Virtual U: Defeating Face Liveness Detection by Building Virtual Models From Your Public Photos

Yi Xu, True Price, Jan-Michael Frahm, and Fabian Monrose

Department of Computer Science, University of North Carolina at Chapel Hill

USENIX Security
August 11, 2016
Face Authentication: Convenient Security
Evolution of Adversarial Models

- **Attack:** Still-image Spoofing
Evolution of Adversarial Models

- **Attack:** Still-image Spoofing
- **Defense:** Liveness Detection
Evolution of Adversarial Models

- **Attack**: Still-image Spoofing
- **Defense**: Liveness Detection
- **Attack**: Video Spoofing
Evolution of Adversarial Models

- **Attack**: Still-image Spoofing
- **Defense**: Liveness Detection
- **Attack**: Video Spoofing
- **Defense**: Motion Consistency
Evolution of Adversarial Models

- **Attack**: Still-image Spoofing
- **Defense**: Liveness Detection
- **Attack**: Video Spoofing
- **Defense**: Motion Consistency
- **Attack**: 3D-Printed Masks
Virtual U: A New Attack

We introduce a new **VR-based attack** on face authentication systems solely **using publicly available photos** of the victim.
Virtual U: A New Attack

1. Input Web Photos
2. Landmark Extraction
3. 3D Model Reconstruction
4. Image-based Texturing
5. Gaze Correction
6. Expression Animation

Viewing with Virtual Reality System
Leveraging Social Media
Landmark Extraction
3D Face Model

Identity Variation (e.g., thin-to-heavyset)

Expression Variation (e.g., frowning-to-smiling)
3D Face Model

Expression Variation (e.g., frowning-to-smiling)

Identity Variation (e.g., thin-to-heavyset)

\[ S = \bar{S} + A^{id}\alpha^{id} + A^{exp}\alpha^{exp} \]
3D Face Model

\[ S = \bar{S} + A^{id} \alpha^{id} + A^{exp} \alpha^{exp} \]
3D Face Model

\[ S = \bar{S} + A_{id}^{\text{id}} + A_{exp}^{\text{exp}} \]
$S = \bar{S} + A_{id} \alpha_{id} + A_{exp} \alpha_{exp}$

3D Face Model
3D Face Model

\[ S = \bar{S} + A^{\text{id}} \alpha^{\text{id}} + A^{\text{exp}} \alpha^{\text{exp}} \]
3D Face Model
3D Face Model
3D Face Model
3D Face Model

Pose $\alpha^{exp}$

Pose $\alpha^{exp}$

Pose $\alpha^{exp}$

Pose $\alpha^{exp}$

$\alpha^{id}$
Multi-Image Modeling

Single image

Multiple images
Texturing

Direct Texturing 2D Poisson Editing
Texturing

Direct Texturing

2D Poisson Editing

3D Poisson Editing
Gaze Correction
Gaze Correction
Virtual U: A New Attack

1. Input Web Photos
2. Landmark Extraction
3. 3D Model Reconstruction
4. Image-based Texturing
5. Gaze Correction
6. Viewing with Virtual Reality System

Expression Animation
Expression Animation

\[ S = \bar{S} + A^{id} \alpha^{id} + A^{exp} \alpha^{exp} \]
VR Display

Printed Marker  VR System  Authentication Device
VR Display
Experiments

- Interaction-based liveness detection
- Motion-based liveness detection
- Texture-based liveness detection

KeyLemon
Mobius
TrueKey
BioID
1U
Experiments

- 20 participants
  - Aged 24 to 44
  - 14 males, 6 females
  - Various ethnicities

- Two tests
  - Indoor photo of the subject in the same environment as registration
  - Publicly accessible photos
    - Anywhere from 3 to 27 photos per person
    - Low-, medium-, and high-quality
    - Potentially strong changes in appearance over time
## Experiments

<table>
<thead>
<tr>
<th>Service</th>
<th>Indoor Image (Single frontal image)</th>
<th>Online</th>
<th>Avg. # Tries</th>
</tr>
</thead>
<tbody>
<tr>
<td>KeyLemon</td>
<td>100%</td>
<td>85%</td>
<td>1.6</td>
</tr>
<tr>
<td>Mobius</td>
<td>100%</td>
<td>80%</td>
<td>1.5</td>
</tr>
<tr>
<td>TrueKey</td>
<td>100%</td>
<td>70%</td>
<td>1.3</td>
</tr>
<tr>
<td>BioID</td>
<td>100%</td>
<td>55%</td>
<td>1.7</td>
</tr>
<tr>
<td>1U</td>
<td>100%</td>
<td>0%</td>
<td>--</td>
</tr>
</tbody>
</table>
Observations

- Medium- and high-resolution photos work best
  - Photos from professional photographers (weddings, etc.)

- Group photos provide consistent frontal views
  - Often lower resolution

- Only a small number of photos required
  - One or two forward-facing photos
  - One or two higher-resolution photos
Experiments

How does resolution affect reconstruction quality?
Experiments

How does rotation affect reconstruction quality?
Experiments

Combining high-res rotation with low-res front-facing?

![Image of a person looking at the camera with a group of people in the background.]

+ 

![Image of a person looking at a device with a face highlighted.]

![Graph showing spoofing success rate vs. yaw angle with bars for KeyLemon, Bold, and TrueKey.]

Yaw Angle (Deg)

Spoofing Success Rate (%)
Experiments

- Virtual U is successful against liveness detection
Experiments

- Virtual U is successful against liveness detection

- Also successful against motion consistency
Experiments

- “Seeing Your Face is Not Enough: An Inertial Sensor-Based Liveness Detection for Face Authentication” (Li et al., ACM CCS’15)
  - Device motion measured by inertial sensor data
  - Head pose estimated from input video
  - Train a classifier to identify real data (correlated signals) versus spoofed video data
### Experiments

<table>
<thead>
<tr>
<th>Training Data (Pos. Data vs. Neg. Data)</th>
<th>Test Result (Accept Rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Real Face</td>
</tr>
<tr>
<td>Real vs. Video</td>
<td>98.0%</td>
</tr>
</tbody>
</table>
## Experiments

<table>
<thead>
<tr>
<th>Training Data (Pos. Data vs. Neg. Data)</th>
<th>Test Result (Accept Rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Real Face</td>
</tr>
<tr>
<td>Real vs. Video</td>
<td>98.0%</td>
</tr>
<tr>
<td>Real vs. Video +VR</td>
<td>67.0%</td>
</tr>
</tbody>
</table>
## Experiments

<table>
<thead>
<tr>
<th>Training Data (Pos. Data vs. Neg. Data)</th>
<th>Test Result (Accept Rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Real Face</td>
</tr>
<tr>
<td>Real vs. Video</td>
<td>98.0%</td>
</tr>
<tr>
<td>Real vs. Video +VR</td>
<td>67.0%</td>
</tr>
<tr>
<td>Real vs. VR</td>
<td>67.0%</td>
</tr>
</tbody>
</table>
Mitigations

- Alternative/additional hardware
  - Infrared imaging (e.g. Windows Hello)
  - Random structured light projection
Mitigations

- Alternative/additional hardware
  - Infrared imaging (e.g. Windows Hello)
  - Random structured light projection

- Improved defense against low-resolution synthetic textures
Conclusion

- We introduce a new VR-based attack on face authentication systems solely using publicly available photos of the victim.

- This attack bypasses existing defenses of liveness detection and motion consistency.

- At a minimum, face authentication software must improve against VR-based attacks with low-resolution textures.

- The increasing ubiquity of VR will continue to challenge computer-vision-based authentication systems.
Thank you!

Questions?
Overview

- Face Authentication
- Virtual U: A VR-based attack
- Evaluation
- Mitigations
- Conclusion
Evolution of Adversarial Models

- **Attack:** Still-image Spoofing
- **Defense:** Liveness Detection
- **Attack:** Video Spoofing
- **Defense:** Motion Consistency
- **Attack:** 3D-Printed Masks
- **Defense:** Texture Detection
3D Face Model

Identity Variation (e.g., thin-to-heavyset)

Expression Variation (e.g., frowning-to-smiling)
3D Face Model

\[ S = \bar{S} + A^{id}\alpha^{id} + A^{exp}\alpha^{exp} \]

Identity Variation (e.g., thin-to-heavyset)

Expression Variation (e.g., frowning-to-smiling)
3D Face Model

\[ S = \bar{S} + A^{id}_{\alpha^{id}} + A^{exp}_{\alpha^{exp}} \]

Reprojection

\[
\min_{p, \alpha^{id}, \alpha^{exp}} \sum_i \left\| s_i - PS_i \right\|^2 + \beta_{id} \left\| \alpha^{id} \right\|^2 + \beta_{exp} \left\| \alpha^{exp} \right\|^2
\]

Summed over all landmarks

Normalization

Pose
3D Face Model
Multi-Image Modeling

Single Image

\[
\min_{P, \alpha^{id}, \alpha^{exp}} \sum_i \|s_i - PS_i\|^2 + \beta_{id} \|\alpha^{id}\|^2 + \beta_{exp} \|\alpha^{exp}\|^2
\]

Multiple Images

\[
\min_{P, \alpha^{id}, \alpha^{exp}} \sum_m \sum_i \|s_{mi} - P_m S_{mi}\|^2 + \beta_{id} \|\alpha^{id}\|^2 + \beta_{exp} \sum_m \|\alpha^{exp}_m\|^2
\]

Sum over all images
Multi-Image Modeling

Corners of the eyes and mouth are **stable** landmarks

Contour points are **variable** landmarks
Multi-Image Modeling

Multiple Images

$$\min_{P, \alpha^{id}, \alpha^{exp}} \sum_m \sum_i \|s_{mi} - P_m s_{mi}\|^2 + \text{norm.}$$

Multiple Images with Landmark Weighting

$$\min_{P, \alpha^{id}, \alpha^{exp}} \sum_m \sum_i \frac{1}{(\sigma_i^s)^2} \|s_{mi} - P_m s_{mi}\|^2 + \text{norm.}$$

Higher weighting for stable landmarks
Experiments

- 20 participants
  - Aged 24 to 44
  - 14 males, 6 females
  - Various ethnicities

- Two tests
  - Indoor photo of the subject in the same environment as registration
  - Publicly accessible photos
    - Anywhere from 3 to 27 photos per person
    - Low-, medium-, and high-quality
    - Potentially strong changes in appearance over time
Experiments

How does rotation affect reconstruction quality?
Experiments

VR System

Google Cardboard

Authentication Device