Balancing Usability and Security in a Video CAPTCHA

Computer Science @ RIT
B.Thomas Golisano College of Computing & Information Sciences

Kurt Alfred Kluever
Google, Inc.
kak@google.com

Richard Zanibbi
Rochester Institute of Technology
rlaz@cs.rit.edu

Symposium on Usable Privacy and Security (SOUPS) 2009
July 15th-17th, 2009, Mountain View, CA, USA

Balancing Usability and Security in a Video CAPTCHA

Google

Kurt Alfred Kluever

Computer Science @ RIT
B.Thomas Golisano College of Computing & Information Sciences

Richard Zanibbi

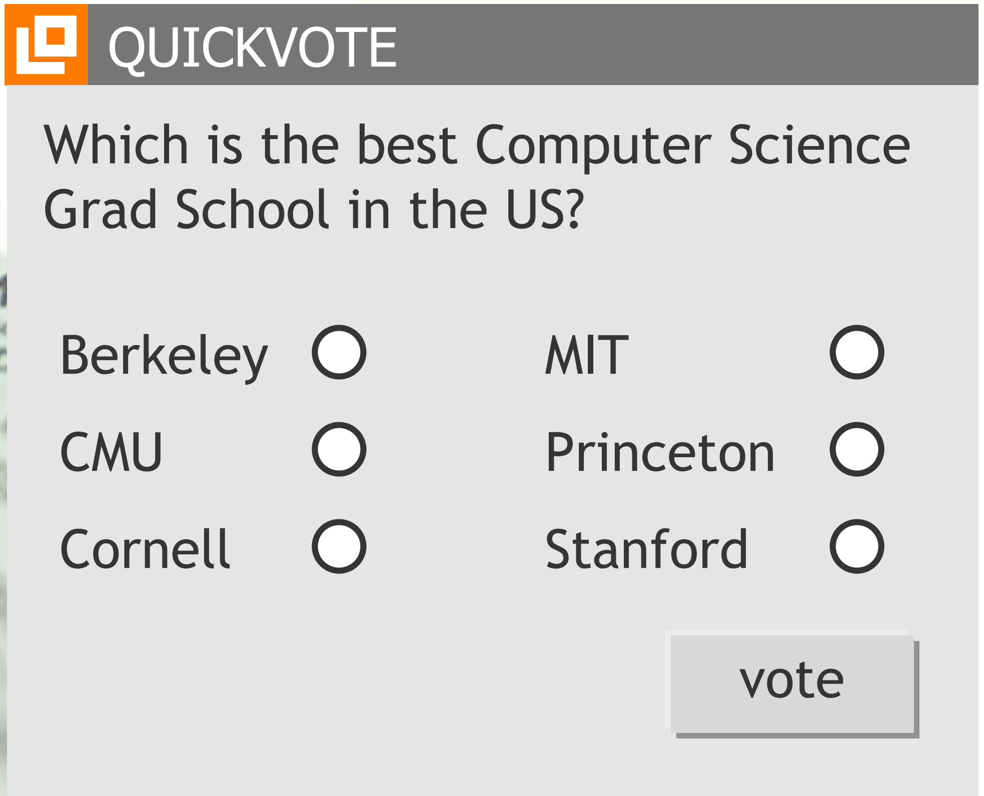
Symposium on Usable Privacy and Security (SOUPS) 2009
July 15th-17th, 2009, Mountain View, CA, USA

Overview

- Motivation + Brief History
- Desirable CAPTCHA Properties
- Video CAPTCHA + Research Goal
- Methodology
  - Data sources
  - Generating + Grading Challenges
- Attack Simulation
- Two User Studies
- Results + Comparison to Existing Work

Motivation: Abuse of Online Services

- Generate accounts to abuse free services
- Send SPAM from free email accounts
- Take advantage of free offers
- Buy hundreds of tickets for scalpers
- Brute force passwords
- Post spam to blogs
- Poison online polls



QUICKVOTE

Which is the best Computer Science Grad School in the US?

Berkeley	○	MIT	○
CMU	○	Princeton	○
Cornell	○	Stanford	○

Desirable CAPTCHA Properties

Automated

- The generation and grading of challenges is automatic

Open

- Underlying databases/algorithms are publicly available

*Usable**

- Frequently passed by humans

*Secure**

- Frequently failed by machines

> “*A CAPTCHA is a program that can generate and grade tests that it itself cannot pass (much like some professors).*” -Luis von Ahn

Existing CAPTCHA Types

Natural language processing

- “What is 4 times the number of legs a kangaroo has?”

Character recognition

- “Type the letters you see in this image.”

Image understanding

- “What are these images of?” / “Is this image upright?”

Automatic speech recognition

- “1-6-3-9-2-7” / Old radio broadcasts

Character Recognition-based

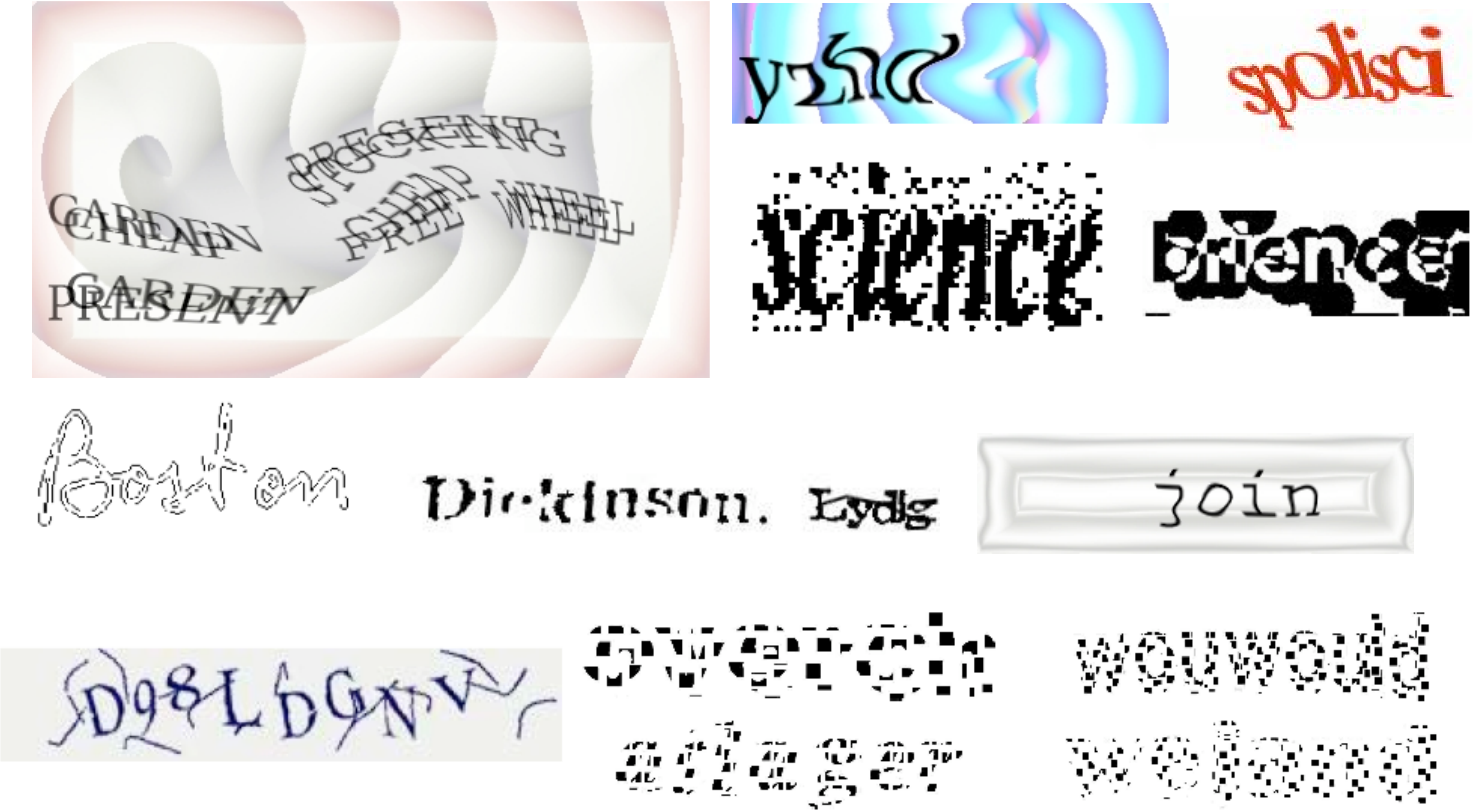
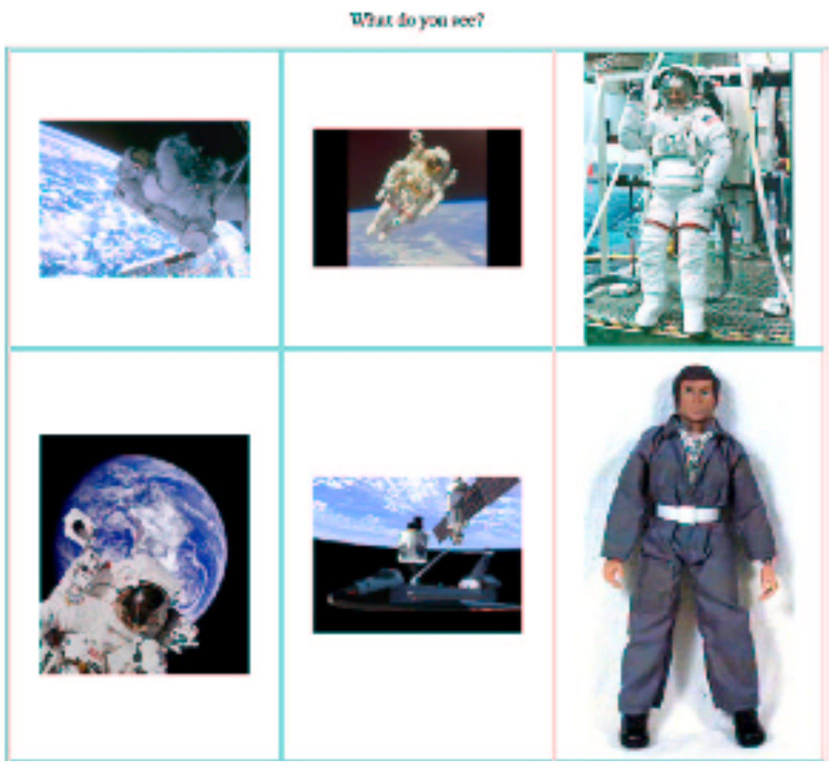
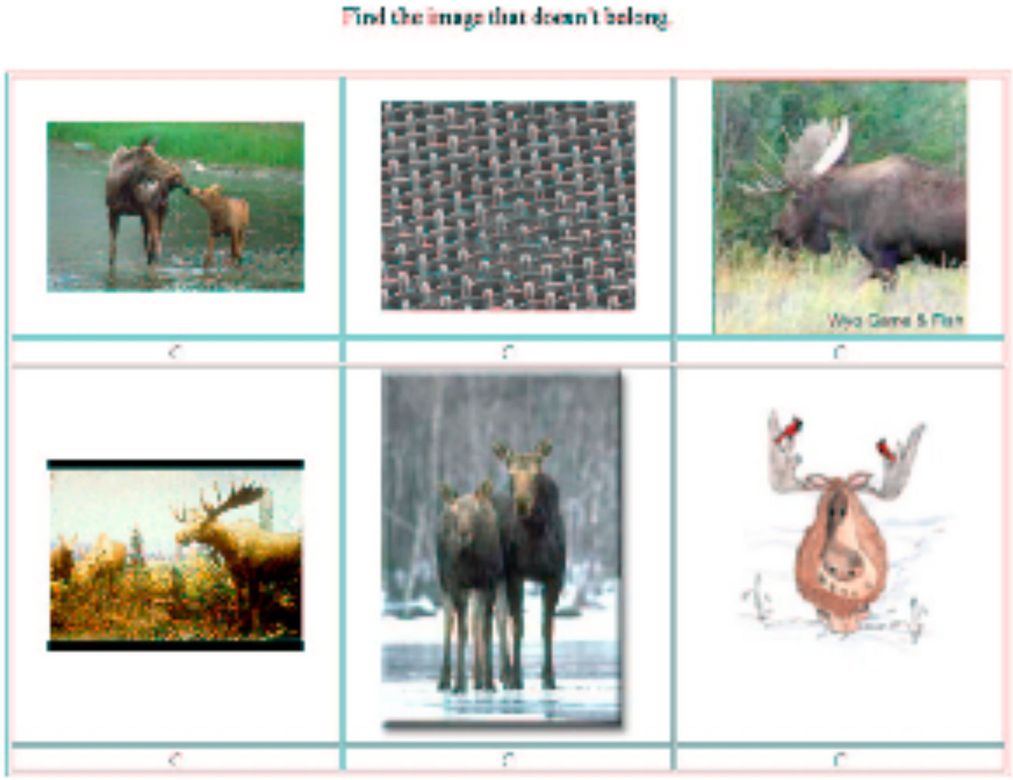


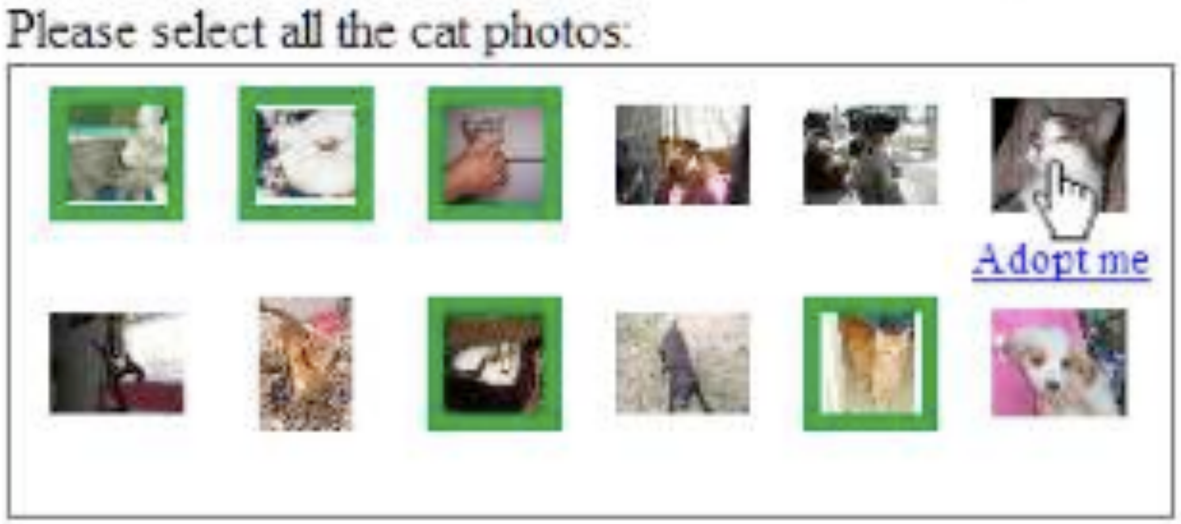
Image Recognition-based



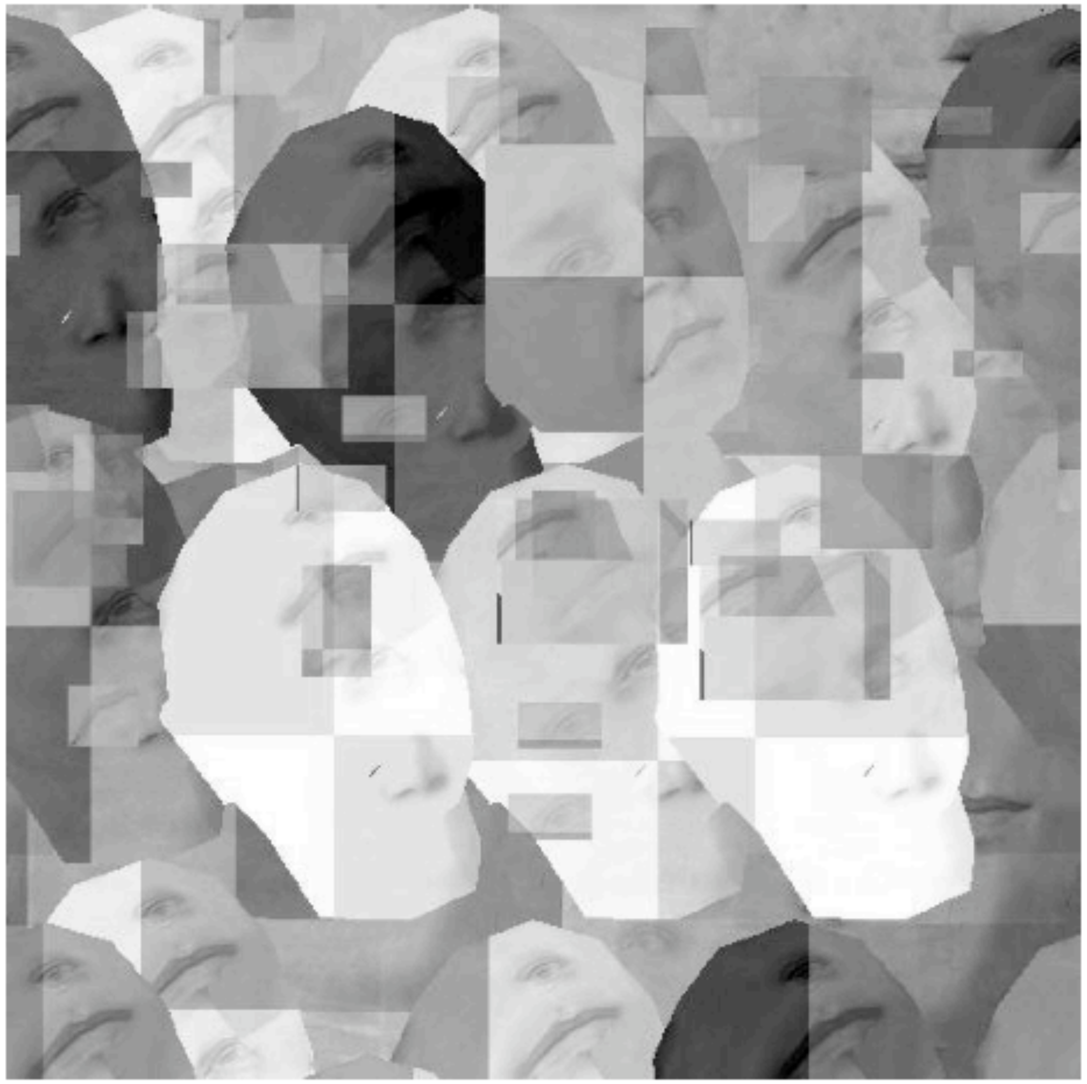
What do you see?



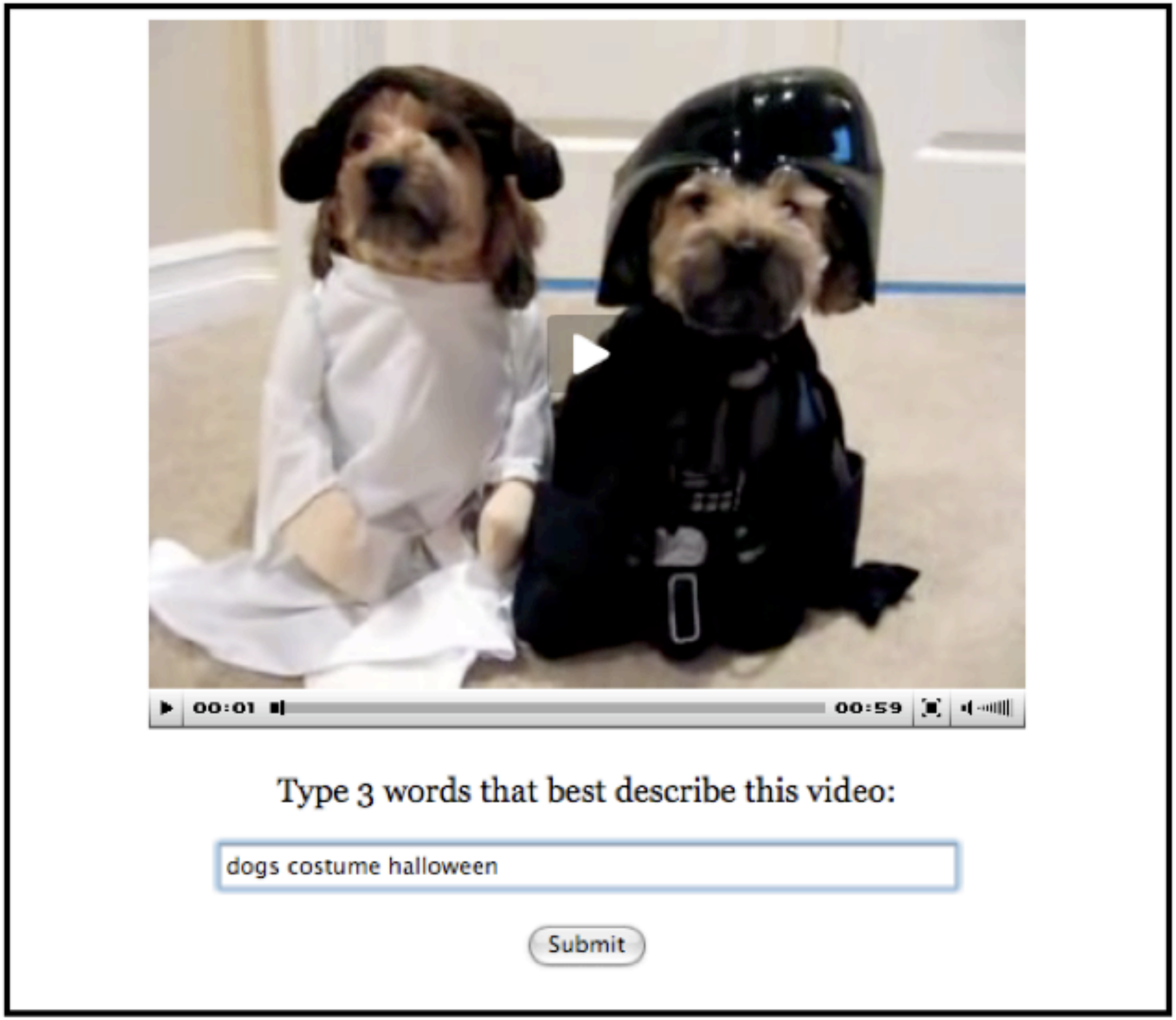
Find the image that doesn't belong.



Please select all the cat photos:



Video CAPTCHA



Type 3 words that best describe this video:

dogs costume halloween

Task:

Submit three tags, aiming to match one in a set of automatically generated ground truth tags.

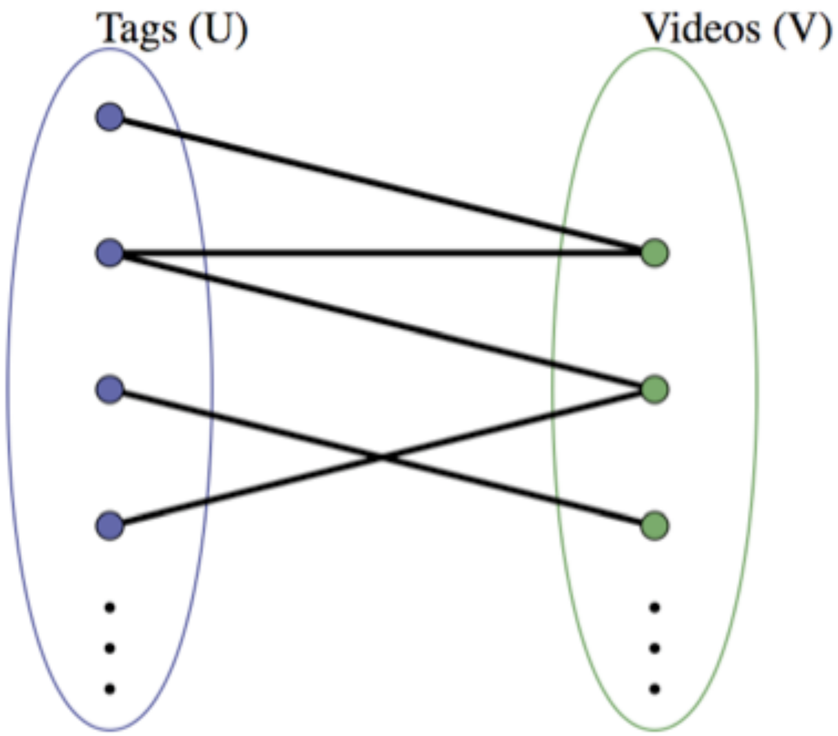
Public Video Dataset: YouTube.com

Generate a random YouTube ID...Good luck

- 64 possible characters; 11 characters long
- > 150 million videos on YouTube (August 2008)

Random walk (randomized local search)

- Query with a dictionary* word
- Randomly choose a video
- Randomly choose a tag
- Repeat for a random depth
  - [1, 100]



Generating Challenges

Use random walk to select a challenge video

From Related Videos set, add *n* additional tags (sorted by cosine similarity over tag sets)

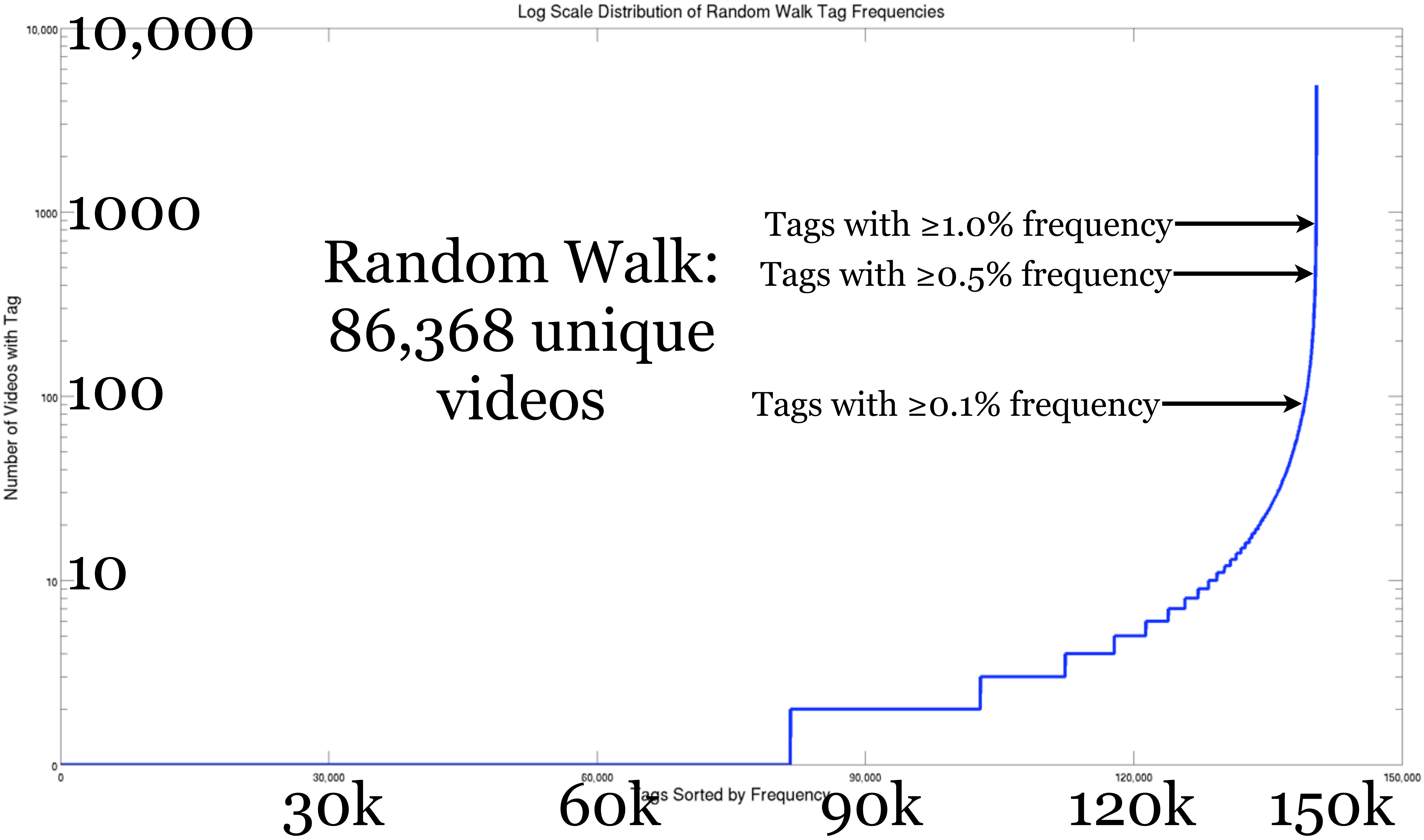
- *black box* algorithm (hard* to compute it ourselves)

Remove tags estimated to be more frequent than a threshold *t*

$$\textsc{Sim}(A, B) = \cos\theta = \frac{A \cdot B}{\|A\|\|B\|}$$

$$\cos\theta = \frac{|A_t \cap R_t|}{\sqrt{|A_t|}\sqrt{|R_t|}}$$

Tag Frequency Distribution



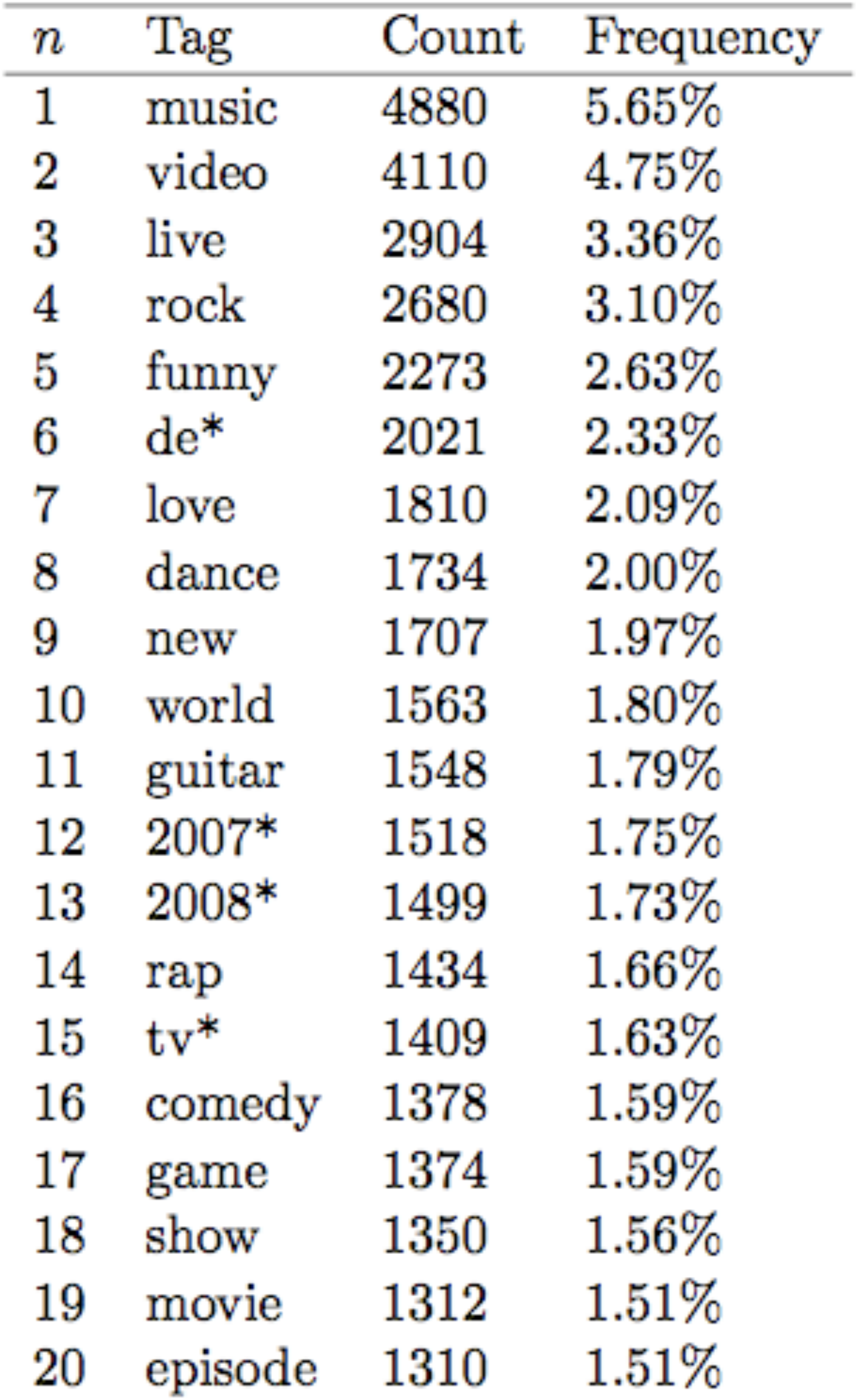
Random Walk: 86,368 unique videos

Tags with ≥1.0% frequency

Tags with ≥0.5% frequency

Tags with ≥0.1% frequency

Estimated Tag Frequencies



n	Tag	Count	Frequency
1	music	4880	5.65%
2	video	4110	4.75%
3	live	2904	3.36%
4	rock	2680	3.10%
5	funny	2273	2.63%
6	de*	2021	2.33%
7	love	1810	2.09%
8	dance	1734	2.00%
9	new	1707	1.97%
10	world	1563	1.80%
11	guitar	1548	1.79%
12	2007*	1518	1.75%
13	2008*	1499	1.73%
14	rap	1434	1.66%
15	tv*	1409	1.63%
16	comedy	1378	1.59%
17	game	1374	1.59%
18	show	1350	1.56%
19	movie	1312	1.51%
20	episode	1310	1.51%

Random walk of 86k YouTube videos

Many tags do not appear in our original dictionary

Grading Challenges

Normalize Input

- Lowercase, no punctuation or stop words, only 3 tags

Stemming

- Add word stems to ground truth (Porter algorithm)
- Adds at most 3 additional tags (‘dogs’ -> ‘dog’)

Levenshtein Edit Distance

- Allows for insertions, deletions, and substitutions
- Normalized threshold of 0.8

$$\textsc{NormalizedLevenshtein}(s_1, s_2) = 1.0 - \frac{\textsc{Levenshtein}(s_1, s_2)}{\textsc{Max}(|s_1|, |s_2|)}$$

Testing the Hypothesis

One may increase *usability* while maintaining *security* against a frequency-based attack in a video CAPTCHA by intelligently extending the set of *user-supplied* and *ground truth* tags.

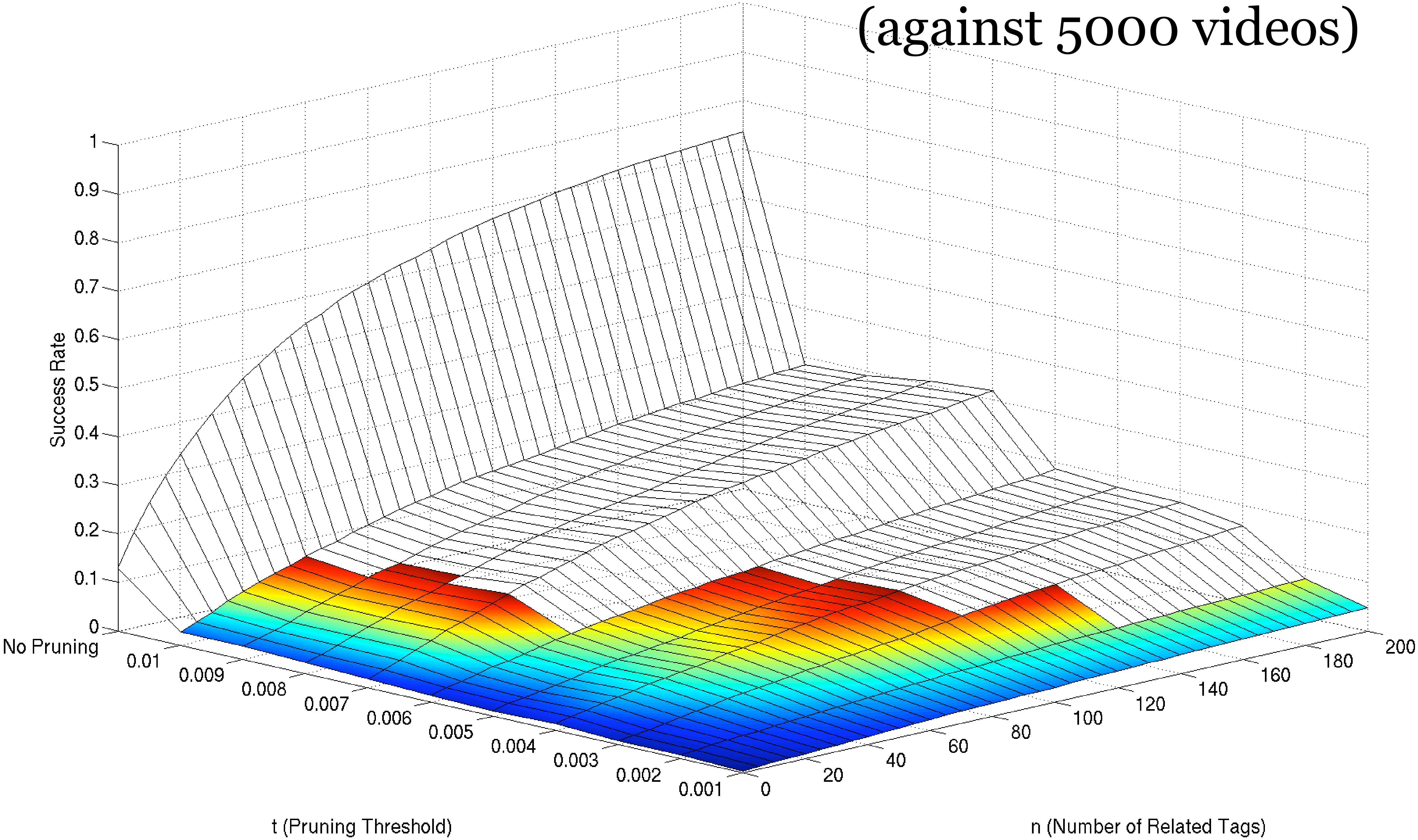
n	Number of related tags added.
t	Pruning threshold.
s	Use stemming?
l	Use inexact match?

Attack Tags Used

t	Best Attack Tags	# Pruned	$\hat{S}_c(A)$
1.0	[music, video, live]	0	0.1377
0.01	[dj, remix, vs]	37	0.0291
0.009	[girl, school, el]	44	0.0256
0.008	[animation, michael, star]	49	0.0237
0.007	[concert, news, day]	67	0.0207
0.006	[fantasy, dragon, rb]	92	0.0179
0.005	[islam, humor, blues]	129	0.0148
0.004	[real, bass, 12]	184	0.0120
0.003	[uk, spoof, pro]	302	0.0090
0.002	[seven, jr, patrick]	570	0.0060
0.001	[ff, kings, ds]	1402	0.0030

Attack Success Rates: Random Walk

(against 5000 videos)



Balancing Usability and Security in a Video CAPTCHA

Kurt Alfred Kluever
Richard Zanibbi

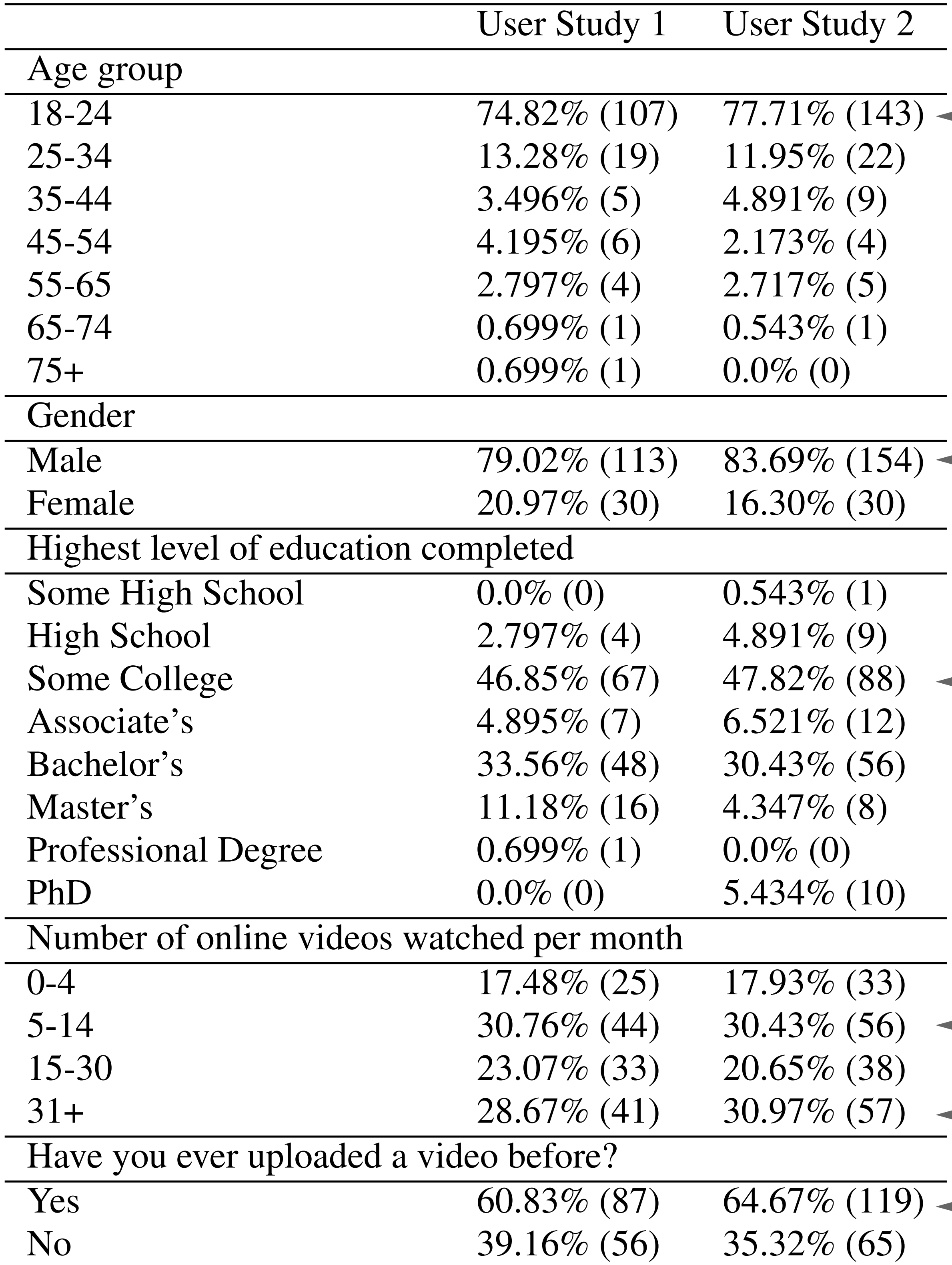
Two User Studies

Emails, flyers, word of mouth

Number of participants

- User Study 1:
  - 233 -> 143 (61.3%)
- User Study 2:
  - 300 -> 184 (61.3%)

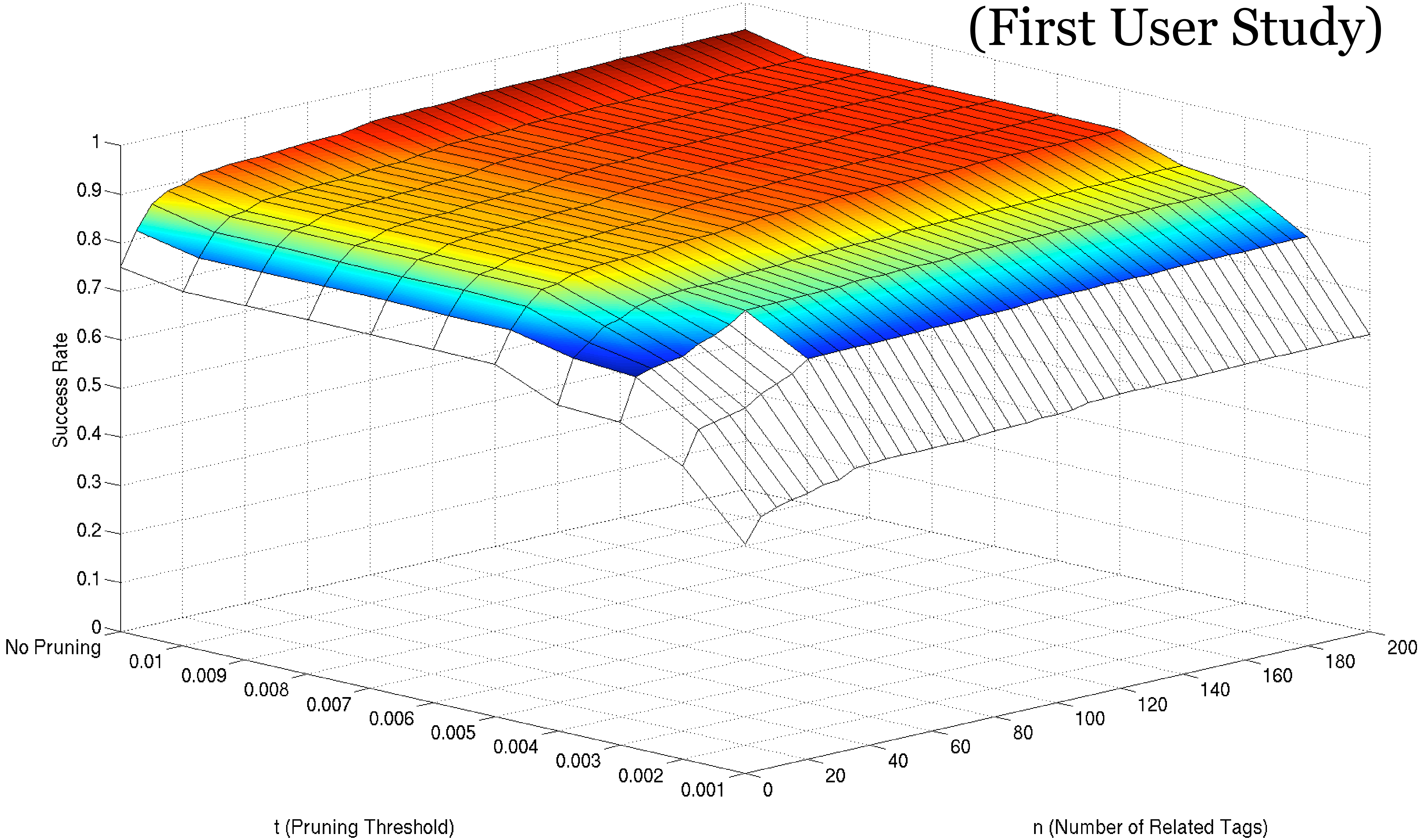
Online collection



	User Study 1	User Study 2
Age group		
18-24	74.82% (107)	77.71% (143)
25-34	13.28% (19)	11.95% (22)
35-44	3.496% (5)	4.891% (9)
45-54	4.195% (6)	2.173% (4)
55-65	2.797% (4)	2.717% (5)
65-74	0.699% (1)	0.543% (1)
75+	0.699% (1)	0.0% (0)
Gender		
Male	79.02% (113)	83.69% (154)
Female	20.97% (30)	16.30% (30)
Highest level of education completed		
Some High School	0.0% (0)	0.543% (1)
High School	2.797% (4)	4.891% (9)
Some College	46.85% (67)	47.82% (88)
Associate's	4.895% (7)	6.521% (12)
Bachelor's	33.56% (48)	30.43% (56)
Master's	11.18% (16)	4.347% (8)
Professional Degree	0.699% (1)	0.0% (0)
PhD	0.0% (0)	5.434% (10)
Number of online videos watched per month		
0-4	17.48% (25)	17.93% (33)
5-14	30.76% (44)	30.43% (56)
15-30	23.07% (33)	20.65% (38)
31+	28.67% (41)	30.97% (57)
Have you ever uploaded a video before?		
Yes	60.83% (87)	64.67% (119)
No	39.16% (56)	35.32% (65)

Human Success Rates: Manual Selection

(First User Study)

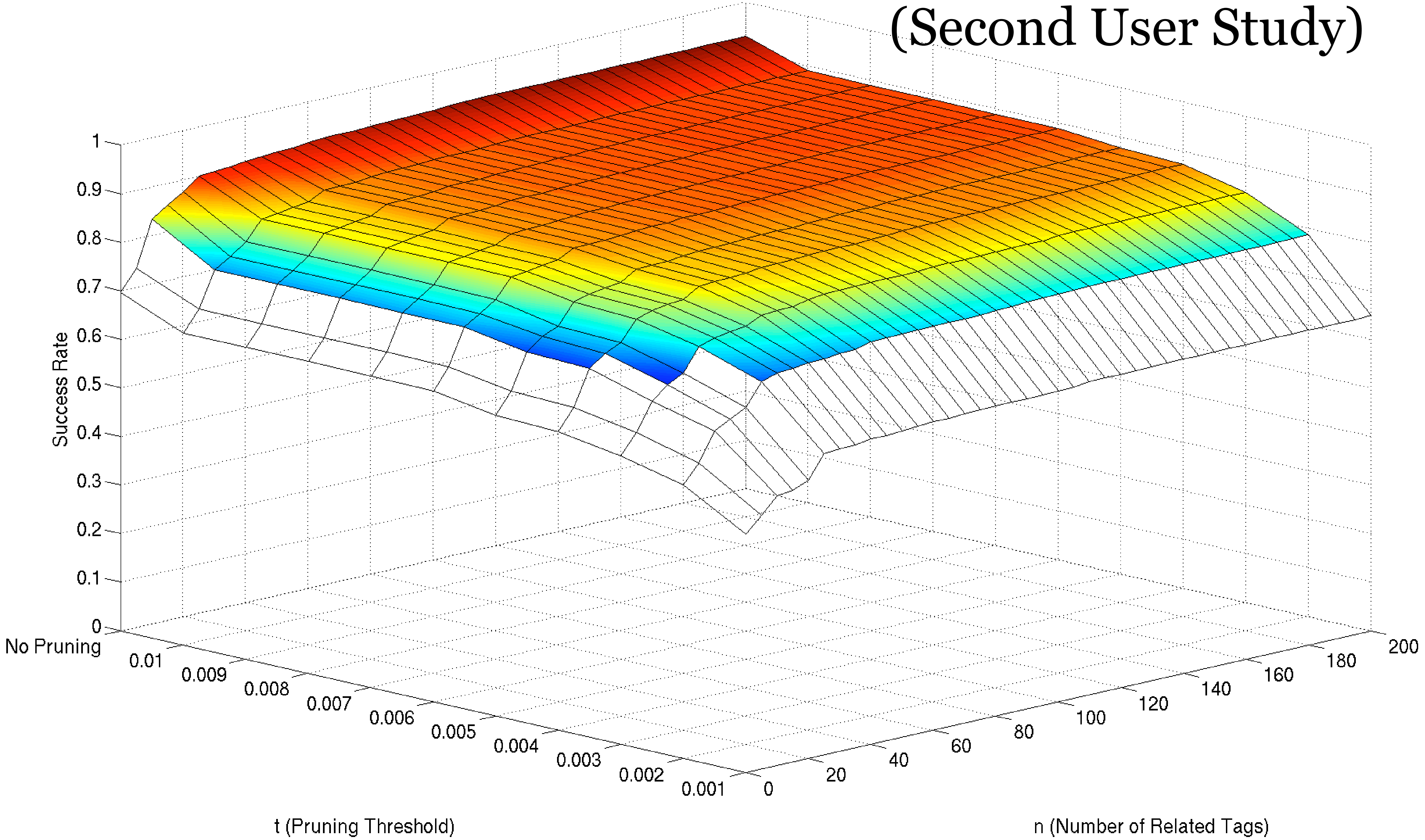


Balancing Usability and Security in a Video CAPTCHA

Kurt Alfred Kluever
Richard Zanibbi

Human Success Rates: Random Walk

(Second User Study)



Balancing Usability and Security in a Video CAPTCHA

Kurt Alfred Kluever
Richard Zanibbi

Completion Times and User Feedback

Completion times (in seconds)

- User Study 1: median = 20.6 ($\mu = 29.7$, $\sigma = 34.7$)
- User Study 2: median = 17.1 ($\mu = 22.0$, $\sigma = 23.6$)

Which task is faster?

- User Study 1: 16%: neither, **64%: text**, 20%: video
- User Study 2: 13%: neither, **60%: text**, 27%: video

Which task is more enjoyable?

- User Study 1: 23%: no pref, 15%: text, **62%: video**
- User Study 2: 22%: no pref, 20%: text, **58%: video**

Comparison with Existing Work

		Success Rates	
CAPTCHA	Type	Human	Machine
Microsoft	Text-based	0.90 [3]	0.60 [28]
Baffletext	Text-based	0.89 [4]	0.25 [4]
Handwritten	Text-based	0.76 [23]	0.13 [23]
ASIRRA	Image-based	0.99 [6]	0.10 [9]
Video	$\tau = \langle 15, 0.003, \mathrm{T}, \mathrm{T} \rangle$	0.77	0.02
	$\tau = \langle 25, 0.006, \mathrm{T}, \mathrm{T} \rangle$	0.86	0.05
	$\tau = \langle 90, 0.006, \mathrm{T}, \mathrm{T} \rangle$	0.90	0.13

Perhaps not a replacement,

but an alternative?

Conclusions

First video-based CAPTCHA and it is:

- *Automated*
- *Open*
- *Usable*
- *Secure*

Usability/security tradeoff

Pass rates are comparable to existing CAPTCHAs

~60% of participants reported that Video CAPTCHAs were more enjoyable than text-based CAPTCHAs

Future Work

Collaborative filtering to improve ground truth tags

- Improve existing tags on poorly labeled videos

Computer vision attacks

- Detect text in video frames, recognize it, submit it

Content-based Video Retrieval attacks

- Look for similar videos in database + submit their tags

Audio analysis attacks

- Extract important words from video + submit them

Further user studies with audio-only or video-only

Thank You

Online Demonstration:

http://sudbury.cs.rit.edu/

Thanks to

Google

xerox

Computer Science @ RIT
B.Thomas Golisano College of Computing & Information Sciences

Questions?

TO COMPLETE YOUR WEB REGISTRATION, PLEASE PROVE THAT YOU'RE HUMAN:
WHEN LITTLEFOOT'S MOTHER DIED IN THE ORIGINAL 'LAND BEFORE TIME,' DID YOU FEEL SAD?
○ YES
○ NO
(BOTS: NO LYING)

Image Credit: xkcd.com