Video-based Morphing Attack Detection

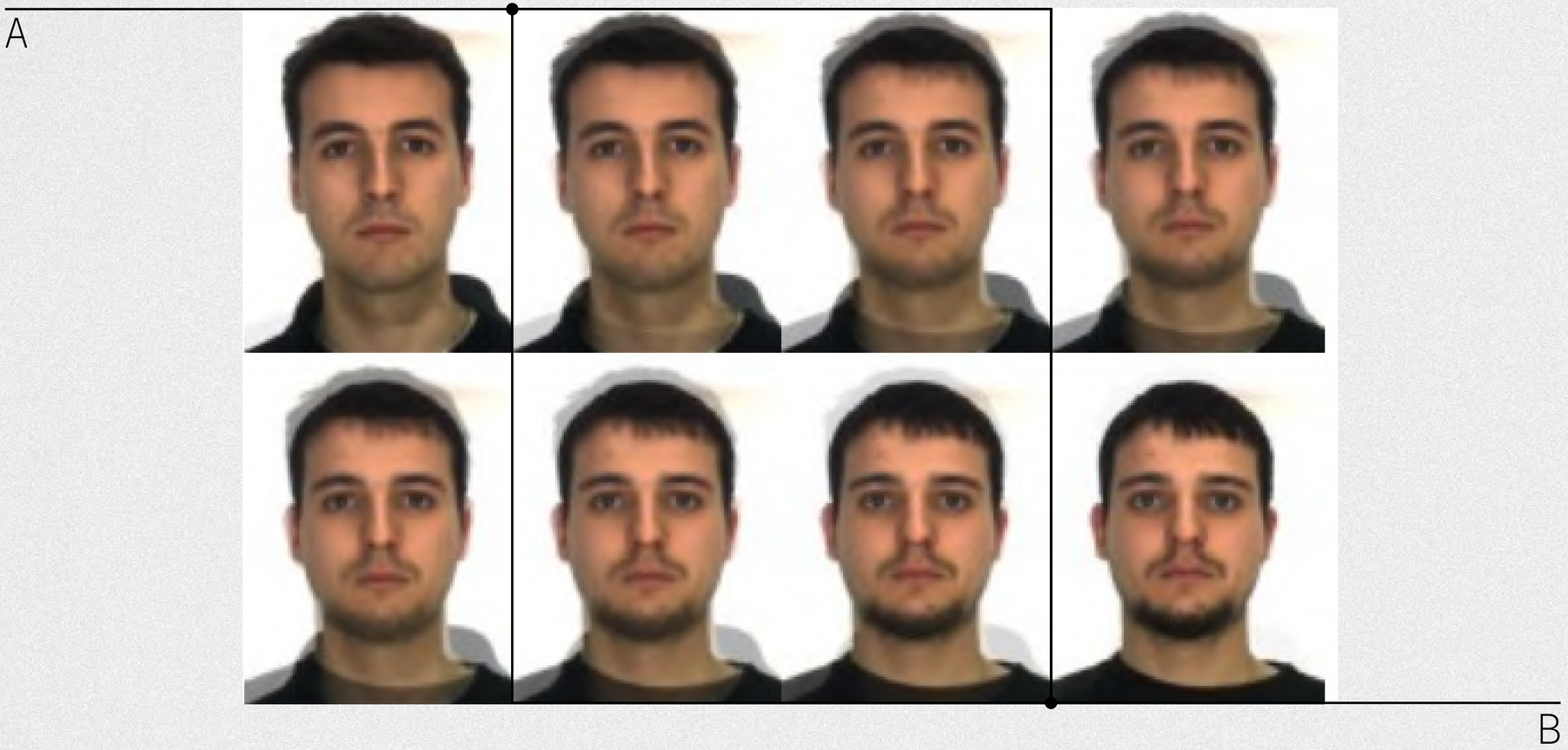
Candidato: Fabio Notaro

Relatrice: Annalisa Franco
Co-relatori: Guido Borghi
Matteo Ferrara

Tesi di Laurea Magistrale in Ingegneria
e Scienze Informatiche
17/07/2025

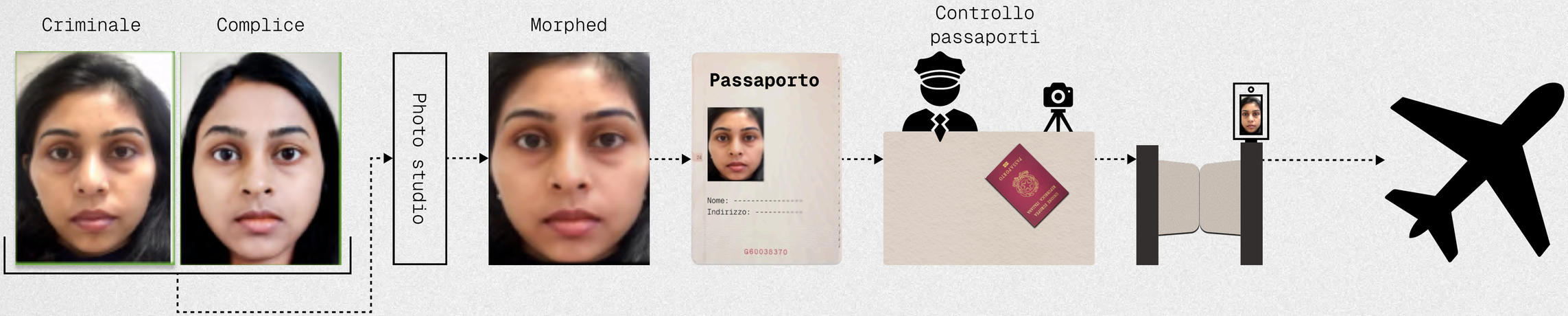
Face Morphing

Tecnica di computer graphics che combina due immagini di volti umani per creare una transizione graduale da un soggetto all’altro. Ciascun frame generato contiene caratteristiche del volto di entrambi i soggetti, in proporzioni diverse.



Problemi

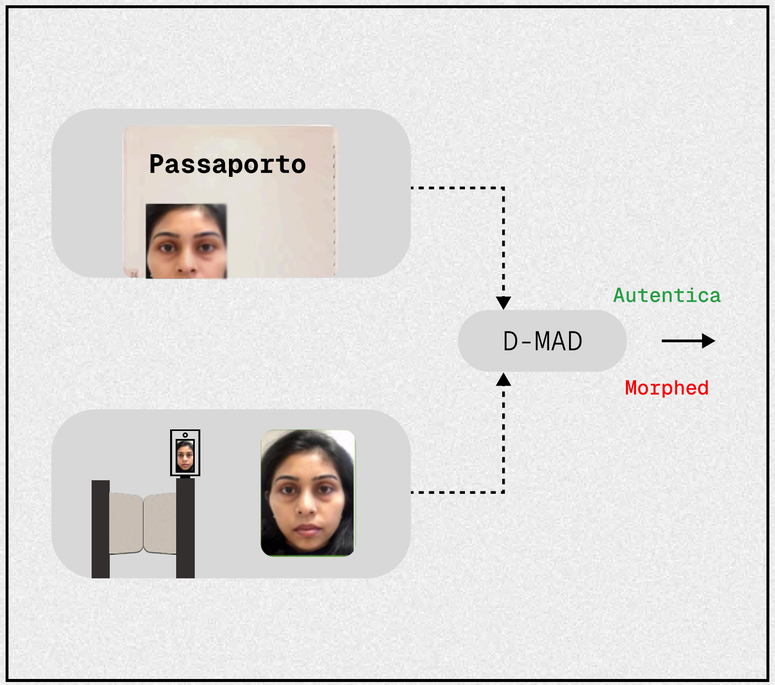
- Elusione dei sistemi di riconoscimento facciale
- Violazione del principio di univocità tra documento e proprietario
- Falsificazione e furto di identità.



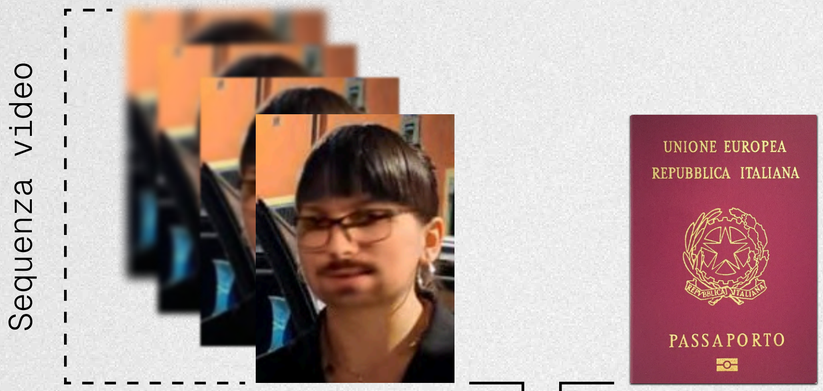
Morphing detection tradizionale

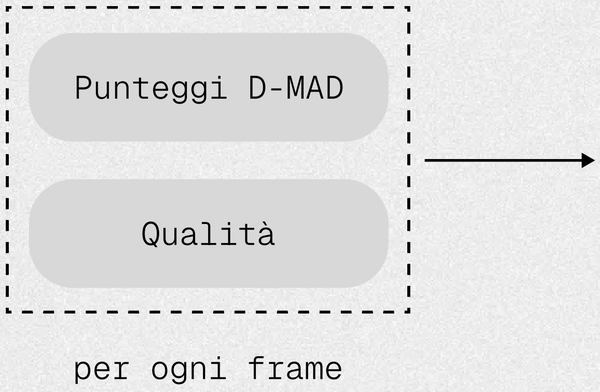
Single-image based MAD

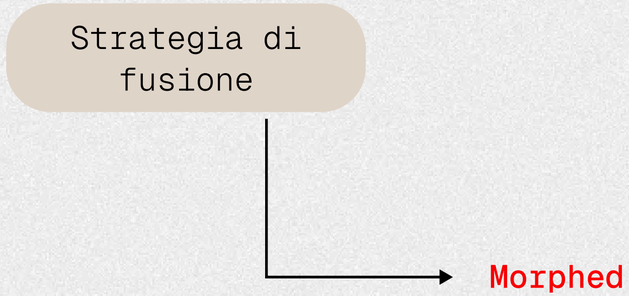
Differential-image based MAD



Video-based MAD







Punti critici:

- comprensione di se e quanto la qualità impatta i punteggi D-MAD
- individuazione di adeguate strategie di fusione
- verifica dell'impatto di condizioni di acquisizione subottime → distanza, scarsa illuminazione e occlusioni.

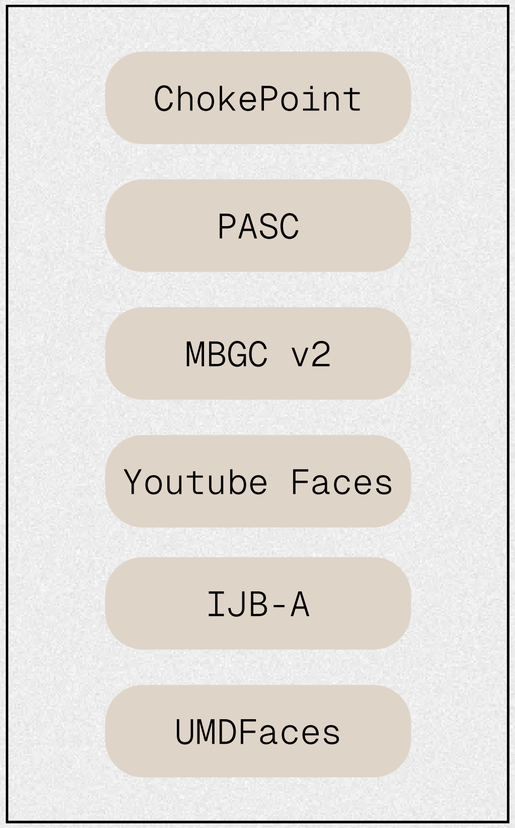
Lavoro di ricerca

- Collezione di un dataset interno all'Università con particolare focus su qualità, variabilità ed eterogeneità → successiva generazione di immagini morphed dei soggetti presenti
- Esperimenti di riconoscimento facciale
- Esperimenti di V-MAD
- Valutazione di metriche specifiche in modo da stimare ed analizzare i risultati ottenuti.

Caratteristiche desiderabili per V-MAD

- Foto frontale ICAO-compliant per ogni soggetto del dataset
- Multipli video aventi condizioni differenti per ogni soggetto del dataset
- Etichettatura chiara e coerente
- Accessibilità e riusabilità
- Alta qualità e risoluzione delle immagini.

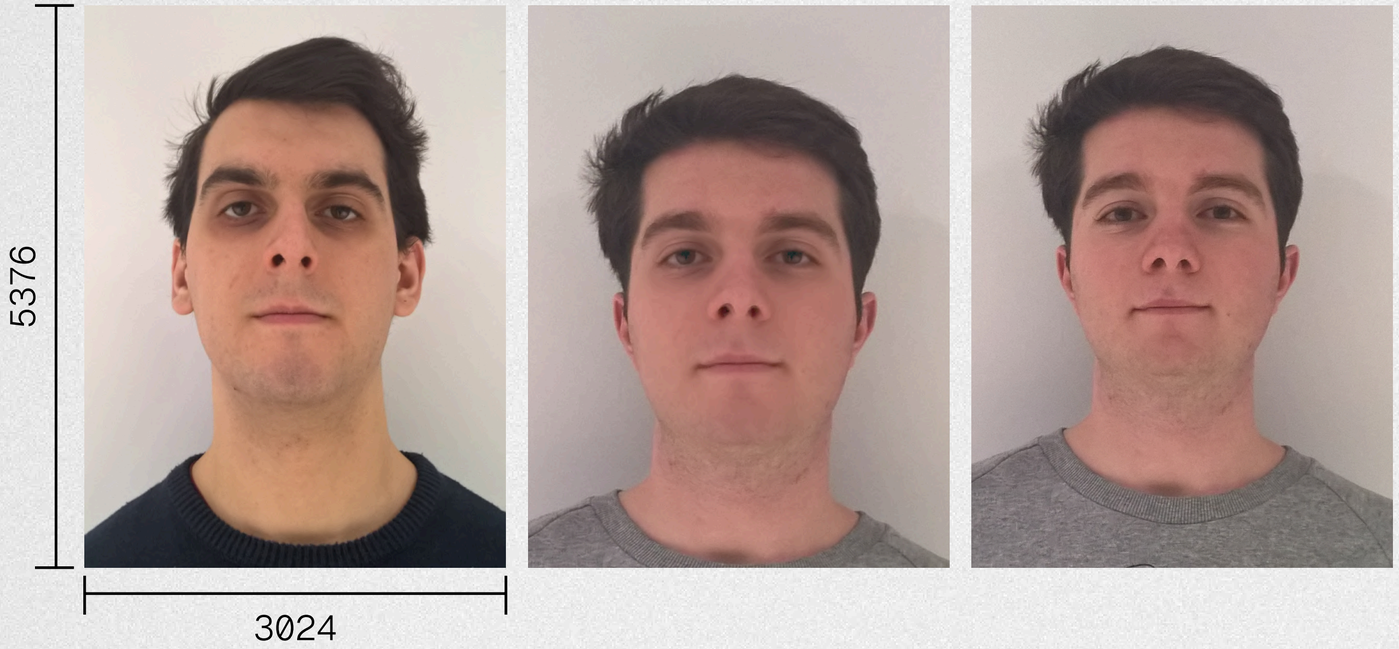
Dataset studiati:



Panoramica dataset GazeWay

65 soggetti

325 immagini morphed

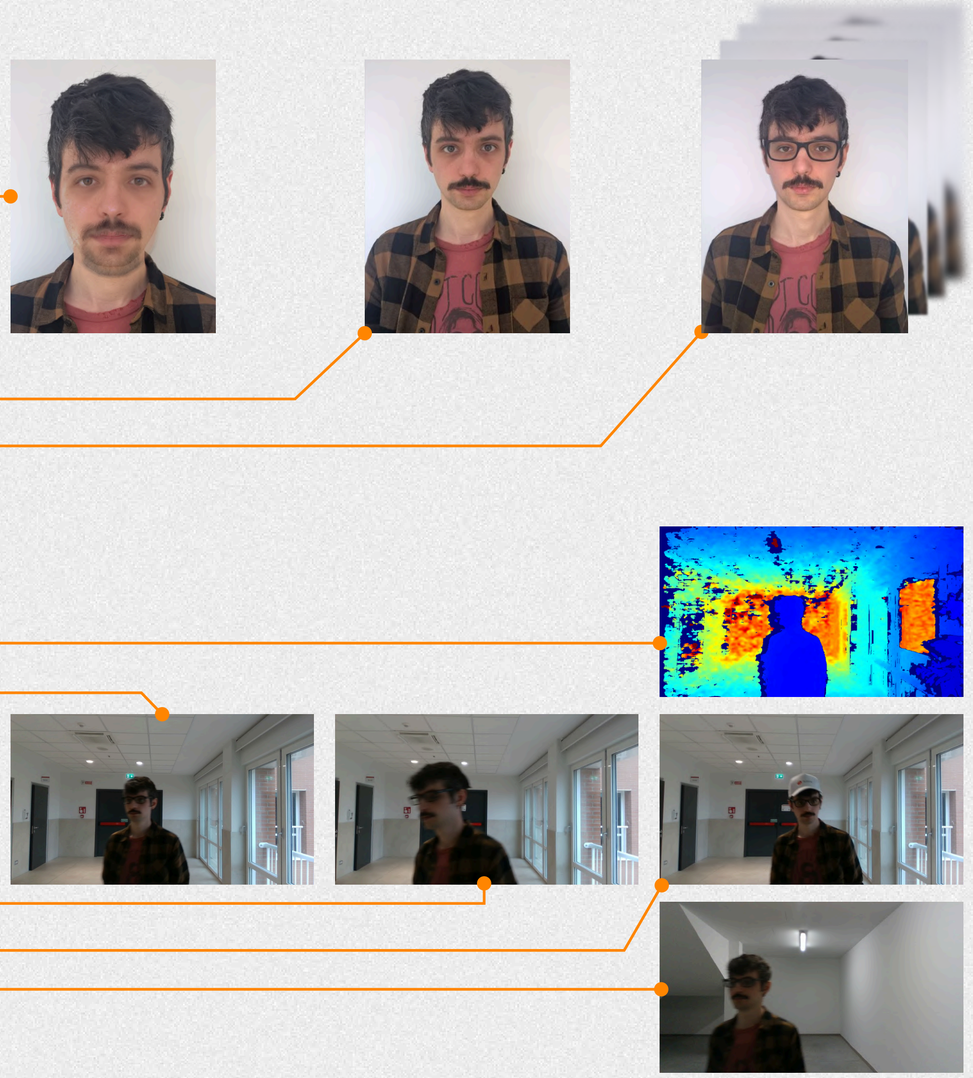


- Foto frontali, sia ICAO-compliant che non, aventi risoluzione 3024x5376 → la ICAO compliance è stata verificata sfruttando strumenti quali BioLab-ICAO-Check Tool, icaonet e BioGaze
- Per ogni soggetto, acquisizione di 6 sequenze video a 30 FPS, decomposte in frame 1024x720, in due differenti ambienti e assumendo tre pose distinte
- Frame video sia in formato RGB che depth
- In totale oltre 331000 immagini.

Struttura dataset GazeWay

Command Prompt

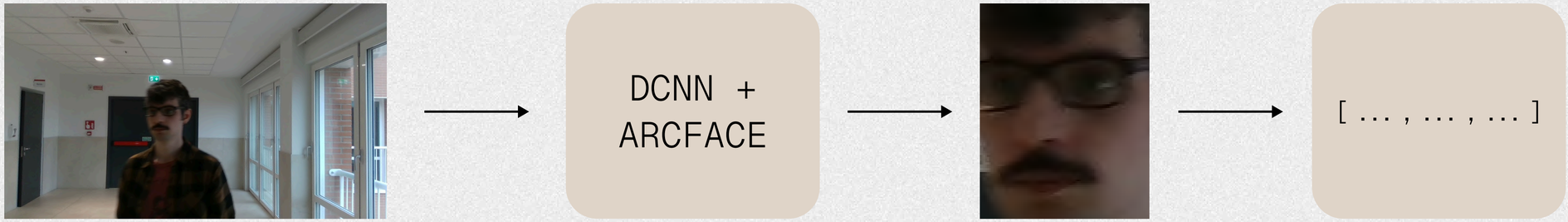
```
GazeWay/

├── morphed
├── ID001
├── ...
├── ID065
    ├── ICAO_photo
    ├── non_ICAO_frontal_photos
    └── video_sequences
        ├── sequence_01
            ├── frontal_gaze
                ├── depth
                └── rgb
            ├── looking_around
            └── looking_around_with_occlusion
        └── sequence_02
            ├── frontal_gaze
            └── ...
```



Face detection ed estrazione embedding

È stato utilizzato ArcFace per rilevare il volto da ogni immagine del dataset e codificarlo tramite un embedding. In particolare ArcFace:

- è una funzione di loss
- richiede una DCNN (solitamente ResNet o RetinaFace) per individuare il volto, ritagliarlo e codificarlo tramite vettore 512-dimensionale
- ha l'obiettivo finale di avere bassa distanza intra-classe ed elevata distanza inter-classe.



Riconoscimento facciale

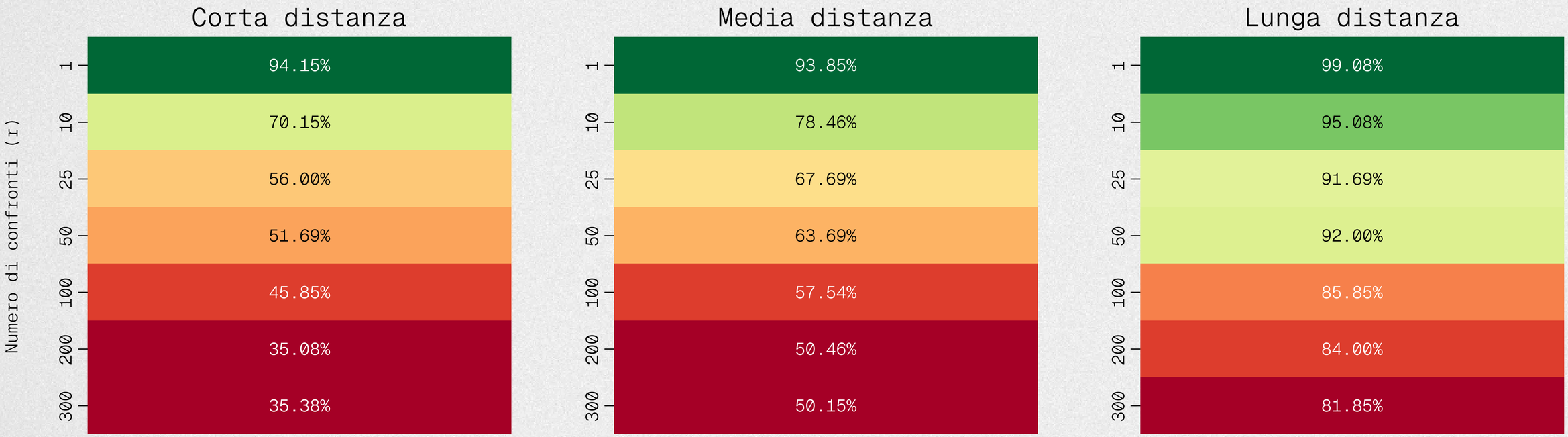
- Obiettivo → per ogni video associazione del soggetto ripreso con relativa foto frontale ICAO-compliant
- Confronti genuine → calcolo distanza coseno tra foto frontale ICAO-compliant di ogni soggetto e tutti i frame video in cui egli compare
- Confronti impostor → calcolo distanza coseno tra la foto frontale ICAO-compliant di ogni soggetto e tutti i frame video in cui compaiono i 6 soggetti a egli più simili
- Metriche analizzate → EER e curve DET, anche in funzione della distanza → i risultati ottenuti mostrano elevata robustezza ed errori prossimi allo 0.

V-MAD

- Obiettivo → verificare le prestazioni sul dataset GazeWay di un classificatore SVM con kernel RBF preaddestrato sul D-MAD
- Confronti genuine → punteggio restituito dal classificatore tra la foto frontale ICAO-compliant di ogni soggetto e tutti i frame video in cui egli compare
- Confronti impostor → punteggio restituito dal classificatore tra ciascuna foto morphed e tutti i frame video dei due soggetti che hanno contribuito a crearla
- Metriche analizzate → matrici MAP, EER e curve DET, anche in funzione della distanza.

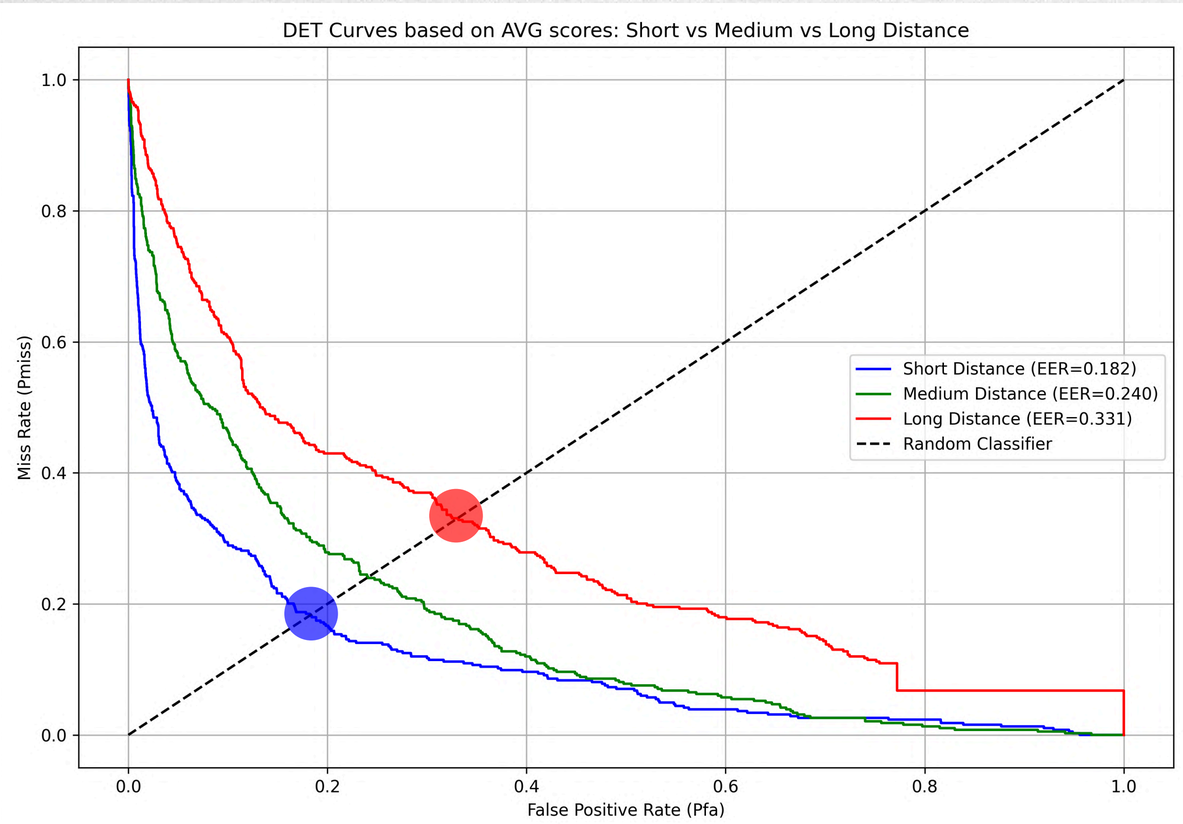
Matrice Morphing Attack Potential

Essa è definita come una matrice dove il generico elemento MAP[r,c] riporta la percentuale di immagini morphed del dataset che riescono a venire confuse con entrambi i soggetti creatori considerando almeno r confronti con coppie di frame dei due soggetti originali.
Le MAP sotto, espresse in funzione della distanza, mostrano chiaramente prestazioni decisamente migliori su distanze medio-corte:

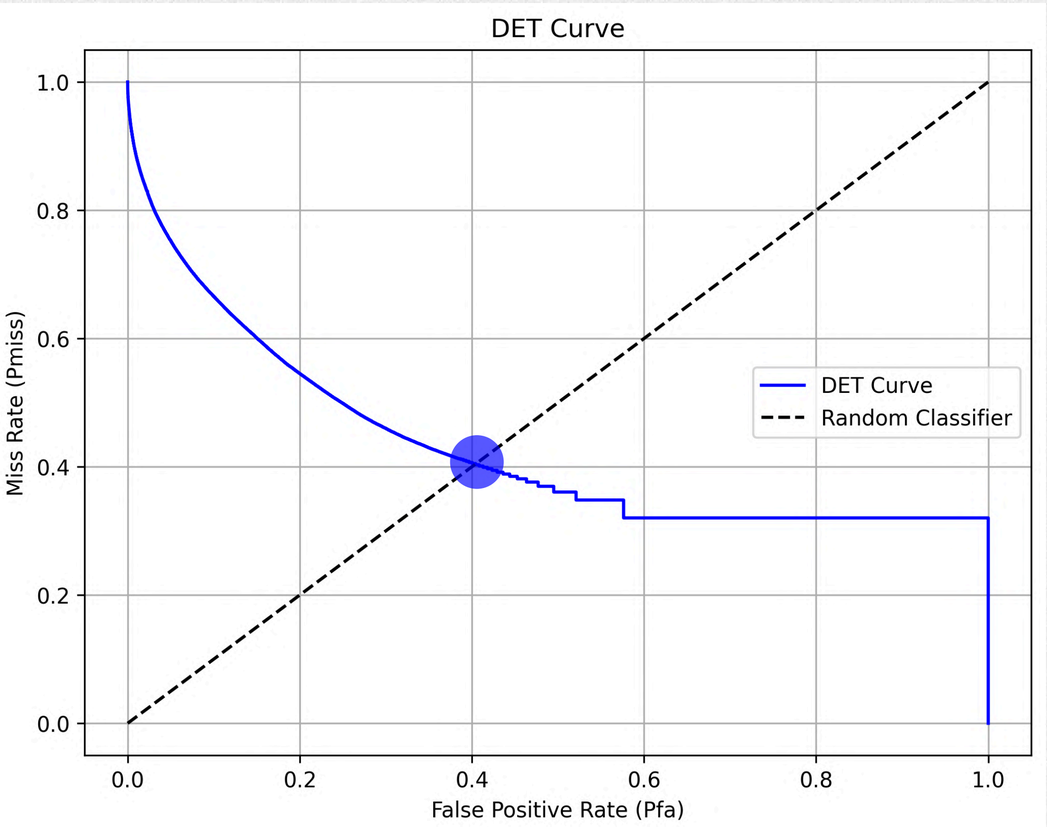


Risultati V-MAD in funzione della distanza

Curve DET considerando solo i punteggi ottenuti da frame in cui il soggetto è stato rilevato a corta distanza (entro 2.5 metri), media distanza (tra 2.5 e 5 metri) e lunga distanza (oltre 5 metri):



Curva DET prodotta senza nessun metodo di aggregazione, che mostra infatti un tasso d'errore molto più elevato rispetto a prima, evidenziando di riflesso l'importanza di adottare tecniche di aggregazione anche semplici:



Conclusioni

Il lavoro di tesi ha permesso di:

- collezionare un dataset adatto a V-MAD
- dimostrare che le migliori prestazioni nel riconoscimento facciale e nel V-MAD si ottengono quando il volto viene rilevato entro 5 metri dalla camera
- dimostrare le potenzialità del V-MAD
- essere tra i primi a condurre esperimenti sul V-MAD, aprendo di fatto nuove prospettive di ricerca nell'ambito della sicurezza biometrica.

Grazie per l’attenzione