Evaluation of Presentation Attack Detection: An Example

Peter Johnson and Stephanie Schuckers
Clarkson University
**Presentation Attacks**

- Spoofing is common term used most in past decade.
- ISO Standards underway:
  - **Presentation Attack** Definition: Presentation of an artefact or human characteristic to the biometric capture subsystem in a fashion that could interfere with the intended policy of the biometric system*

- **Why?**
  Posing as another individual
  - Positive ID applications
  Hiding your identity
  - Negative ID applications
  - May form ‘new’ identity for positive ID

*from: ISO/IEC CD 30107-1, Information Technology — Biometrics -- Presentation Attack Detection
Fingerprint Presentation Attacks

- **Cooperative**
  Characteristic captured directly from individual with assistance (e.g. finger mold)

- **Latent**
  Characteristic captured indirectly through lifting a latent sample

- **Synthetic**
  Synthetic characteristic, not mapped to real person (e.g. synthetic fingerprint)

Feng and Jain, Advances in Biometrics article, 2011 [1].

Presentation Attack Testing on Conventional Systems

- Matsumoto et al., 2002 [3]
  Testing acceptance rate of gelatin and silicone fingers (in terms of matching)

- Thalheim et al., 2002 [4]
  Tested various techniques for spoofing biometric systems
  Reactivating latent print and fingerprint on adhesive film

- Galbally et al., 2010 [5]
  Optical and thermal sweeping sensors shown to be vulnerable to direct (presentation) attacks

Presentation Attack Detection (PAD)

- Presentation Attack Detection (PAD) *
  Automated determination of a presentation attack
- Examples of PAD
  Liveness detection (failure)
  Artefact detection
  Altered biometric detection
  Others terms that have been used: anti-spoofing, biometric fraud, spoof detection, authenticity detection, etc.

*from: ISO/IEC CD 30107-1, Information Technology — Biometrics -- Presentation Attack Detection
Challenge

- Presentation Attack Detection is a component of biometric system.
- In many applications, a successful presentation attack is a combination of failure of the PAD subsystem and matching a stored biometric.
- Previous research on fusion of PAD subsystem and matcher [7]
- Need for common understanding of metrics which measure the fusion of PAD and match scores
Objective

• Give an example of performance results for
  -PAD alone
  -Fusion of PAD and match scores
• Provide dataset of PAD scores and match scores for use in additional research
Traditional Metrics for Biometric Evaluation (Live Finger Input)

Data Capture Subsystem
- Live Finger Presentation
  - Biometric Characteristics
    - Biometric Capture Sensor
      - Reject
        - Failure To Acquire

Presentation Attack Detection Subsystem
- Liveness Detection Module

Signal Processing Subsystem
- Quality Check
  - Segmentation Feature Extraction
    - Reference Creation
      - Reference

Comparison Subsystem
- Comparison
  - Comparison Score
    - Decision Subsystem
      - Match?
        - Decision (Reject/Accept)
          - False Reject
          - False Accept

Decision Subsystem
- Reject
  - Data Storage Subsystem
    - Enrollment Database
      - Reference

Failure To Enroll
- Probe
  - Reference
Additional Metrics (Spoof Input)

- Liveness detection methods treated as two class problem
- Evaluation in literature focuses specifically on liveness detection module only

Accept/Reject

Attack Presentation Classification Rate

Normal Presentation Classification Rate
Additional Metrics (Spoof Input)

- Liveness detection module will contribute to decision to reject
- Other modules (e.g. quality) may contribute
- During testing specific reason for rejection may not be known
- Need clarification in terminology for system testing (this slide) and liveness detection module testing (last slide)
What about matching? (Spoof Input)

- Spoof finger may not be rejected by earlier modules
- If spoof matches stored reference, a successful presentation attack has occurred.
Presentation Attack Detection Dataset

• Algorithms are often referred to as liveness detection algorithms

• Dataset includes scores from two PAD algorithms
  Algorithm 1: Intensity analysis of fingerprint image [8]
  Algorithm 2: Combination of multiple algorithms
  • Intensity [8]
  • Valley noise analysis [9]
  • Ridge signal analysis [10]

• A PAD score is determined for the probe image of each pair of fingerprints that is matched
Fingerprint Matching

- Fingerprint matching was conducted using the VeriFinger fingerprint matching SDK [11]

- Genuine match scores:
  Matching of two different fingerprint images from the same subject and same finger
  Every match score was calculated from a pair of fingerprint images that were collected on different days

- Imposter match scores:
  Matching of two different fingerprint images from two different subjects and same finger

- Spoof match scores:
  Matching of two different fingerprint images from the same subject and same finger
  Gallery image is from a live finger and probe image is from a spoof finger
Fingerprint Score Dataset

- A fingerprint dataset consisting of 50 subjects, two fingers each is used for the following analysis
  - The dataset is split into two subsets: 25 subjects for training and 25 subjects for testing
  - 8019 total live images
  - 2705 total spoof images
  - Images collected from right thumb (R1) and right index finger (R2) for each subject
- Dataset is available by request on the CITeR website: http://www.clarkson.edu/citer/research/collections/index.html

<table>
<thead>
<tr>
<th>Subset</th>
<th>Number of Subjects</th>
<th>Number of Live Images</th>
<th>Number of Spoof Images</th>
<th>Normal Presentation—Genuine</th>
<th>Normal Presentation—Imposter</th>
<th>Presentation Attack (Genuine)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>25</td>
<td>R1: 2,187, R2: 1,896</td>
<td>R1: 724, R2: 491</td>
<td>519,198</td>
<td>911,476</td>
<td>106,943</td>
</tr>
<tr>
<td>Testing</td>
<td>24</td>
<td>R1: 2,153, R2: 1,783</td>
<td>R1: 749, R2: 561</td>
<td>381,182</td>
<td>976,161</td>
<td>132,075</td>
</tr>
</tbody>
</table>
Performance Metrics – Matching

• Performance Metrics:

  False match rate (FMR): percentage of fingerprint pairs from different people (imposters) that match

  False non-match rate (FNMR): percentage of fingerprint pairs from the same person/finger (genuine) that do not match

  True match rate (TMR): TMR = 100 – FNMR

• Matching threshold is selected from training set performance and tested on the testing set

  Matching threshold = 30
  FRR = 0.59%
  FAR = 0.003%
Performance Metrics – PAD

- **Performance Metrics:**
  
  **Normal Presentation Classification Rate (NPCR):** percentage of normal presentations (live fingerprints) that are accepted as normal presentations
  
  **Attack Presentation Classification Rate (APCR):** percentage of attack presentations (spoof fingerprints) correctly classified as attack presentations
  
  **Attack presentation error rate (APER):** percentage of attack presentations that are accepted as normal presentations (100 – APCR)
The biometric system combines the Comparison Subsystem (matching) with the Presentation Attack Detection Subsystem (liveness)

The system needs to be able to utilize information passed from both modules to make a single decision (accept or reject)

New error terms must be applied with the addition of Presentation Attack Detection

**Performance Metrics:**

- **False accept rate (FAR):** Percentage of imposters accepted by the system
- **False reject rate (FRR):** Percentage of genuine users rejected by the system
- **True accept rate (TAR):** TAR = 100 – FRR
- **Spoof false accept rate (SFAR):** Percentage of spoof samples that are accepted by the system (i.e. by matching and PAD)
# Decision Matrix & Metrics

## Type of Test

<table>
<thead>
<tr>
<th>Decision</th>
<th>Presentation Attack Genuine</th>
<th>Normal Presentation Genuine</th>
<th>Normal Presentation Imposter</th>
</tr>
</thead>
</table>
| Presentation Attack Match | **FRR*** | | **
| Presentation Attack Non-Match | **FRR*** | | |
| Normal Presentation Non-Match | **FRR*** | | |
| Normal Presentation Match | SFAR | | FAR |

*Incorrectly rejected by PAD OR Matcher
**Correctly rejected but for the wrong reason (PAD)
Joint Distributions of Match and PAD (Liveness) Scores (Liveness Algorithm 2)
Fingerprint System with Presentation Attack Detection (PAD) – Series Implementation

- Fingerprint image is first passed through PAD module
- If rejected by PAD, sample is rejected by entire system

**Data Capture Subsystem**
- Live Finger
- Spoof

**Presentation Attack Detection Subsystem**
- Liveness Detection Module

**Signal Processing Subsystem**
- Quality Check
- Segmentation Feature Extraction
- Reference Creation

**Comparison Subsystem**
- Comparison
  - Comparison Score
- Match?
- Decision (Reject/Accept)

**Decision Subsystem**
- Spoof False Accept
- False Reject (Live)
- False Accept (Live)

**Data Storage Subsystem**
- Enrollment Database
- Biometric Claim
- Reference

**Biometric Characteristics**
- Biometric Capture Sensor
Fingerprint System with Presentation Attack Detection (PAD) – Parallel Implementation

- Fingerprint is passed to signal processing subsystem regardless of PAD output
- Comparison subsystem makes decision based on both scores

Data Capture Subsystem

- Live Finger
- Spoof

Biometric Characteristics

Biometric Capture Sensor

Presentation Attack Detection Subsystem

- Liveness Detection Module

Signal Processing Subsystem

- Quality Check
- Segmentation Feature Extraction

Reference Creation

Decision Subsystem

- Match?
- Decision (Reject/Accept)

Comparison Subsystem

- Comparison
- Comparison Score

Data Storage Subsystem

- Enrollment Database

Biometric Claim

Spoof False Accept
False Reject (Live)
False Accept (Live)
Performance with PAD in Series (Liveness Algorithm 1)

- No spoofing: FRR = FAR = 0.10%
- Liveness fusion: FRR = FAR = SFAR = 11.58%
- No liveness (Threshold = 0): FRR = SFAR = 25.40%
- SFAR = 98.02%
Series System Decision Boundary

Matching Threshold = 43
Liveness Threshold = 552
Parallel Fusion

- **Parallel fusion:**
  Comparison subsystem performs some fusion function $f$ on the match score $S_m$ and liveness score $S_l$.
  Simplest example is the sum rule
  $$f = S_m W_m + S_l W_l$$

- **Weights** are calculated based on individual performance, such that $\sum_i W_i = 1$
  $$W_i = \frac{1 - 2EER_i}{2 - (2EER_i + 2EER_j)}, i \neq j$$

- **Score $S$** is first transformed to normalized score $S_N$ using min-max normalization
  $$S_N = \frac{S - \min(S)}{\max(S) - \min(S)}$$

**Graph**
- FRR = FAR = 9.07%
- SFAR = 11.12%
Sum Rule Fusion Decision Boundary

Fusion Decision Boundary

Density

Genuine
Spoof
Imposter

Liveness Score
Match Score

1000 800 600 400 200 0
1200 1000 800 600 400 200 0

0 1200
1000 800 600 400 200 0

1.5 x 10^{-3}
1.5
1.0
0.5
0
Performance Comparison

Training

• Thresholds are chosen based on the training set

• System 1: No liveness
  Matching Threshold = 30
  FRR = 0.1%
  FAR = 0.1%
  SFAR = 98.02%

• System 2: Liveness in series
  Matching threshold = 43
  Liveness threshold = 552
  FRR = 11.58%
  FAR = 11.58%
  SFAR = 11.58%

• System 3: Liveness in parallel
  Fusion threshold = 0.3083
  FRR = 9.07%
  FAR = 9.07%
  SFAR = 11.12%
Performance Comparison Testing

- Performance of three systems is evaluated on the testing set

- **System 1: No liveness**
  - Matching Threshold = 30
  - FRR = 0.59%
  - FAR = 0.003%
  - SFAR = 98.35%

- **System 2: Liveness in series**
  - Matching threshold = 43
  - Liveness threshold = 552
  - FRR = 3.55%
  - FAR = 0%
  - SFAR = 9.49%

- **System 3: Liveness in parallel**
  - Fusion threshold = 0.3083
  - FRR = 5.75%
  - FAR = 3.33%
  - SFAR = 9.41%
Summary

• Performance metrics for PAD system
  
  Normal Presentation Classification Rate (NPCR): percentage normal presentations that are accepted as normal presentations
  Attack Presentation Classification Rate (APCR): percentage of attack presentations correctly classified as attack presentations

• Performance metrics for combination of PAD subsystem and Comparison subsystem
  
  False accept rate (FAR): Percentage of imposters accepted by the system
  False reject rate (FRR): Percentage of genuine users rejected by the system
  Spoof False Accept Rate (SFAR) -- Percentage of spoof samples that are accepted by the system (i.e. by matching and PAD)

• The training and testing datasets are available by request for download for further experimentation
  
  http://www.clarkson.edu/citer/research/collections/index.html
Summary -con-

- Two distinct implementations of presentation attack detection in a fingerprint recognition system have been examined
  Series: Detecting fingerprint liveness prior to matching and filtering out spoof samples
  Parallel: Detecting fingerprint liveness alongside matching and implementing a fusion function in the comparison subsystem
- The series implementation resulted in a significant reduction in performance regarding live fingers
  FRR dropped from 0.59% to 3.55% on testing set
- The simple sum rule fusion did not improve upon the series result
  Sum rule still provides a linear decision boundary
  A more complex (nonlinear) decision boundary fitted to the score densities is likely to improve performance
References


