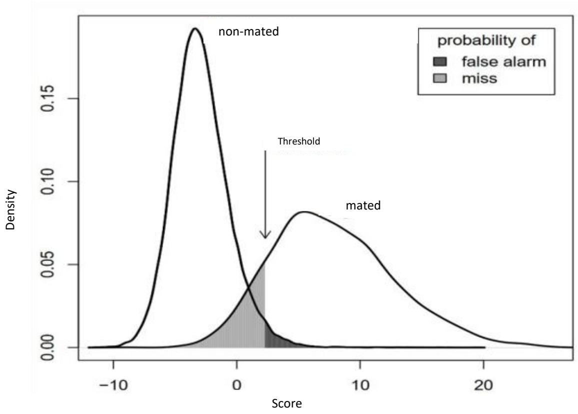


A sanitized shared data (trial set)
B gathered pictures with (enrollment set)
Considering a linkage function $\mathrm{LF}(a, b) = s$ $a \in A$ and $b \in B$
A pair (a, b) is called a trial, it is either mated H (i.e., same patients) or non-mated $\overline{H}$

A quick word on privacy metrics [Maouche et al. INTERSPEECH 20']

Different Metrics are applied:

$EER = P_{fa}(t^*) = P_{miss}(t^*)$
$EER \in [0, 0.5]$
Higher means more errors
$\Longrightarrow$ More Privacy
Random Guess = 0.5



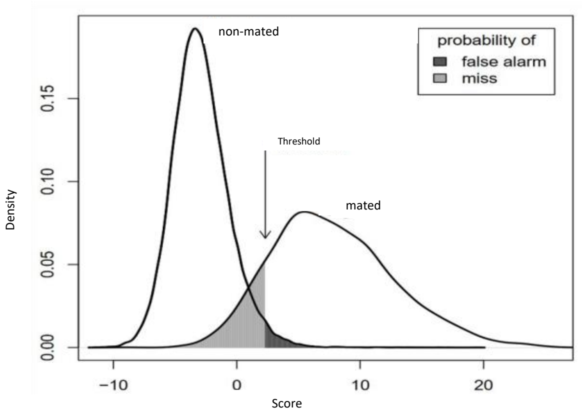
[DA Van Leeuwen & N Brümmer 2007]

A quick word on privacy metrics [Maouche et al. INTERSPEECH 20']

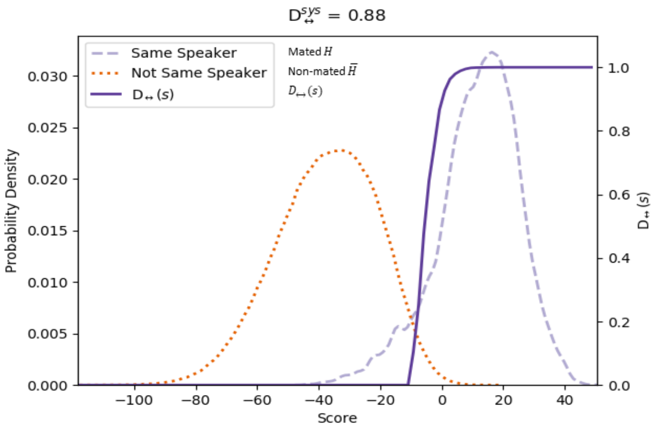
Different Metrics are applied:

$EER = P_{fa}(t^*) = P_{miss}(t^*)$
$EER \in [0, 0.5]$
Higher means more errors
$\Longrightarrow$ More Privacy
Random Guess = 0.5

$\text{Link}(s) = p(H \mid s) - p(\overline{H} \mid s) \cong D_{\leftrightarrow}(s)$
$Link \in [0, 1]$
Higher means more linkability
$\Longrightarrow$ Less Privacy
Random Guess = 0.0



[DA Van Leeuwen & N Brümmer 2007]



[Gomez-Barrero, et al. TIFS'2017]

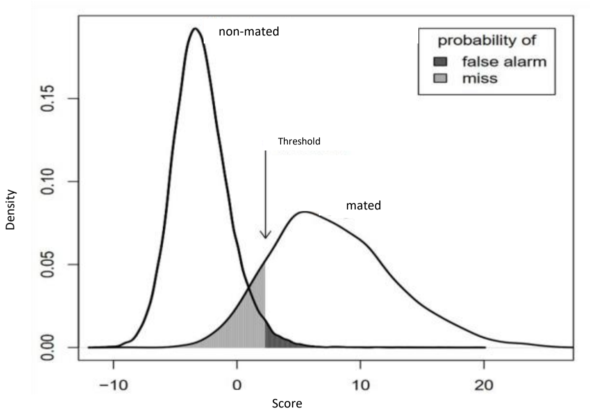
A quick word on privacy metrics [Maouche et al. INTERSPEECH 20']

Different Metrics are applied:

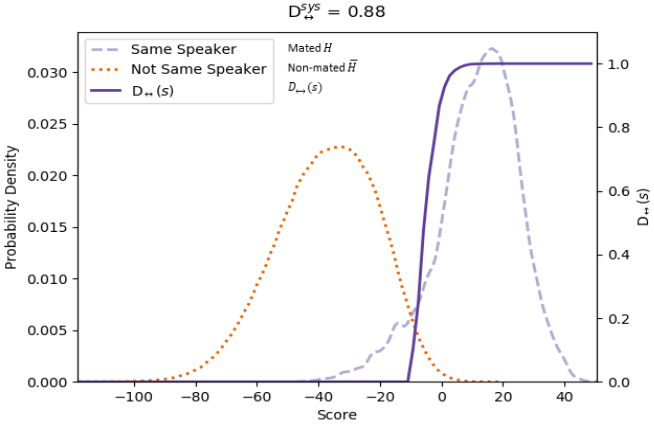
$EER = P_{fa}(t^*) = P_{miss}(t^*)$
$EER \in [0, 0.5]$
Higher means more errors
$\Longrightarrow$ More Privacy
Random Guess = 0.5

$\text{Link}(s) = p(H \mid s) - p(\bar{H} \mid s) \cong D_{\leftrightarrow}(s)$
$Link \in [0, 1]$
Higher means more linkability
$\Longrightarrow$ Less Privacy
Random Guess = 0.0

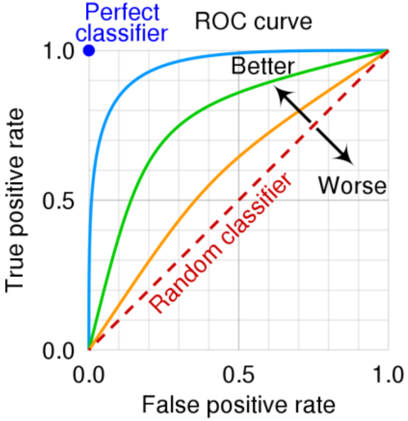
$\text{AUCROC} = \frac{1}{N_H N_{\bar{H}}} \sum_{s^H} \sum_{s^{\bar{H}}} 1(s^H > s^{\bar{H}})$
$\text{AUCROC} \in [0, 1]$
Higher means more linkability
$\Longrightarrow$ Less Privacy
Random Guess = 0.5



[DA Van Leeuwen & N Brümmer 2007]

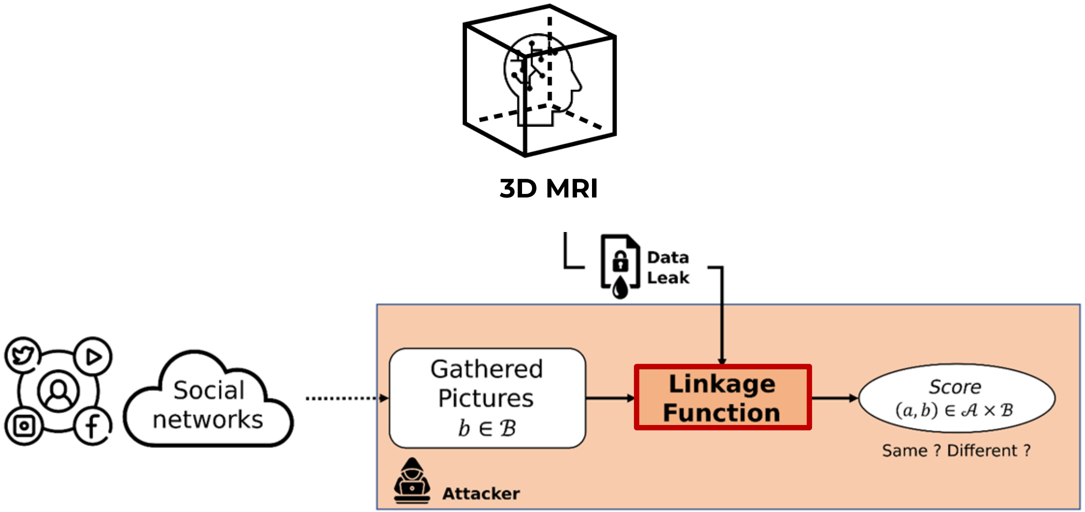


[Gomez-Barrero, et al. TIFS'2017]



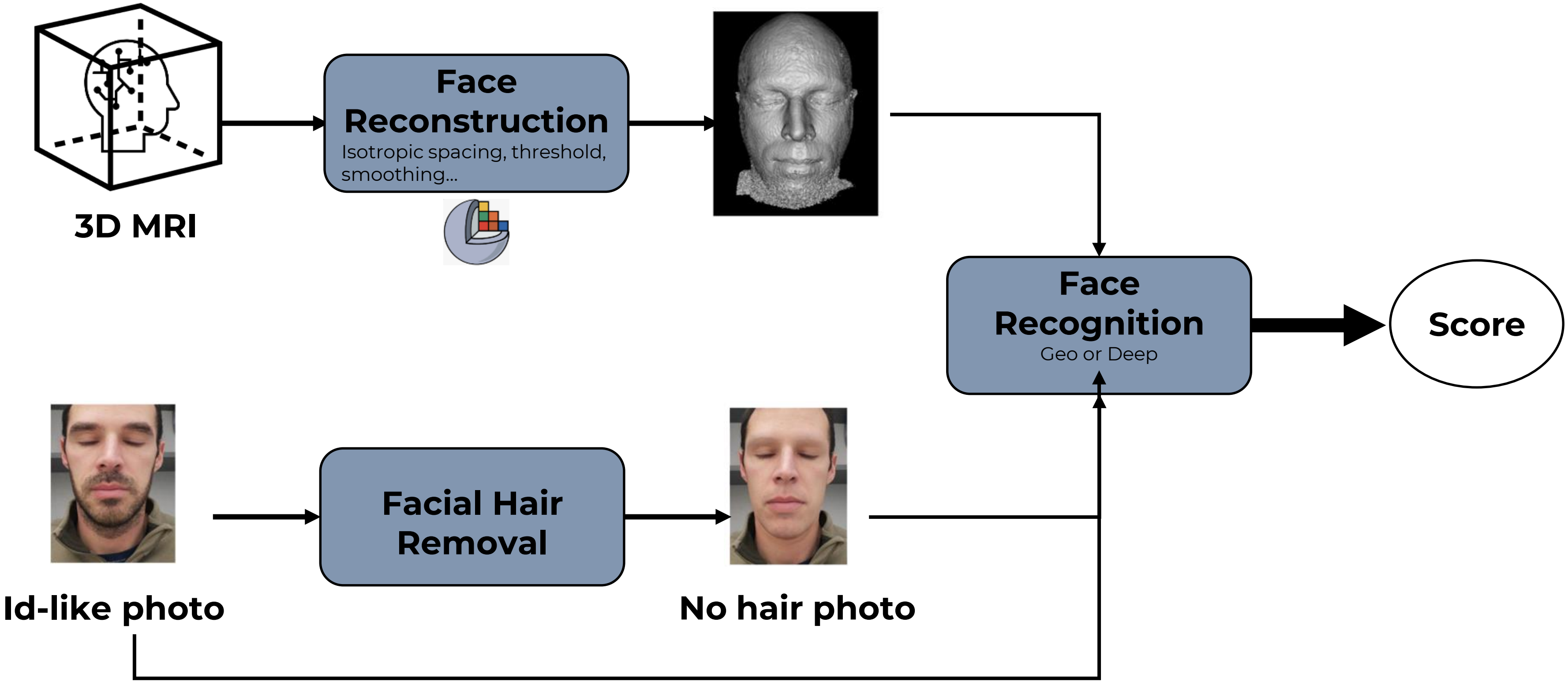
cmglee, MartinThoma
CC-BY-SA-4.0

Linkage Function



Linkage Attacks
Face recognition

General Algorithm



Face recognition

Geo Attack

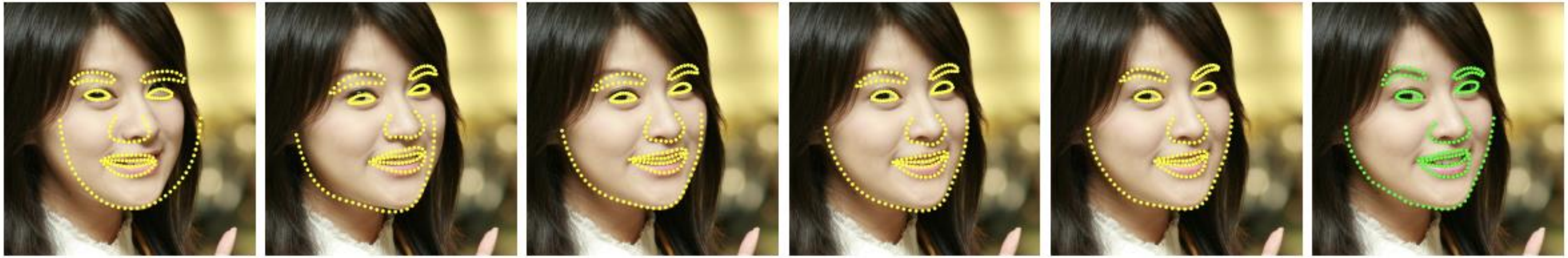
Extracting face's landmarks
We detect landmark on the 2D images using a model of the dlib package.
We compute geometrical features (distances, angles, ratios)

Deep Attack

deepFace

Facial recognition model (VGG-Face)
We extract embeddings from 2D images (photo or 2D MRI reconstruction)
We compute the distances between those vectors.

Geometrical features (Geo)



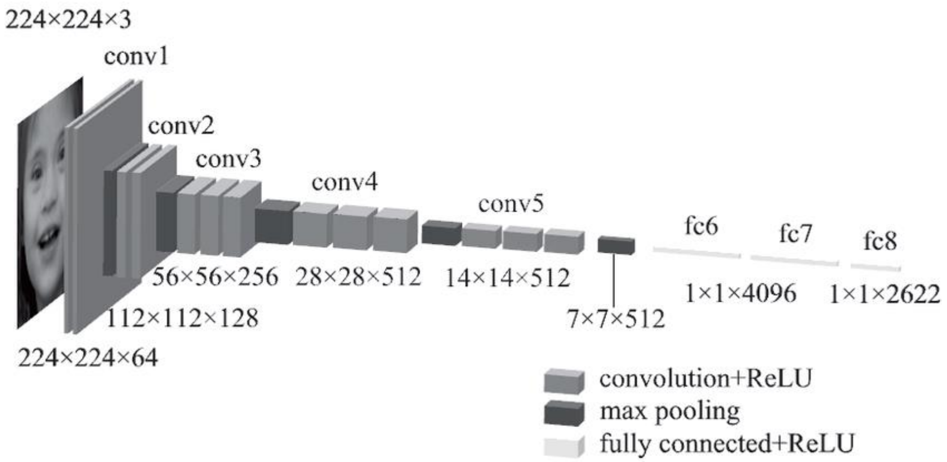
(a) $T = 0$ (b) $T = 1$ (c) $T = 2$ (d) $T = 3$ (e) $T = 10$ (f) Ground truth

Kazemi, CVPR, 2014

- **Cascaded regressors**
- Progressively refined and accurate facial landmark localization

http://dlib.net/face_landmark_detection.py.html

Deep learning features (Deep)

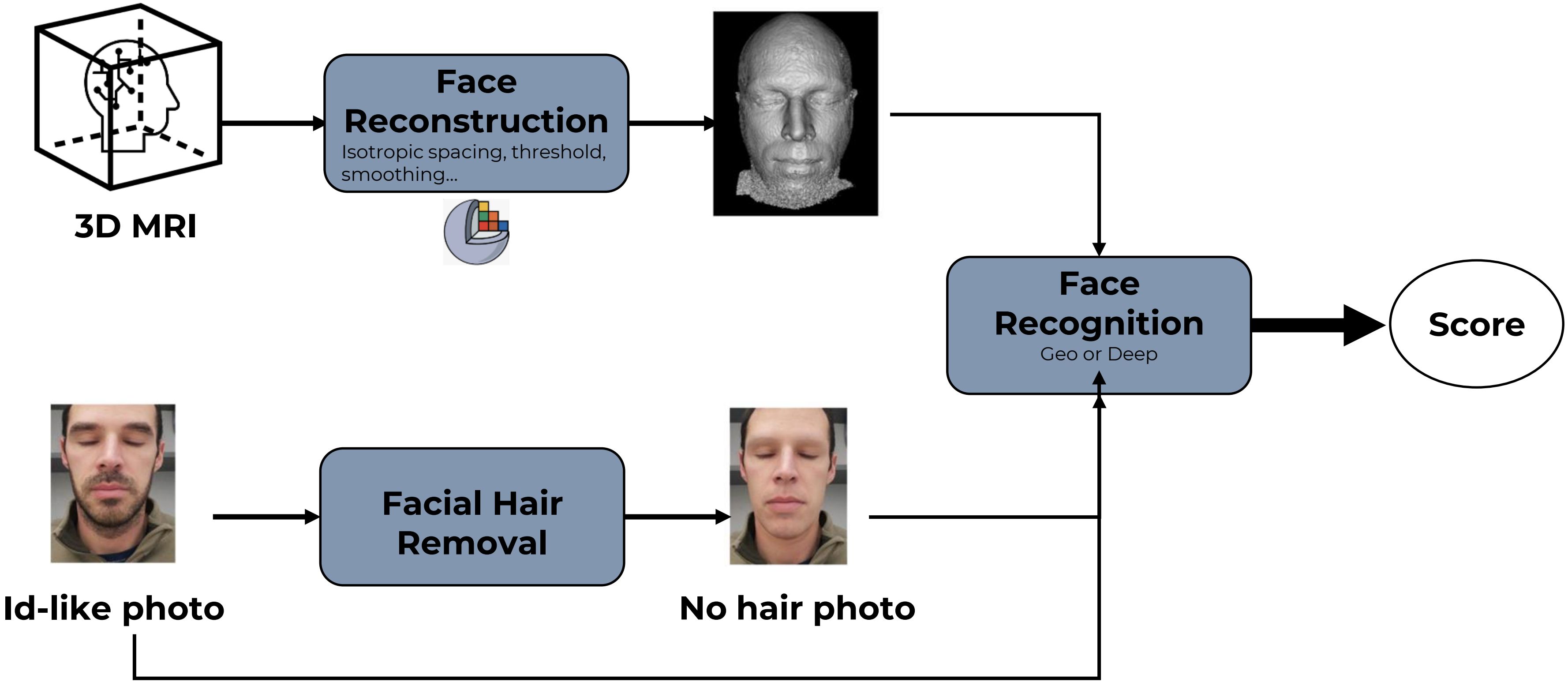


VGG Face [Parkhi, CVPR 2015]

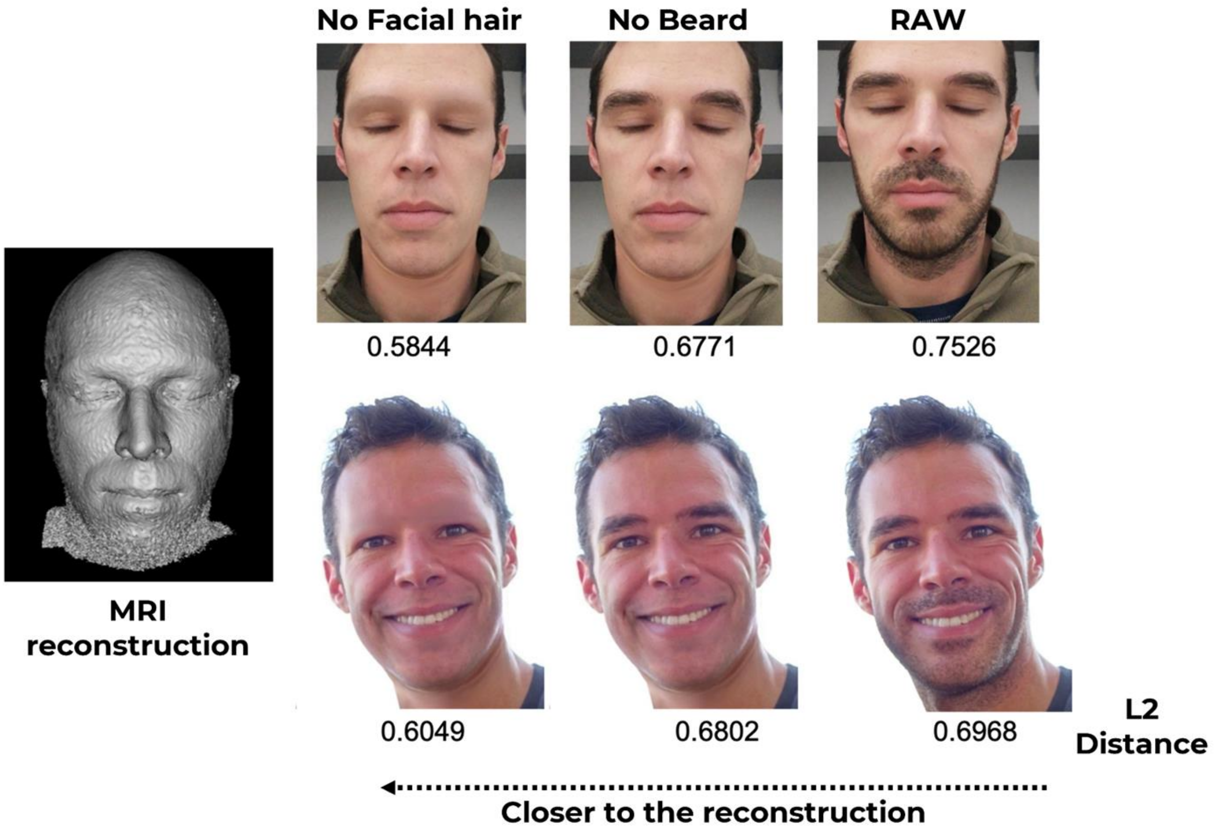
- Deep convolutional network trained on an **extensive dataset of facial images**
- Passes facial images through its layers
- Extracts high-level features at multiple abstraction levels

https://pypi.org/project/deepface/

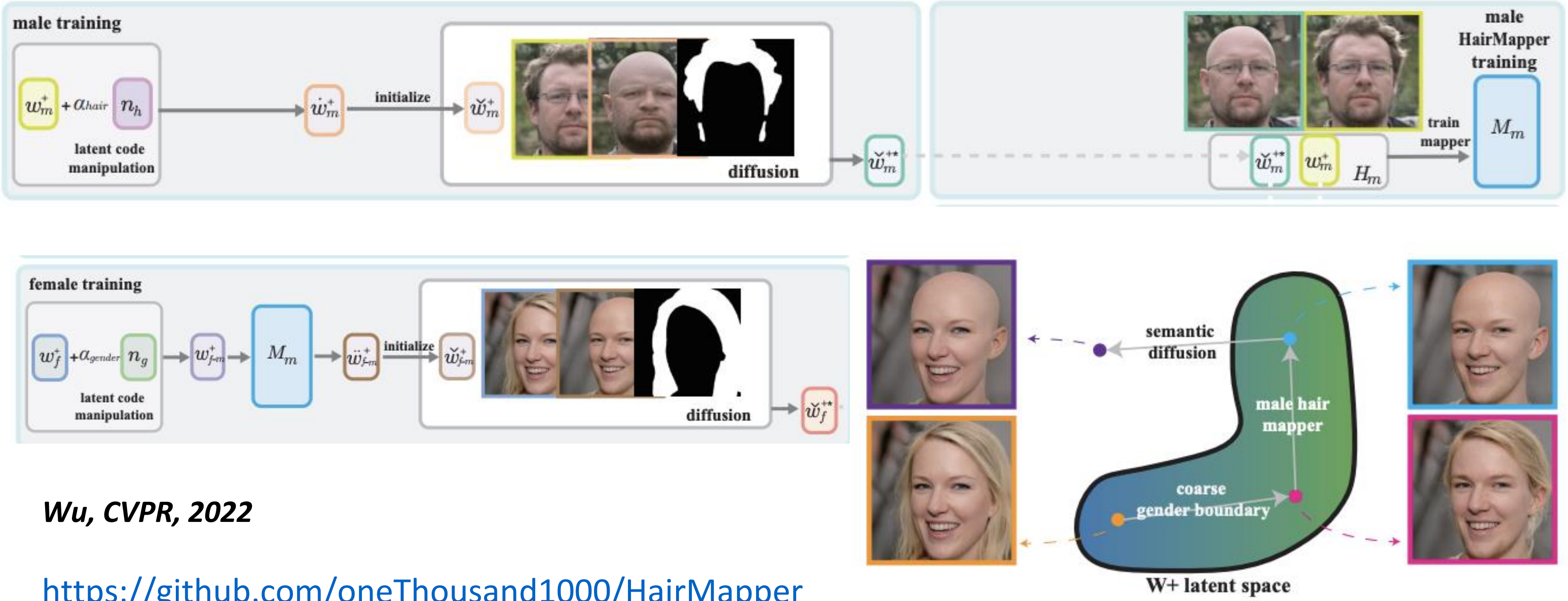
General Algorithm



Facial Hair Removal



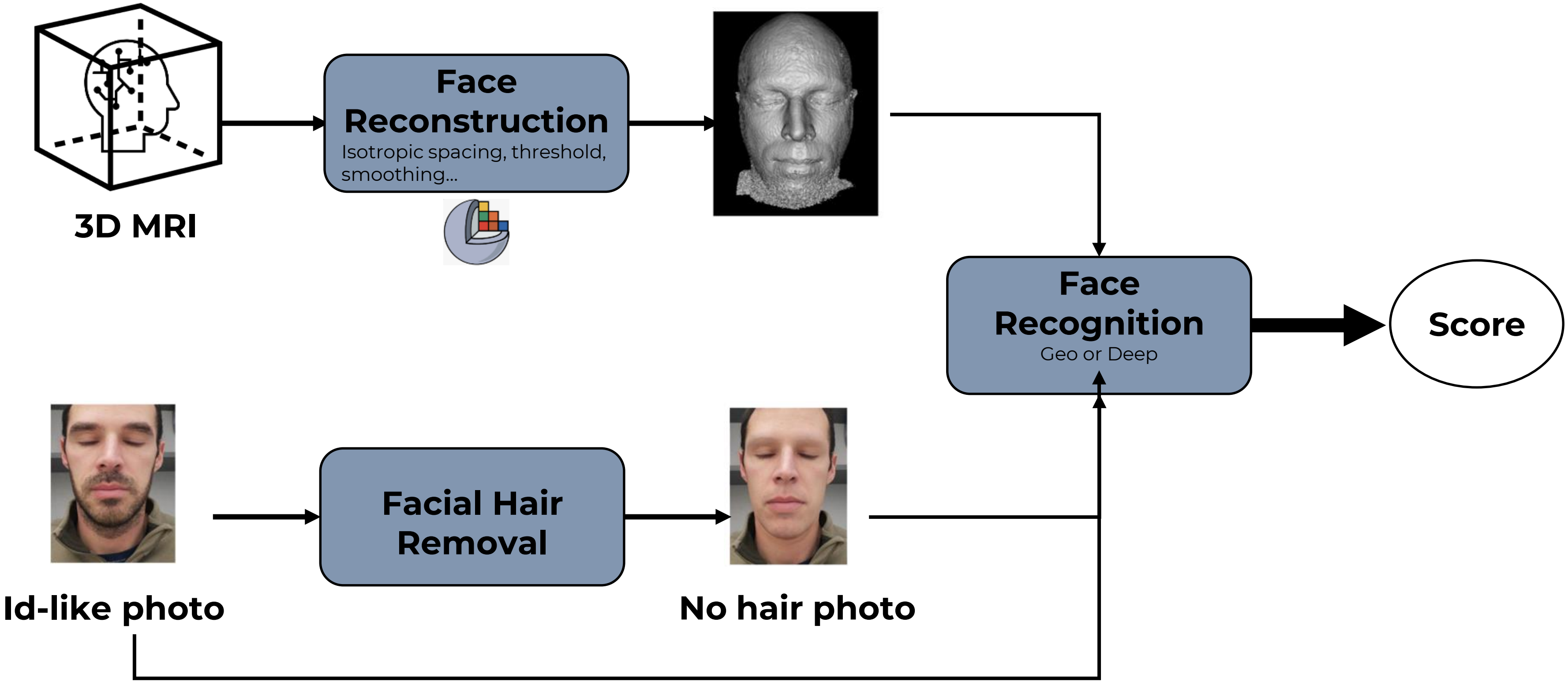
Facial hair removal



Wu, CVPR, 2022

https://github.com/oneThousand1000/HairMapper

General Algorithm



General Evaluation

Dataset

Data type: T2-weighted sagittal MRI imaging Turbo Spin Echo + photograph collected

#participants: 49 healthy volunteers

Age: 18-50

Location: HCL Lyon - Corentin Dauléac

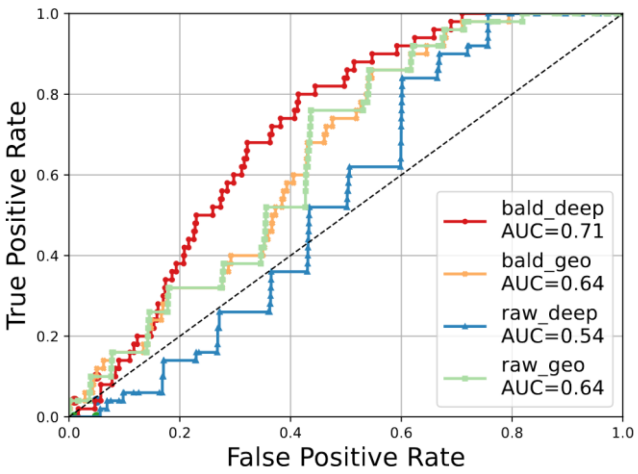
Dates: 02-04/2022

Each volunteer provided their informed consent to participate in the study and to be part of this work.

+ Social network photographs with Label Faces in the Wild (LFW) dataset (500 persons)

Results – Linkage attacks

Method	Facial hair	EER Max 50%	AUC Max 1	Linkability Max 1
Deepface	Raw	41	.54	.07
Deepface	Bald	**32**	**.71**	**.18**
Geo	Raw	36	.64	.11
Geo	Bald	38	.64	.13



Results are better than random

-> privacy leakage

Deepface is highly sensitive to facial hair

Removing hair increase the attack!

No impact of removing hair on Geo method

Results – Attribute inference

Protocol

We train using LFW a model on top of VGG-embeddings to infer sensitive attributes from ID images or MRI reconstructions.

Attributes & Metrics

Attribute	Task	#Modalities	Metric	Domain	Random Guess	Worst Privacy
Age	Regression		R^2	[0,1]	0	1
Gender	Classification	2	Accuracy	[0,1]	0.5	1
Ethnicity	Classification	6	Accuracy	[0,1]	0.17	1

Results – Attribute inference

Type	Age	Gender	Ethnicity
ID images	0.4	1.0	0.4
MRI reconstruciton	0.6	0.4	0.8

For Both Age & Ethnicity, the MRI leaks more information.

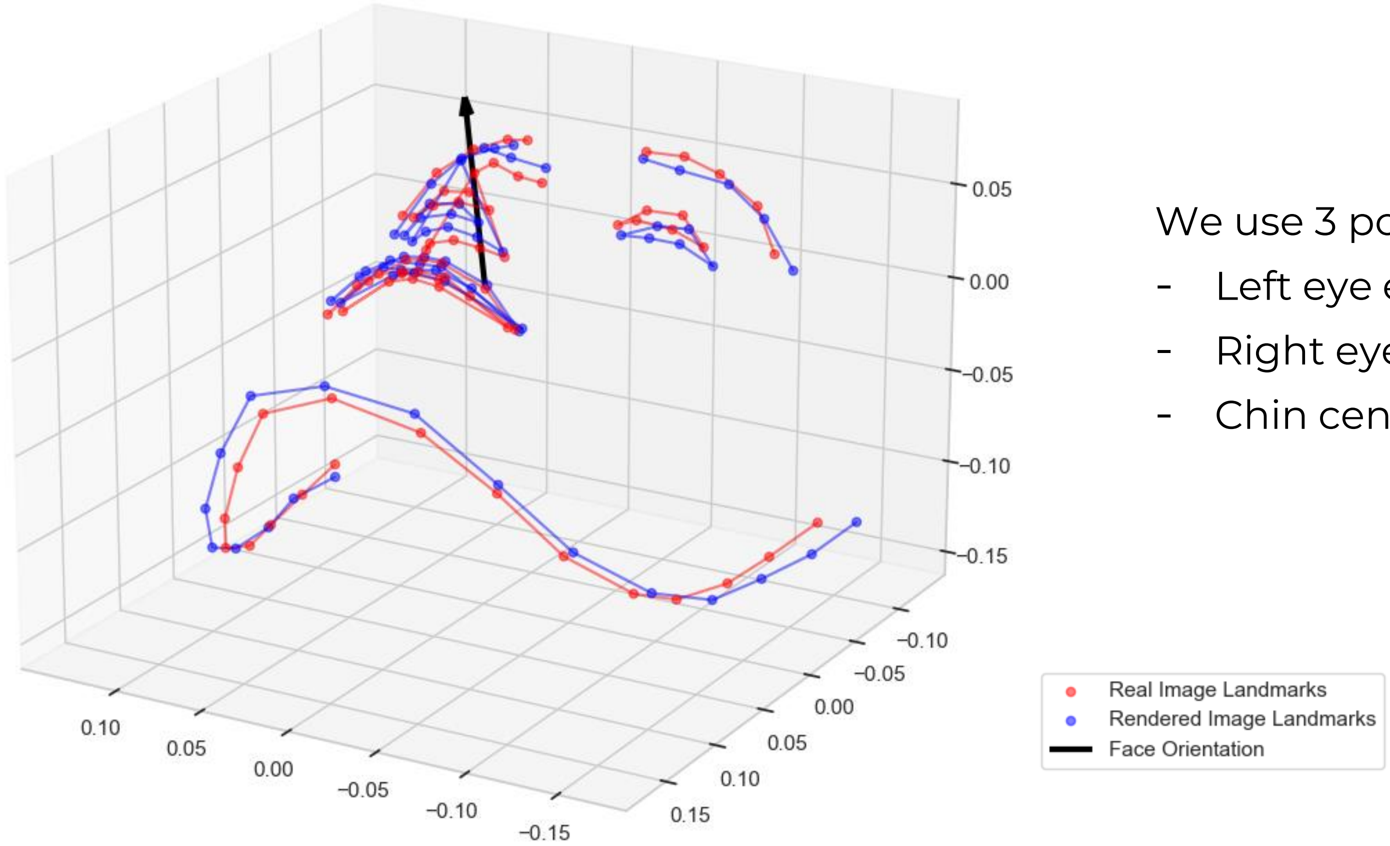
For Gender using photos was more efficient

This experience needs more investigation

(e.g., per class precision)

A lot of questions still remain…

Estimation of the orientation

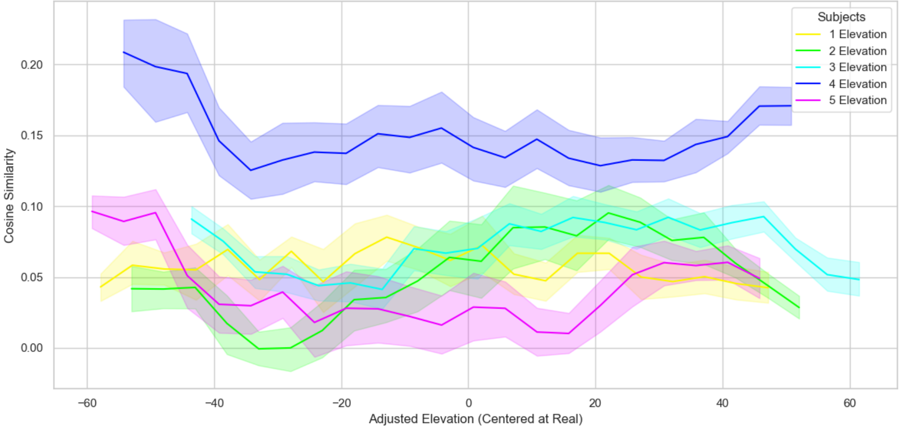


We use 3 points :

- Left eye extremity
- Right eye extremity
- Chin center extremity

Importance of orientation of the reconstruction

ID of Reconstructed MRI: 4



ID of Reconstructed MRI: 1

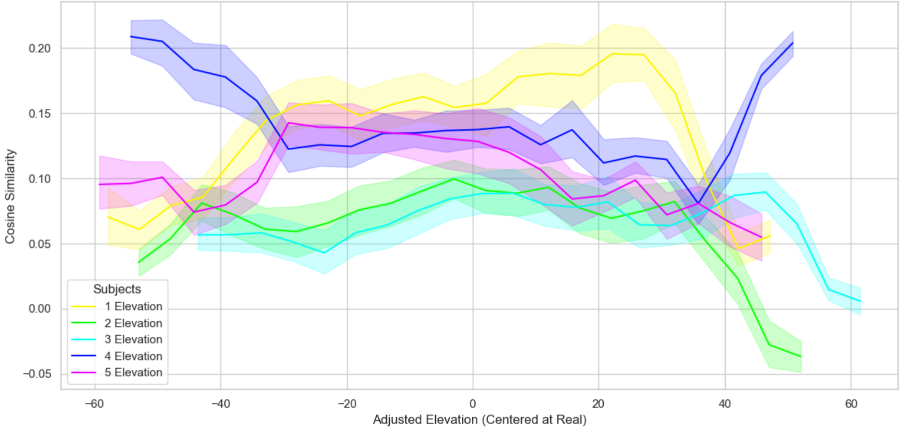


Figure: Similarity depending on the azimuth of the capture.

[GhostFace model, cosine similarity]

In some cases the comparison is stable whatever the choice of azimuth/elevation.

Some other cases: the capture angles have a significant importance

Many other open questions

Better models

Other models than VGG Face (e.g., GhostFaceNet)

Abandoning landmarks?

Switching to 3D landmarks.

More freedom == more errors?

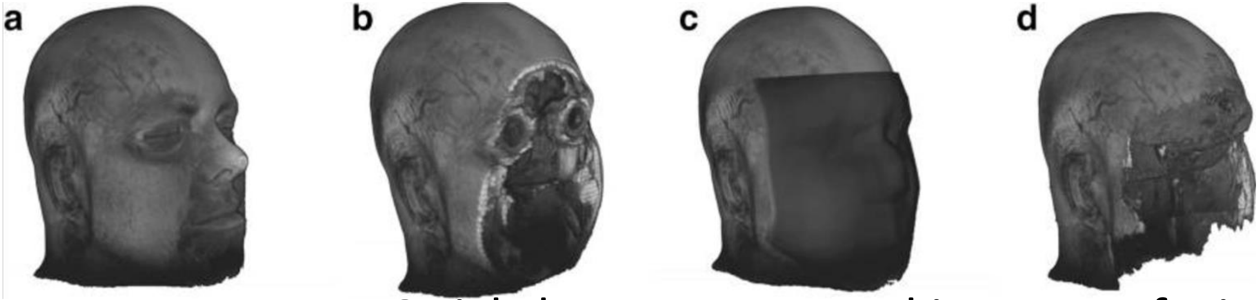
Bridging the gap even more between reconstruction and photos

Black & white reconstructions compared to color photographs

Conclusion

Conclusion

- **What have we discussed ?**
  - We advocate for an evaluation protocol based on verification
  - Designed attacks to highlight the vulnerability of sharing MRI data
  - Illustration of the impact of hair removal in MRI re-identification
  - Many room for improvement
- **Main goal still in sight**
  - Evaluation of the attacks on defacing techniques



Quickshear
Schimke, 2011

FaceMasking
Milchenko, 2013

Defacing
Bischoff-Grethe, 2007

Thank you

Questions ?