LivDet 2011- Fingerprint Liveness Detection Competition 2011

Final Report

David Yambay, Stephanie Schuckers Clarkson University

Luca Ghiani, Gian Luca Marcialis, Fabio Roli, University of Cagliari, Italy

CITeR Spring 2011©

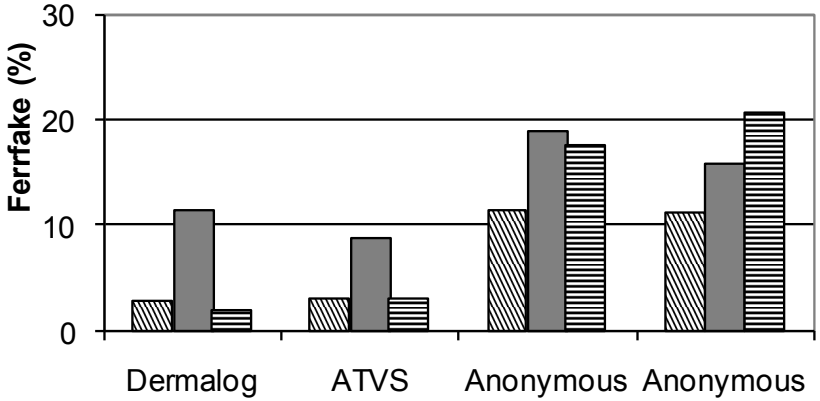
Problem

- Fingerprint vulnerable to artificial reproductions made of silicone, gelatin, Play-Doh, etc.
- Liveness detection proposed to check the vitality of fingers
- Many detection approaches published and tested on their home-made live and spoof databases

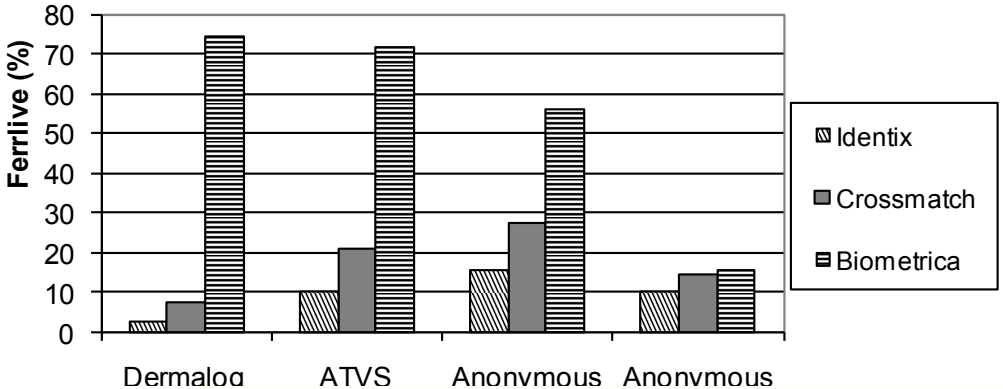
Liveness Detection Competition—LivDet 2009

- *First liveness detection competition at ICIAP 2009 with a public liveness database*
- Collaboration with Univ. of Cagliari
- Focusing on software-based fingerprint liveness
- Scanners used: CrossMatch, Identix, Biometrika
- 2000 live and spoof samples for each scanner
- Four participants

Ferrfake for Submitted Algorithms
rate of misclassified fake fingerprints



Ferrlive for Submitted Algorithms
rate of misclassified live fingeprints



LivDet 2011

- Second Liveness Detection competition—LivDet 2011
- The focus of this competition expanded from that of the first competition
- There are two parts for entrants
  - Part 1: Algorithms – similar to LivDet 09 with expanded spoof types
  - Part 2: Systems – Submission of hardware systems

Our Approach
LivDet II Competition—Algorithms

- Open to academic and industrial institutions
- Supply public fingerprint liveness database
  - Four optical sensors (Biometrika, Digital Persona, ItalData, Sagem)
  - Live database with different quality levels
  - High quality spoof database made of five different materials
    - Playdoh, Gelatin, Silicone and Woodglue on all devices
    - Latex on Digital Person and Sagem
    - Ecoflex on Biometrika and ItalData
- Setup server for downloading training dataset after signing license agreement
- Build the performance evaluation structure (experimental protocol) for the participants
- Accept submissions for algorithms as Win32 console applications
- Process the executable application file on the test dataset from different submitted algorithms
- Present the competition results on conference in 2011 (e.g. Biometric Consortium) and future journal
- Dataset made available to researchers after competition

Device Characteristics

Dataset	Sensor	Model No.	Resolution (dpi)	Image Size
#1	Biometrika	FX2000	500	315x372
#2	Digital Persona	4000B	500	355x391
#3	ItalData	ET10	500	640x480
#4	Sagem	MSO300	500	352x384

- Resolution was kept consistent across datasets
- Image size was allowed to vary

Device Characteristics - Live

Dataset	Sensor	Live Training Samples	Live Testing Samples	Number of Subjects	Number of Fingers	Images per Finger
#1	Biometrika	1000	1000	100	2	10
#2	Digital Persona	1000	1000	100	2	10
#3	ItalData	1000	1000	100	2	10
#4	Sagem	1000	1000	56	2	-

- 10 images were collected per finger per subject
- Sagem images per finger varied from subject to subject

Device Characteristics - Spoof

Dataset	Sensor	Ecoflex Training	EcoFlex Testing	Number of Subjects
#1	Biometrika	200	200	20
#2	Digital Persona	0	0	0
#3	ItalData	200	200	20
#4	Sagem	0	0	0
Dataset	Sensor	PlayDoh Training	PlayDoh Testing	Number of Subjects
#1	Biometrika	0	0	0
#2	Digital Persona	200	200	20
#3	ItalData	0	0	0
#4	Sagem	200	200	40
Dataset	Sensor	Gelatin Training	Gelatin Testing	Number of Subjects
#1	Biometrika	200	200	20
#2	Digital Persona	200	200	25
#3	ItalData	200	200	20
#4	Sagem	200	200	40

- Number of subjects varied per dataset based on quality of spoof images
- Playdoh and Ecoflex were used on only two of the datasets

Device Characteristics - Spoof

Dataset	Sensor	Latex Training	Latex Testing	Number of Subjects
#1	Biometrika	200	200	20
#2	Digital Persona	200	200	20
#3	ItalData	200	200	20
#4	Sagem	200	200	20
Dataset	Sensor	Silicone Training	Silicone Testing	Number of Subjects
#1	Biometrika	200	200	20
#2	Digital Persona	200	200	20
#3	ItalData	200	200	20
#4	Sagem	200	200	20
Dataset	Sensor	Wood Glue Training	Wood Glue Testing	Number of Subjects
#1	Biometrika	200	200	20
#2	Digital Persona	200	200	20
#3	ItalData	200	200	20
#4	Sagem	200	200	20

- Performed a visual inspection of spoof images
- Rejected images that were missing portions of the image or were of extreme poor quality

Our Approach
LivDet II Competition—Systems

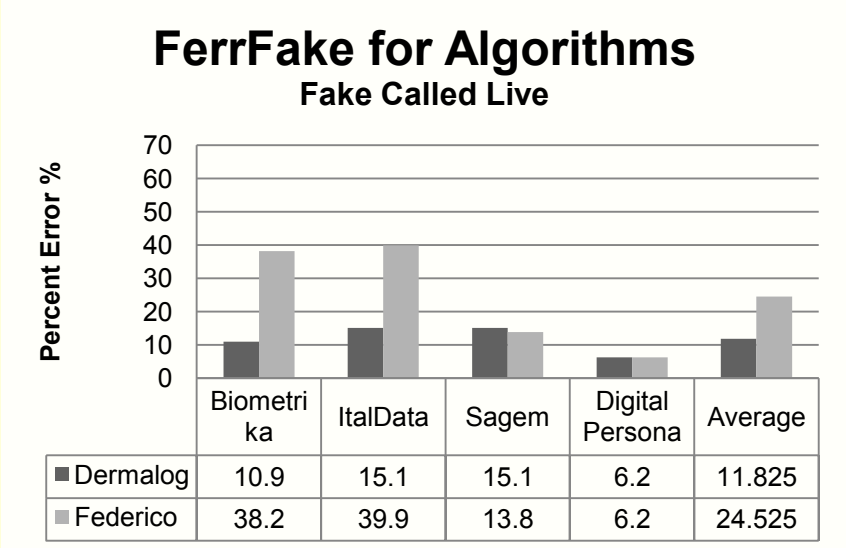
- Open to academic and industrial institutions
- Trained systems to be submitted for evaluation
- Accept submitted hardware/software systems
- System input (two modes: enrollment and verification)
  - Fingerprint placed on sensor
- System output
  - Collected image
  - Corresponding match score and liveness score for each image output
  - Failure to acquire
- Laboratory staff will systematically attempt to spoof the system and also collect corresponding live data
  - 750 attempts for five different materials (Play-Doh, gelatin, silicon, Body Double, and latex)
  - 3 images per spoof, 2 fingers per subject, 25 subjects
  - 500 live attempts from 50 people
  - 5 images per finger, 2 fingers per subject, 50 subjects
- Build the performance evaluation structure (experimental protocol) for the participants
- Present the competition results at Biometric Consortium and future journal

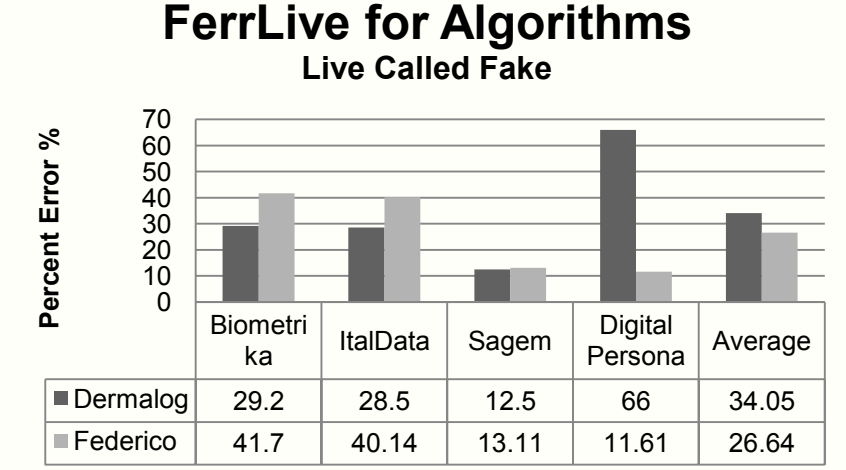
Submissions

- Four submissions were received for each of the two parts of the competition.
- Part 1: Algorithm Submissions
  - Dermalog Identification Systems GmbH (Dermalog)
  - Federico II University (Federico)
- Part 2: System Submissions
  - Dermalog
  - Greenbit Biometric Systems
- Dermalog submitted a revised algorithm after the closure of the competition due to an error in their program (for Digital Persona dataset only).

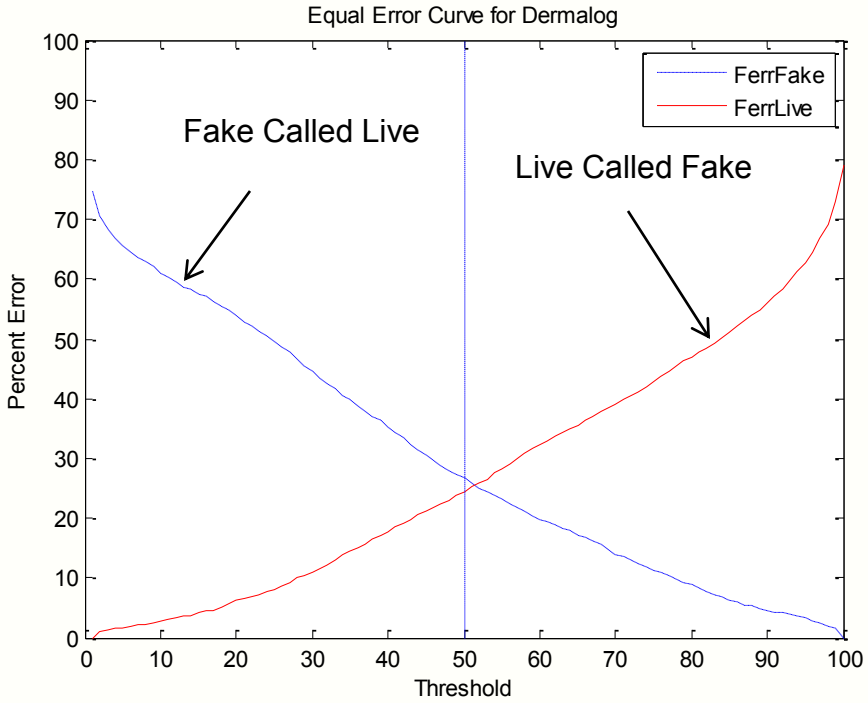
Part 1: Algorithm Results

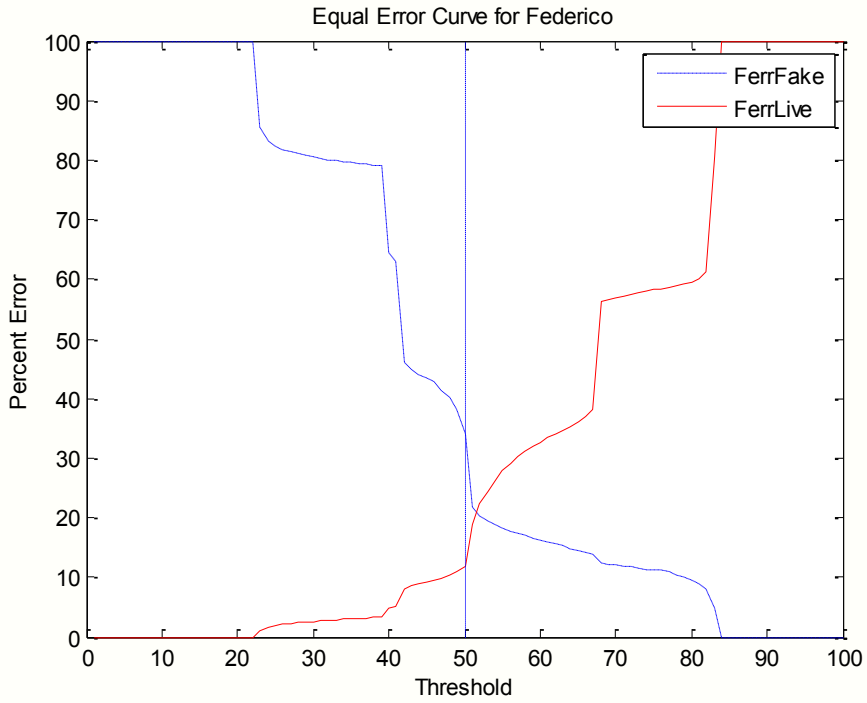
- Threshold value for testing was set at 50%
- Frederico had the best results on a single dataset with the Digital Persona Dataset
- Overall, Dermalog had the best results with an overall classification error rate of 22.9% compared to Frederico's 25.6%





Part 1: Overall Classification Error Rate





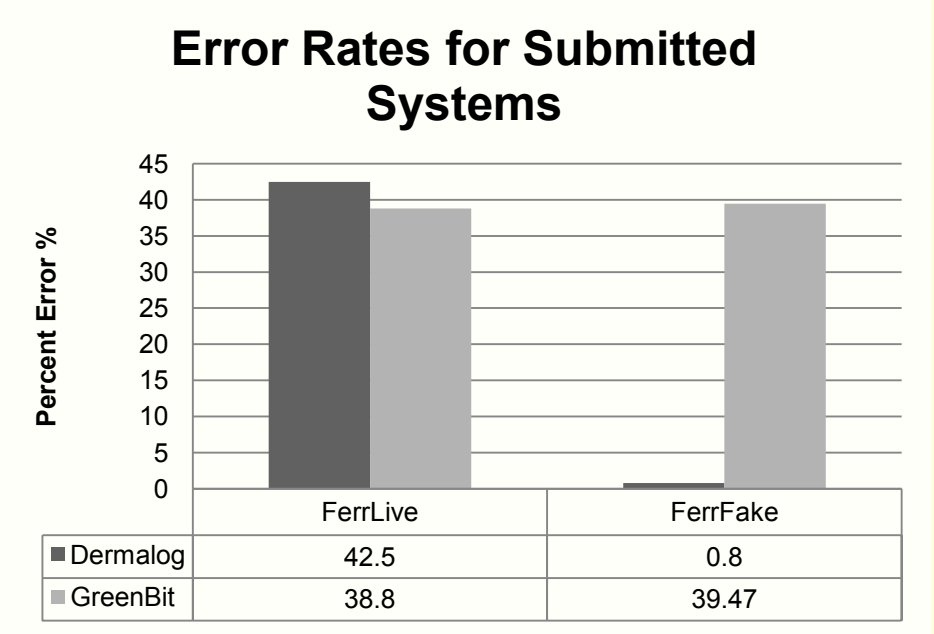
The equal error rate is near a threshold of 50 for both algorithms.

Part 1: Algorithm Results

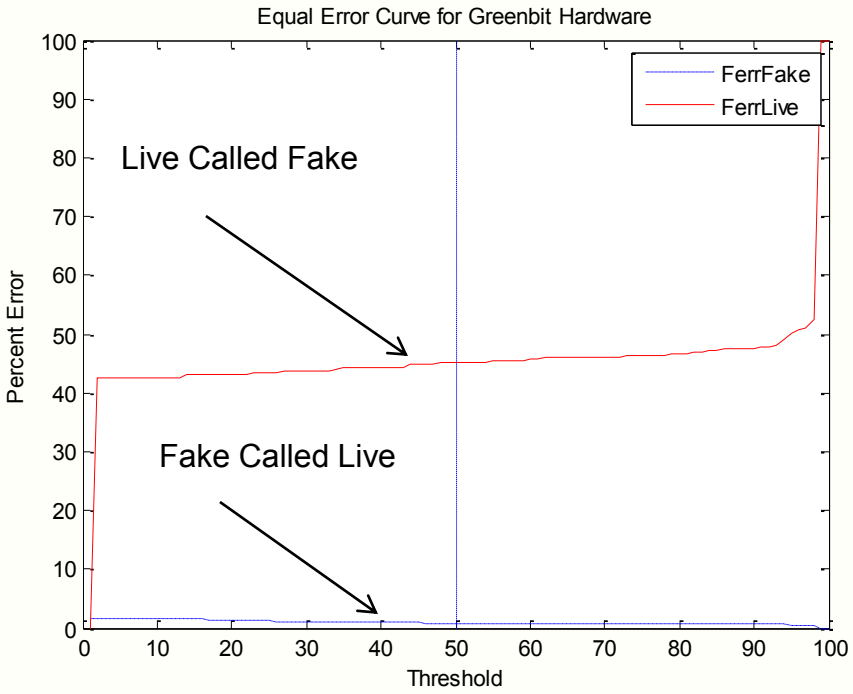
- Both Algorithms had a 0% failure to enroll rate
- Dermalog had a processing time approximately 10x faster than that of Federico
- Dermalog processed images at an average elapsed time of 0.28 seconds per image
- Federico processed images at an average elapsed time of approximately 3 seconds per image

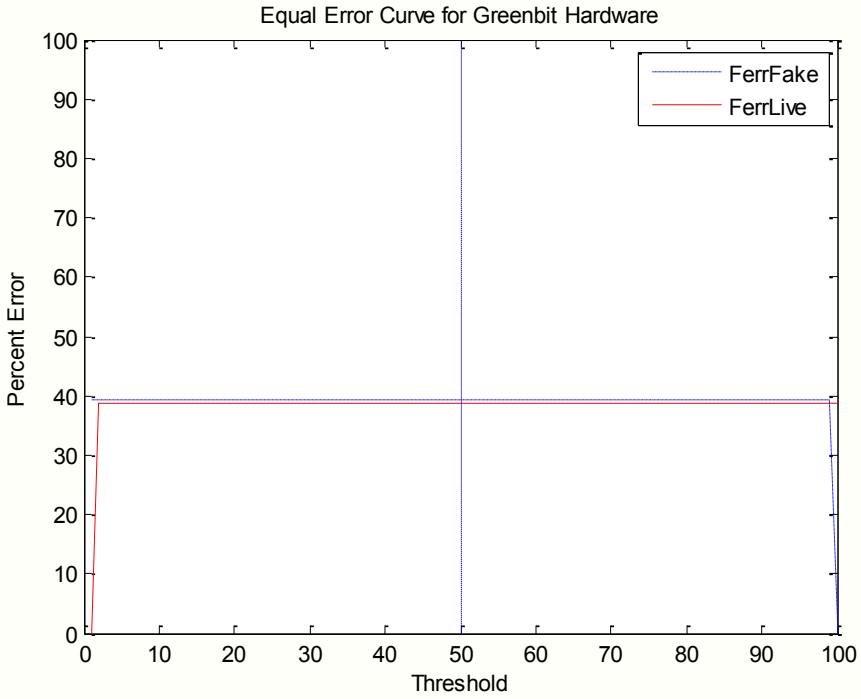
Part 2: System Results

- Dermalog received the overall best results in Part 2: Systems
- Dermalog has classification error rates of 0.8% FerrFake and 42.5% FerrLive
- Greenbit had consistent errors, but overall higher
- Greenbit has classification error rates of 39.5% FerrFake and 38.8% FerrLive



Part 2: Equal Error Rate Curves

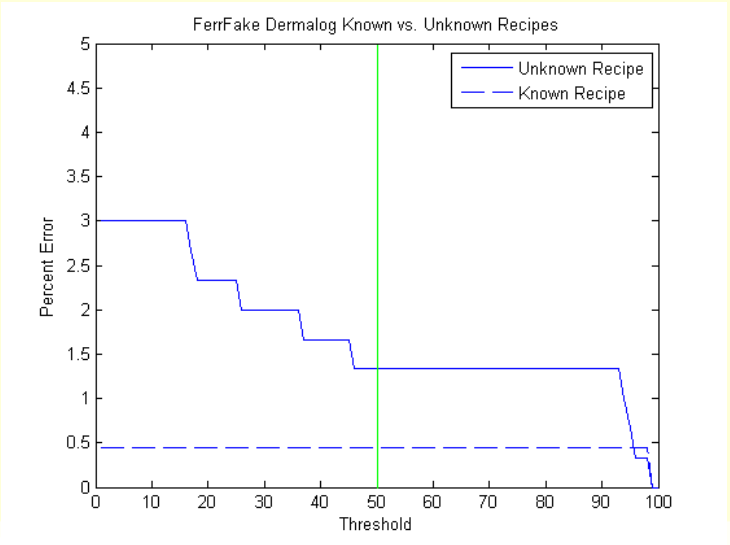


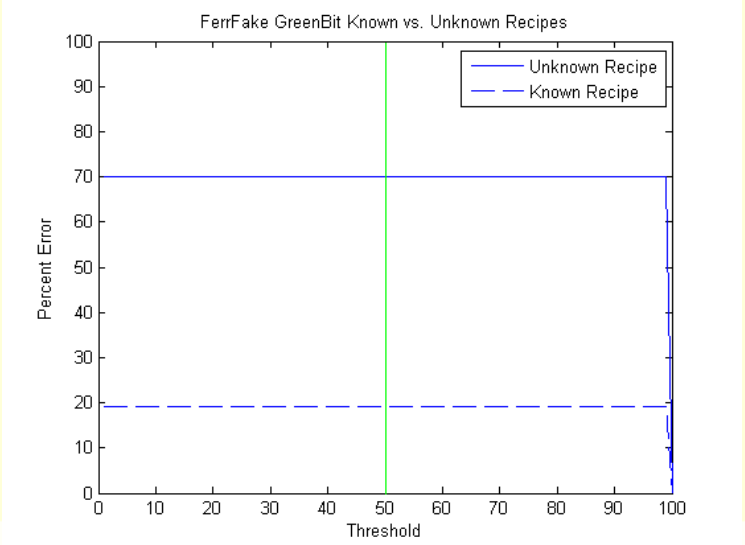


- Changing the threshold does not significantly change the results for the system.

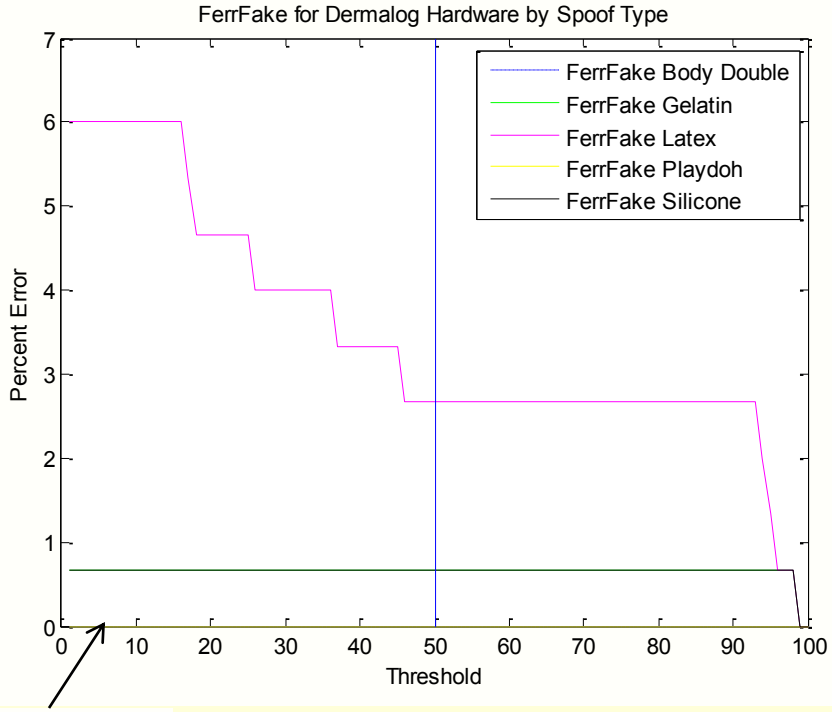
Part 2: Known vs. Unknown Recipes

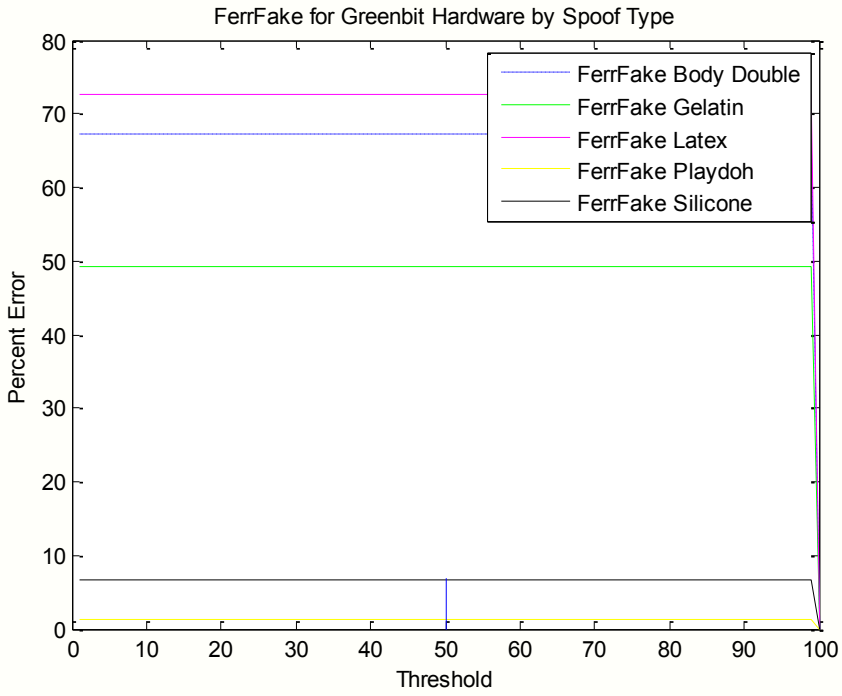
- The spoofing side of Part 2: Systems consisted of 5 different spoof recipes, 3 known and 2 unknown
- Unknown recipes had larger error rates than known
- Both systems had error rates approximately 3.5x larger for the recipes that were unknown compared to known.





Part 2: FerrFake Per Spoof Type





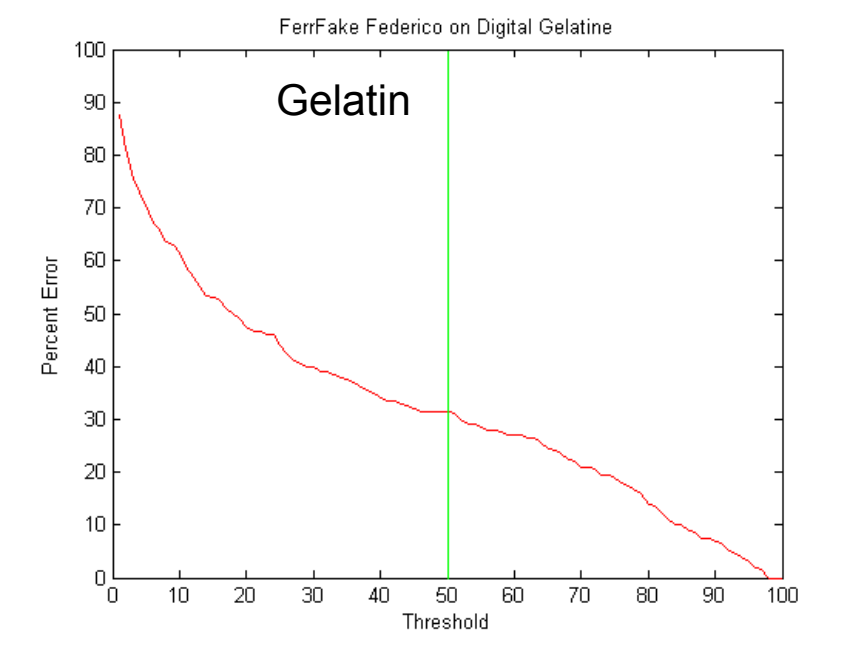
Three Lines at 0%

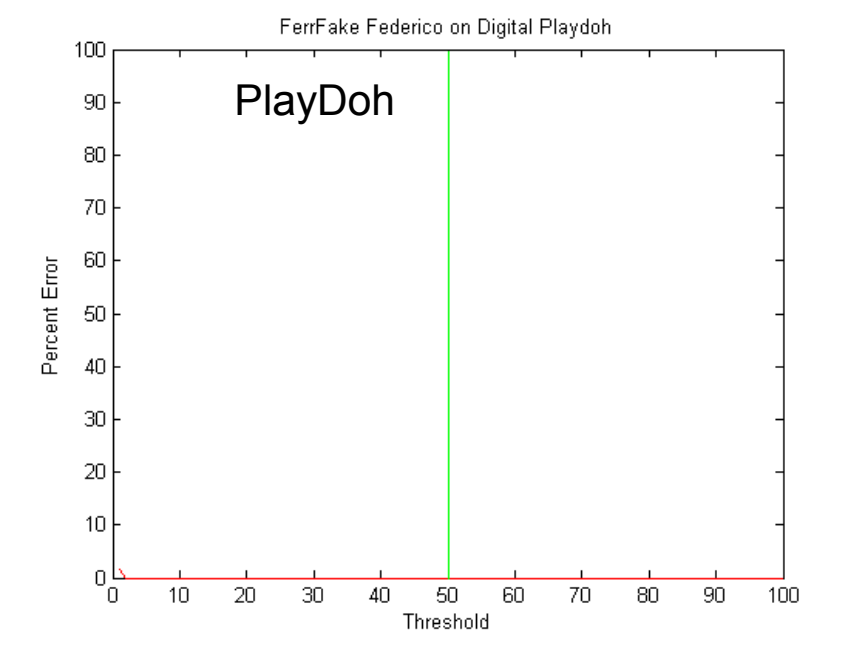
- Unknown Recipes (Latex, Body Double) had much higher error overall error rates than known recipes (Gelatin, Playdoh, Silicone)

Discussion of Part 1: Algorithm Results

- In Part 1: Algorithms, the algorithms had generally low scores for the Sagem and Digital Persona Dataset
- The algorithms had generally high scores for the Biometrika and ItalData
- Each submitted algorithm had certain spoof materials that they were strong against and some that they were weaker against
- This can seem to cause the higher error rates that we are seeing for overall error rates.

Part 1: Example of FerrFake Per Spoof Material (Federico)





- Federico Algorithm for Digital Persona Dataset:
  - 0% FerrFake on Silicone, Playdoh and Wood Glue
  - 30% FerrFake on Gelatin and Latex
  - Overall 6.2% FerrFake

Discussion of Part 2: System Results

- Both systems had unexpectedly high FerrLive scores
- Dermalog seemed to have an advantage against spoofs being a heated scanner as opposed to the non-heated GreenBit
- The heated scanner was able to melt some of the spoofs, specifically gelatin, rendering them useless

Part 2: Example Rejected and Accepted Live Images

Clarkson UNIVERSITY defy convention

Rejected Images on Dermalog

Accepted Images on Dermalog

Part 2: Example Spoof Images

Clarkson UNIVERSITY defy convention

Dermalog

A B C

GreenBit

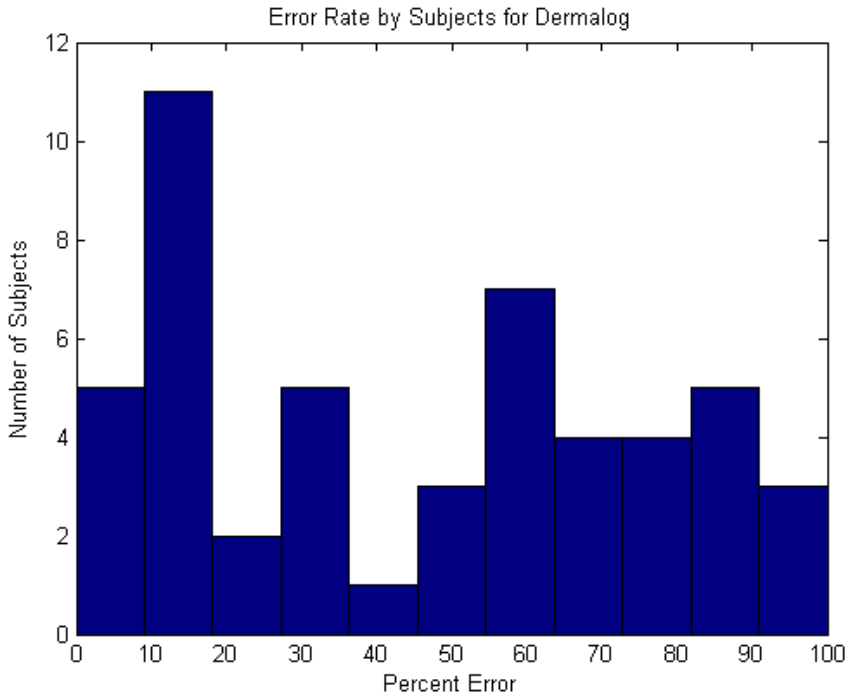
A B C

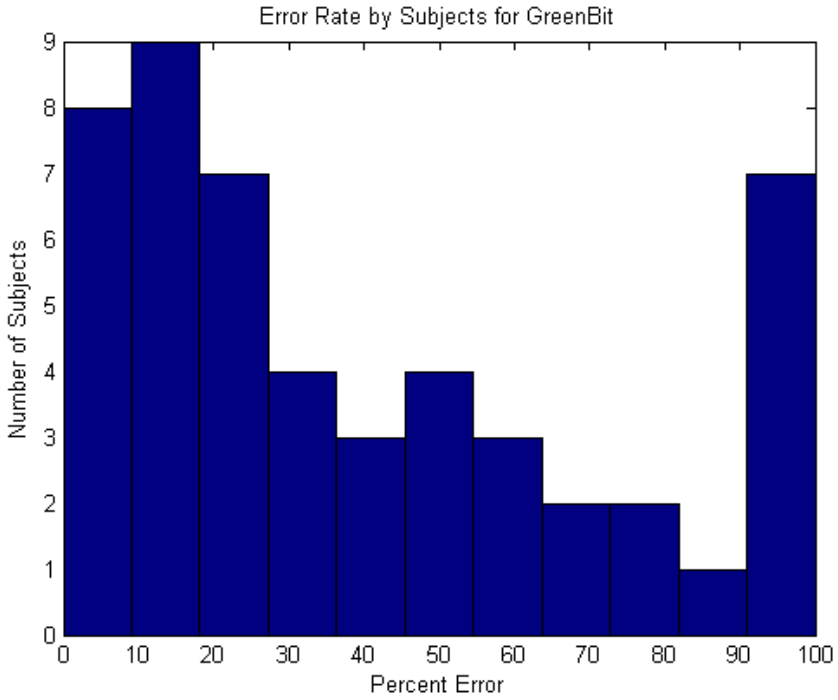
D E F

D E F

Images from Left to Right for both systems. A: Live, B: Body Double, C: Gelatin, D: Latex, E: Playdoh, F: Silicone

Part 2: Histograms of System Errors per Subject





- Histogram of number of error per subject
- No distinct pattern for errors across subjects

Conclusions

- Best overall results were shown by Dermalog in both Part 1 and Part 2 of the competition
- It is hoped that this competition will be continued in order to promote the state of the art in Liveness Detection
- Creating effective solutions are an important step in minimizing the vulnerability of spoof attacks

Current and Next Steps

- Process the datasets using quality matchers NFIQ and VeriFinger
- Apply match and decision level fusion techniques to both the algorithm and system datasets
- One algorithm submission was not originally received and will be tested against the datasets and results reported at a future time

© Clarkson University