usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

CAPTCHA

EJ Jung

usfCS UNIVERSITY of SAN FRANCISCO department of computer science

CAPTCHA

- stands for Completely Automated Public Turing test to tell Computers and Humans Apart
- Reverse Turing test
  - Turing test: how to tell an intelligent computer apart
  - from Wikipedia
    - it proceeds as follows: a human judge engages in a natural language conversation with one human and one machine, each of which try to appear human; if the judge cannot reliably tell which is which, then the machine is said to pass the test.
  - remember Blade Runner?
- Human Interactive Proof

usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

Turing test example

- Imagine that two players are playing Jeopardy over the Internet by typing in answers.
- In one window, a real human person answers.
- In the other, Watson answers.
- Would you be able to tell which is which?

usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

Robots can do more and faster

- Botnets can do even more
- Crawlers may ignore robot.txt
- Bots leave malicious contents as comments, postings, emails and collect informations
- Web spam is legal (spam is not)
  - btw, http://www.ncsl.org/programs/lis/CIP/hacklaw.htm
  - http://www.usfca.edu/its/about/policies/aup/

usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

Motivation for attack

- Search engine
  - more links, higher ranking
  - e.g. Google's page rank
- Advertisement
  - mimic "word of mouth"
- Phishing
  - disguise as suggestions and recommendations

usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

Motivation Beyond the Web

- Prevent dictionary attacks in any password system (Pinkas & Sander)
  - after failures, ask for CAPTCHA and the password
- Deter massive attacks
  - botnets may not pass CAPTCHA
  - humans are much slower
  - ask for CAPTCHA for any suspicious activity

usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

Precursors

- Unpublished manuscript by Moni Naor first mentions automated Turing test in 1997, but not proposed or formalized.
- Alta Vista patent in 1998 first practical example of using slightly distorted images of text to deter bots.
  - broken later by OCR

usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

Definition

- In 2000, formalized by Luis von Ahn, Manuel Blum & Nicholas J. Hopper of Carnegie Mellon; John Langford of IBM
- "A CAPTCHA is a cryptographic protocol whose underlying hardness assumption is based on an AI problem."
- www.captcha.net

- Advancing AI and security together
  - battle of breaking and improving

General Approaches

- Text (ASCII/Unicode)
- Image
- Speech
- Animation
- 3-D
- Combinations of all above

ASCII/Unicode ©4Pt¢h4

- Change text to look-alike: SPAM is $P4M. Fools simplest text matching.
- Accented or non-English chars: Spám
- Chars to words: uce@ftc.gov --> uce at ftc dot gov
- URL/HTML entities: COPY becomes ¢0Ρ¥ or %430P%59
- Better than nothing, but easy to crack

- This is not technically CAPTCHA

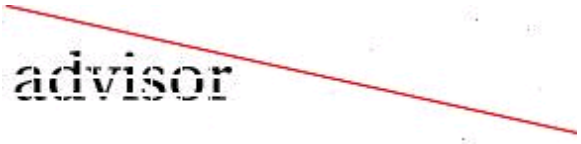
usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

Text Based CAPTCHAs

- Gimpy, ez-gimpy
  - Pick a word or words from a small dictionary
  - Distort them and add noise and background
- Gimpy-r
  - Pick random letters
  - Distort them, add noise and background
- Simard's HIP
  - Pick random letters and numbers
  - Distort them and add arcs

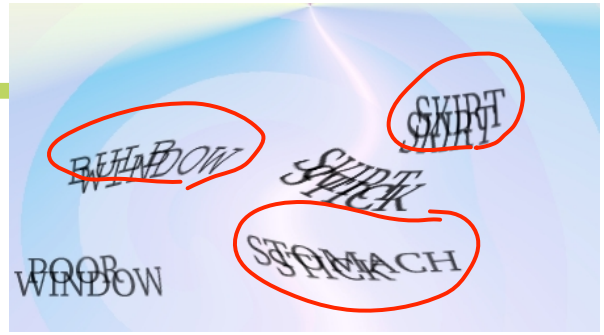
usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

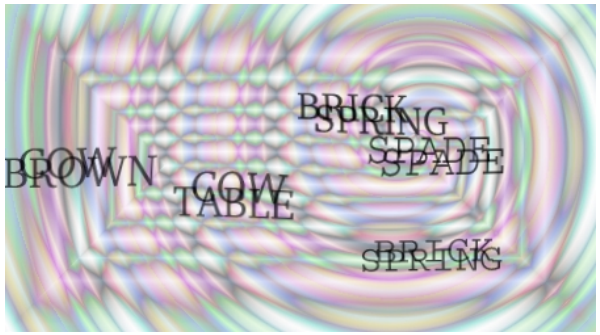
Text Based CAPTCHAs

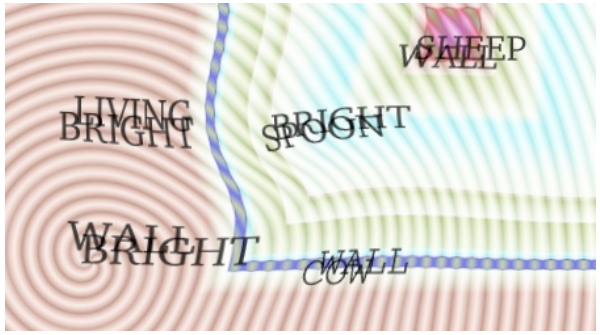


Gimpy

- First generation
  - Pick a word from dictionary
  - Random placement, font, distortion, background pattern
  - Overlapping words serve as noise.
- Frequently cracked and improved.
  - http://www.cs.sfu.ca/~mori/research/gimpy/
- In current version, 5 pairs of overlapped words. User identifies 3 words.

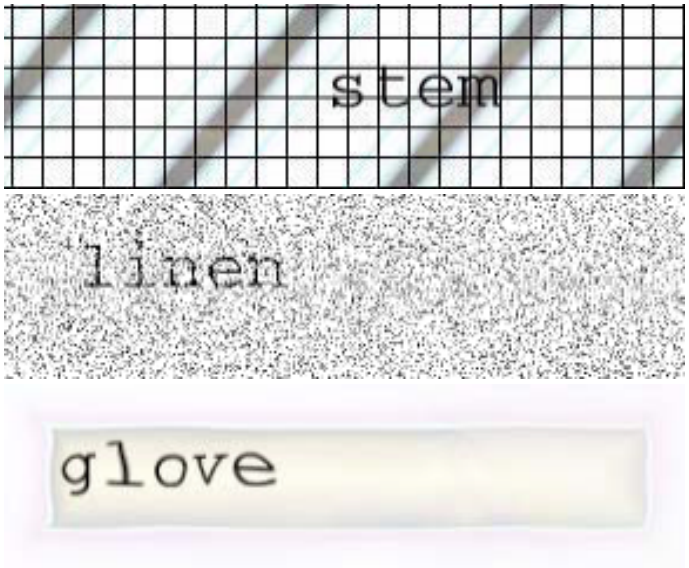






EZ-Gimpy

- Pick a word or words from a small dictionary
- Distort them and add noise and background



- 99% success in breaking
  - Distortion Estimation Techniques in Solving Visual CAPTCHAs, CVRP 2004

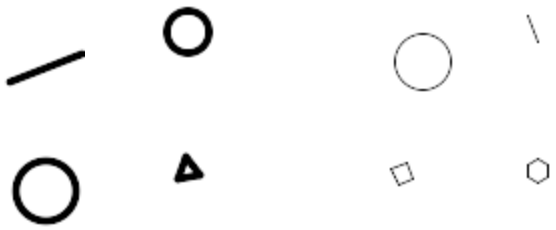
Gimpy-r

- Pick random letters
- Distort them, add noise and backgroun

- 78% success in breaking Gimpy-r
  - Distortion Estimation Techniques in Solving Visual CAPTCHAs, CVRP 2004

Bongo

- Visual pattern recognition puzzle
- Example: thick vs. thin
- User is presented with a new block and needs to pick left or right



usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

Pix

- Image recognition with keywords
- Procedure
  - display four images with the same keyword
  - provide a random set of keywords to choose from
  - user needs to pick the common keyword

usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

ESP-Pix

LIVESTOCK

Choose a word that relates to all the images.

TIP: You can type the first letter of a word and then use the down arrow to find it.

Submit

usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

Beating CAPTCHA

- OCR-base attacks
  - http://sam.zoy.org/pwntcha/
  - *Pretend We're Not a Turing Computer but a Human Antagonist*
- Heuristics
  - vary position, warp, noise, background, colors, overlap, randomness, font, angles, language,
- Accessibility problem for vision-impaired users
  - audio as well as visual
  - http://www.w3.org/TR/turingtest/

Classification-based approach

- Text-based CAPTCHA Strengths and Weaknesses [Bursztein,Martin,Mitchell CCS2011]
- Classify the given image to one of the words in synthetic corpus

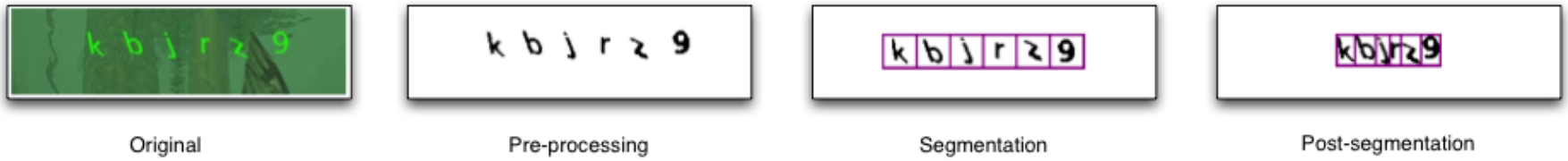


Figure 5: Example of the Blizzard pipeline

Figure 10: Example of the Slashdot pipeline

Real-World Captchas Summary

Scheme	Recall	Precision	Anti-segmentat
Authorize	84%	66%	background con
Baidu	98%	5%	collapsing
Blizzard	75%	70%	background con
Captcha.net	96%	73%	background confusion
CNN	50%	16%	line
Digg	86%	20%	line
eBay	95%	43%	collapsing
Google	0%	0%	collapsing
Megaupload	n/a	93%	collapsing
NIH	87%	72%	background confusion
Recaptcha	0%	0%	collapsing
Reddit	71%	42%	background confusion
Skyrock	30%	2%	background confusion
Slashdot	52%	35%	lines
Wikipedia	57%	25%	n/a

/total guess

/tp+fn

Table 2: Real world captchas summary

usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

Speech CAPTCHA

- Spell in synthesized or recorded voices
- Voice recognition vs. user's miss rate
- Use with visual CAPTCHA for increased accessibility
  - may help attackers guess correctly

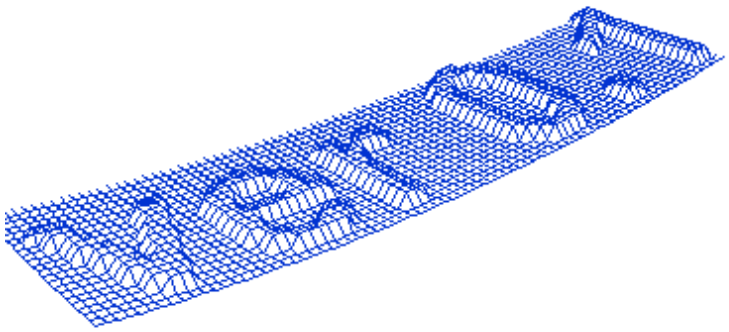
usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

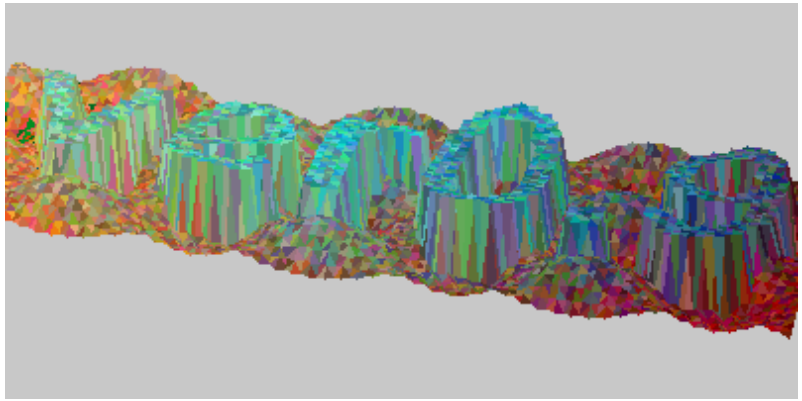
Animated CAPTCHA

- Can use Flash, MPEG, animated GIF
- Often combined with speech
- Weaknesses of Image CAPTCHA apply
- Usually *easier* to crack due to extra data for pattern matching to analyze
- Much higher processor and traffic load
- Not practical in most cases

3D

- tEABAG_3D
  - http://www.ocr-research.org.ua/index.php?action=teabag
- Renders the password in 3D image
- More difficult to crack then 2D images
- More resources on server
  - high load graphic processing
- Can be combined with other methods





usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

Beating CAPTCHA by humans

- Man-in-the-middle
  - copy CAPTCHA from the target
  - post on the attacker's website
  - forward the answer to the target
- CAPTCHA factory
  - http://taint.org/2008/03/05/122732a.html
- Reuse the session id
  - http://www.puremango.co.uk/cm_breaking_captcha_115.php

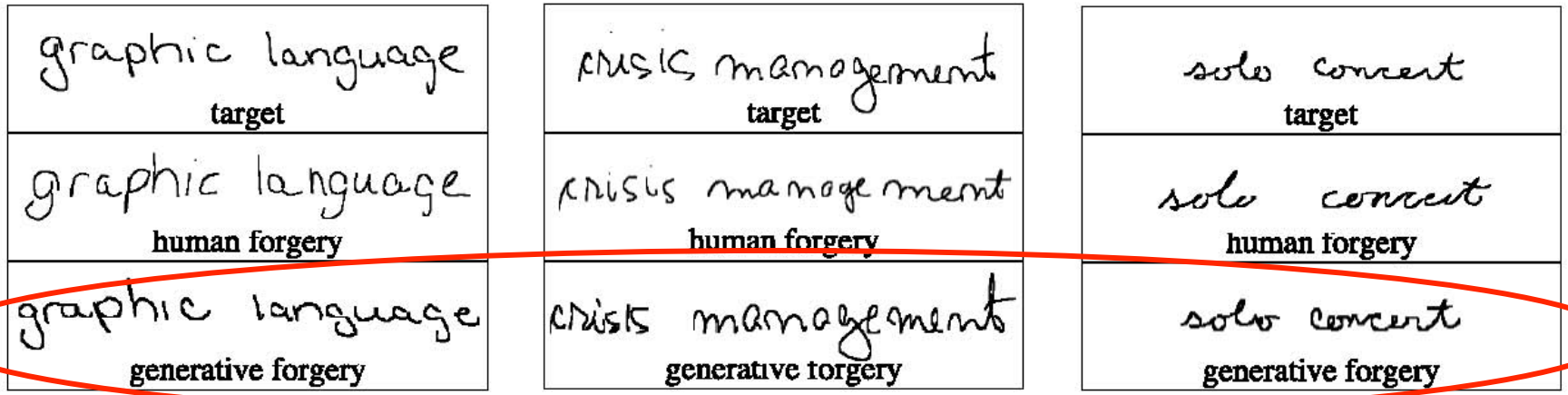
usfCS
UNIVERSITY of SAN FRANCISCO
department of computer science

Adopt CAPTCHA for yourself?

- Free software
  - http://www.google.com/recaptcha
  - http://captcha.net

Forging Handwriting

[Ballard, Monrose, Lopresti]



Generated by computer algorithm trained on handwriting samples

Cloning a Finger

UNIVERSITY of SAN FRANCISCO
department of computer science

[Matsumoto]

Making an Artificial Finger from a Residual Fingerprint

Materials

A photosensitive coated Printed Circuit Board (PCB)
"10K" by Sanhayato Co., Ltd .

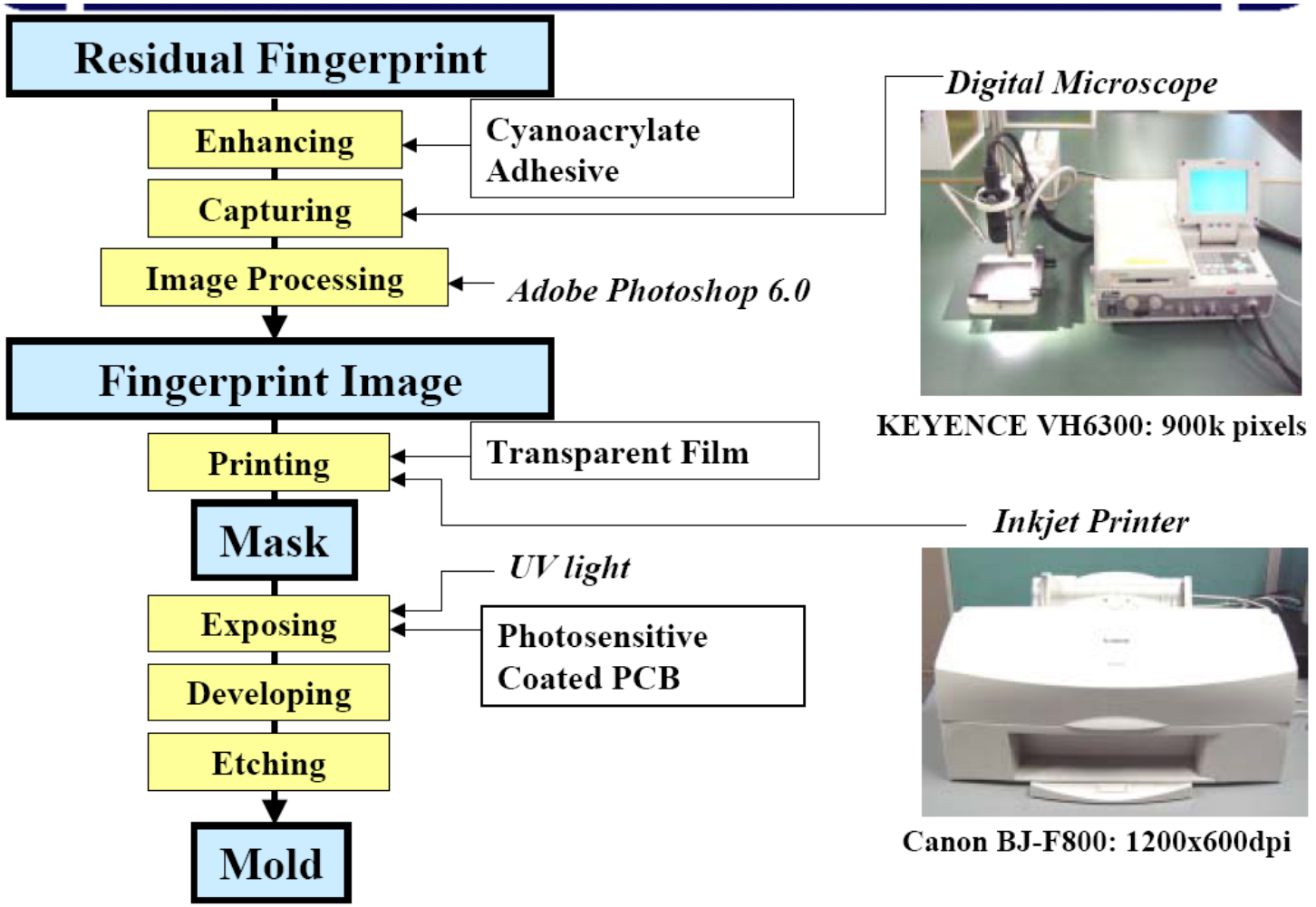
320JPY/sheet

Solid gelatin sheet
"GELATINE LEAF "
by MARUHA CORP

200JPY/30grams

Cloning Process

[Matsumoto]

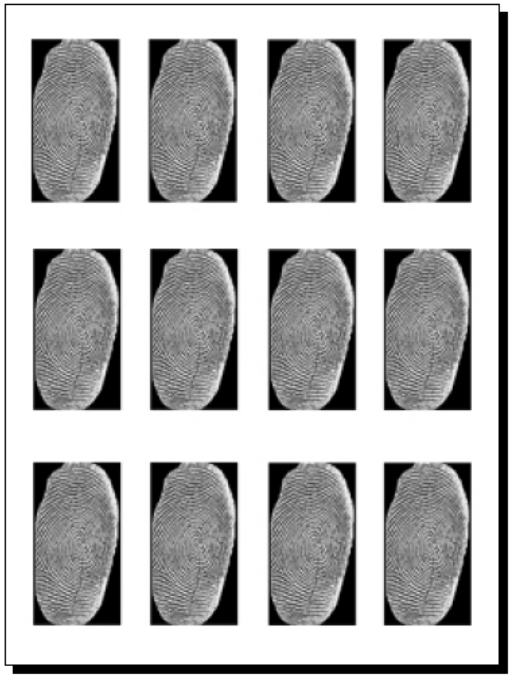


Fingerprint Image

[Matsumoto]

An Enhanced Fingerprint

A Fingerprint Image



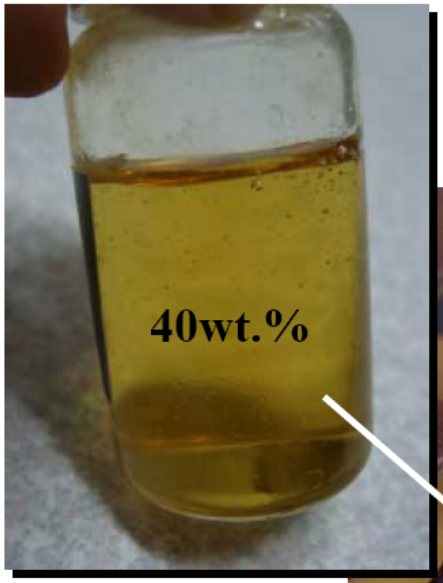
A Mask with Fingerprint Images

Molding

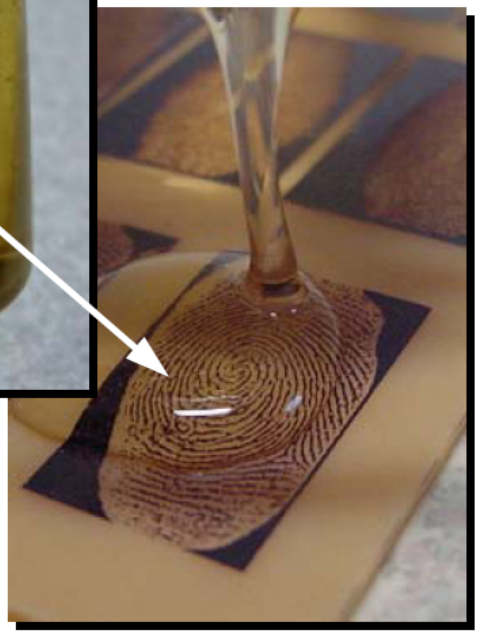
UNIVERSITY of SAN FRANCISCO
department of computer science

[Matsumoto]

Gelatin Liquid



Drip the liquid onto the mold.

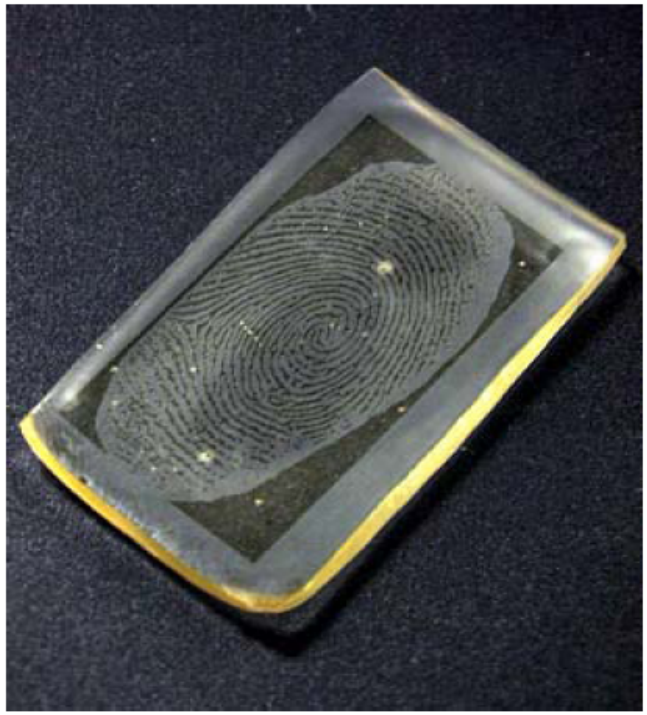


Put this mold into a refrigerator to cool, and then peel carefully.

The Mold and the Gummy Finger

[Matsumoto]

Mold: 70JPY/piece
(Ten molds can be obtained in the PCB.)



Gummy Finger: 50JPY/piece

Side By Side

[Matsumoto]

Pores can be observed.

Enhanced Fingerprint

Captured Fingerprint Image of
the Gummy Finger
with the device H (a capacitive sensor)

Play-Doh Fingers

[Schuckers]

- Alternative to gelatin
- Play-Doh fingers fool 90% of fingerprint scanners
  - Clarkson University study
- Suggested perspiration measurement to test "liveness" of the finger