# Vulnerability of Face Recognition to Deep Morphing

## Morphing with DeepFakes

[redacted]

Oct 10th 2019

[redacted]

Idiap Research Institute

# Outline

Morphing and Deepfakes

Database

Vulnerability assessment

Detection of Deep Morphs

Conclusion and Future Works

Thank You

# Outline

Morphing and Deepfakes

Database

Vulnerability assessment

Detection of Deep Morphs

Conclusion and Future Works

Thank You

# Morphing attacks

- A blend of two faces with features resembling both people

# Morphing attacks



## Latest developments


- M. Ferrara, A. Franco, and D. Maltoni. *The magic passport*, BTAS (2014).
- U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch. *Face recognition systems under morphing attacks: A survey*, IEEE Access (2019).
- A. Makrushin, T. Neubert, and J. Dittmann. *Automatic generation and detection of visually faultless facial morphs*, VISIGRAPP (2017), SciTePress.
- C. Seibold, W. Samek, A. Hilsmann, and P. Eisert. *Detection of face morphing attacks by deep learning*, IWDW (2017).
- R. S. S. Kramer, M. O. Mireku, T. R. Flack, and K. L. Ritchie. *Face morphing attacks: Investigating detection with humans and computers*, Cognitive Research: Principles and Implications, 2019.

# Deepfakes

- A GAN model is trained from images of two people
- Trained model can convert faces of one person to another

# Deepfake images/videos



## Deep learning and fake

- Started in social media to swap faces of celebrities in videos
- They are relatively easy to generate
- The aim is to generate realistic 'digital copy' of a person
- Deepfakes can mimic facial expressions and movements

# Deepfake images/videos



## Latest developments

- I. Korshunova, W. Shi, J. Dambre, and L. Theis. *Fast face-swap using convolutional neural networks*, ICCV (2017).
- Y. Nirkin, I. Masi, A. T. Tuan, T. Hassner, and G. Medioni. *On face segmentation, face swapping, and face perception*, FG (2018).
- H. X. Pham, Y. Wang, and V. Pavlovic. *Generative adversarial talking head: Bringing portraits to life with a weakly supervised neural network*, arXiv (2018).

# Deepfakes are the new morphs

Steve Buscemi to Jennifer Lawrence

Nicolas Cage to Amy Adams

- Generated Deepfake is also a blend
- But no control on how much of each person is in the result

# Outline

Morphing and Deepfakes

Database

Vulnerability assessment

Detection of Deep Morphs

Conclusion and Future Works

Thank You

# Database of deep morph videos



## Can generated faces fool algorithms?

- Selected 16 pairs of people in VidTIMIT database
- Swapped faces within each pair
- 320 low quality tampered videos with 64x64 GAN model
- 320 higher quality videos with 128x128 model

Download database:
https://www.idiap.ch/dataset/deepfaketimit

# Low and higher quality deep morphs

64x64 Model

128x128 Model

# Example 1, higher quality model

Original Donor | Original Target | Generated face

# Video example 1, higher quality model



Original Donor

Original Target

Generated face

# Example 2, higher quality model

Original Donor | Original Target | Generated face

# Video example 2, higher quality model



Original Donor

Original Target

Generated face

# Outline

Morphing and Deepfakes

Database

Vulnerability assessment

Detection of Deep Morphs

Conclusion and Future Works

Thank You

# Is face recognition vulnerable to genera

VGG-based Face Recognition

FaceNet-based Face Recognition

- 85% to 95% of Deepfakes fool face recognition

# Outline

Morphing and Deepfakes

Database

Vulnerability assessment

Detection of Deep Morphs

Conclusion and Future Works

Thank You

# Detection of deep morph videos



## General approach

- Split database into Train and Test sets
- Apply techniques used in presentation attack detection

## Baselines

- Apply PCA+LDA on pixels directly
- 139 Image quality measures (IQM) as features and PCA+LDA as classifier
- IQM features with SVM classifier

# Baseline systems for low and high quality

| Database | Detection system | EER (%) | FRR@FAR10% (%) |
|---|---|---|---|
| LQ deep morph | Pixels+PCA+LDA | 39.48 | 78.10 |
| | IQM+PCA+LDA | 20.52 | 66.67 |
| | IQM+SVM | **3.33** | **0.95** |
| HQ deep morph | IQM+SVM | **8.97** | **9.05** |

- Higher quality deep morphs are harder to detect!

# Demo of assessment and detection

- Face recognition is almost always fooled
- Detection is wrong for half of the frames

# Outline

Morphing and Deepfakes

Database

Vulnerability assessment

Detection of Deep Morphs

Conclusion and Future Works

Thank You

# Conclusions and future plans



## Conclusions

- Deepfakes are promising as new morphs
- They fool face recognition
- The higher quality, the harder to detect

## Future work

- Can we control the balance of two faces in deep morphs?
- Can deep morphs spoof humans?
- Need detection methods independent of generator models

# Outline

Morphing and Deepfakes

Database

Vulnerability assessment

Detection of Deep Morphs

Conclusion and Future Works

Thank You

# Thank you for your attention

## Idiap Research Institute (Switzerland)

- **Biometrics Security and Privacy group**
  Reproducible Research in:
  - Face (2D/3D, multi-spectral, heterogeneous), Speaker, Vein and multi-modal fusion
  - Soft-biometrics: gender, age and heart-beat estimation (**rPPG**)
  - Adversarial attacks (e.g. Presentation Attacks, Morphing, Deepfakes, . . . )
  - Attack detection (PAD)
- **FIDO Accredited Biometric Laboratory in Europe**

fido™ CERTIFIED

## Contact

[redacted]

# Thank you for your attention!

[redacted]

Idiap Research Institute, Martigny, Switzerland