A Simple Generic Attack on Text Captchas

Haichang Gao (Xidian University), Jeff Yan (Lancaster University), Fang Cao, Zhengya Zhang, Lei Lei, Mengyun Tang, Ping Zhang, Xin Zhou, Xuqin Wang and Jiawei Li (Xidian University)

Introduction

CAPTCHA:

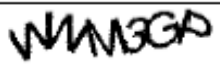
Completely Automated Public Turing Test to Tell Computers and Humans Apart .

Gabor Filters

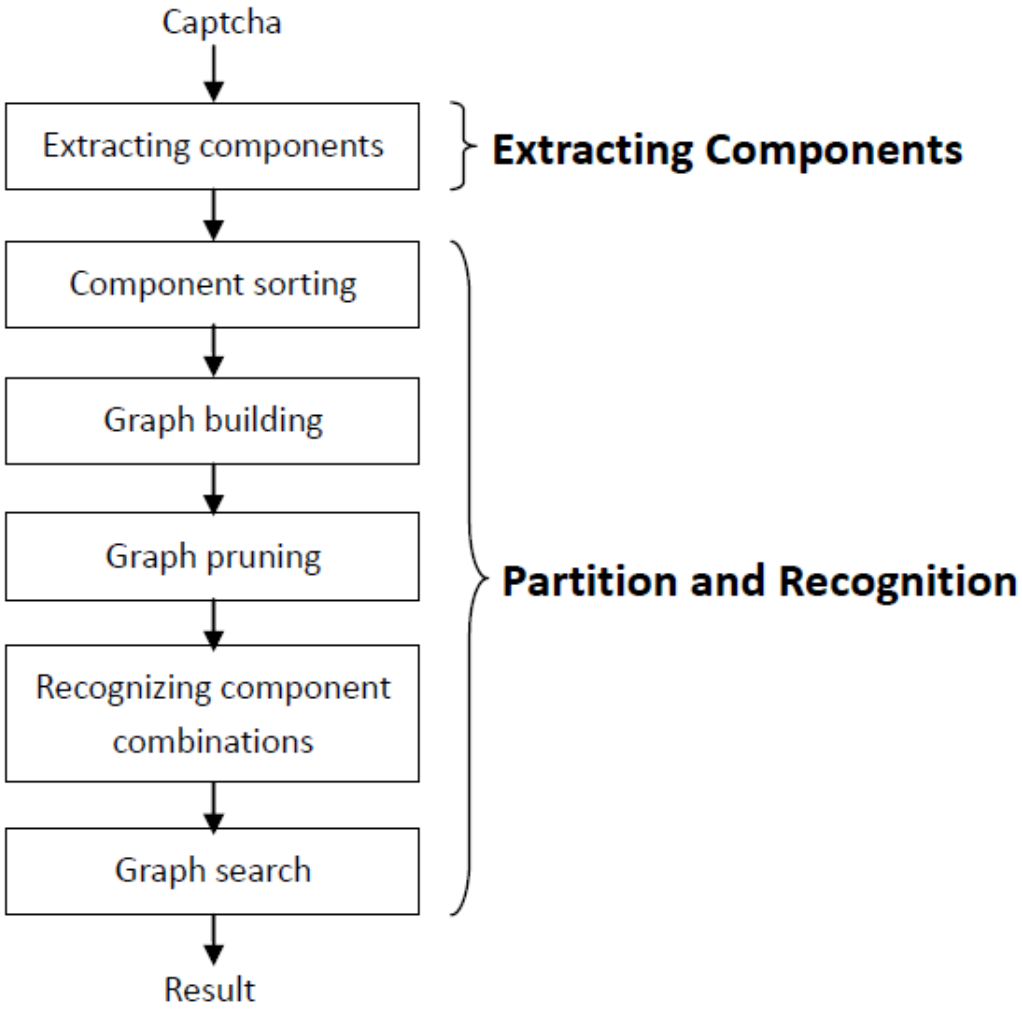
Dennis Gabor laid their theoretical foundations in 1946.

The temporal (1-D) Gabor filters
The Spatial (2-D) Gabor filters
Log-Gabor filters
2D Log-Gabor filters

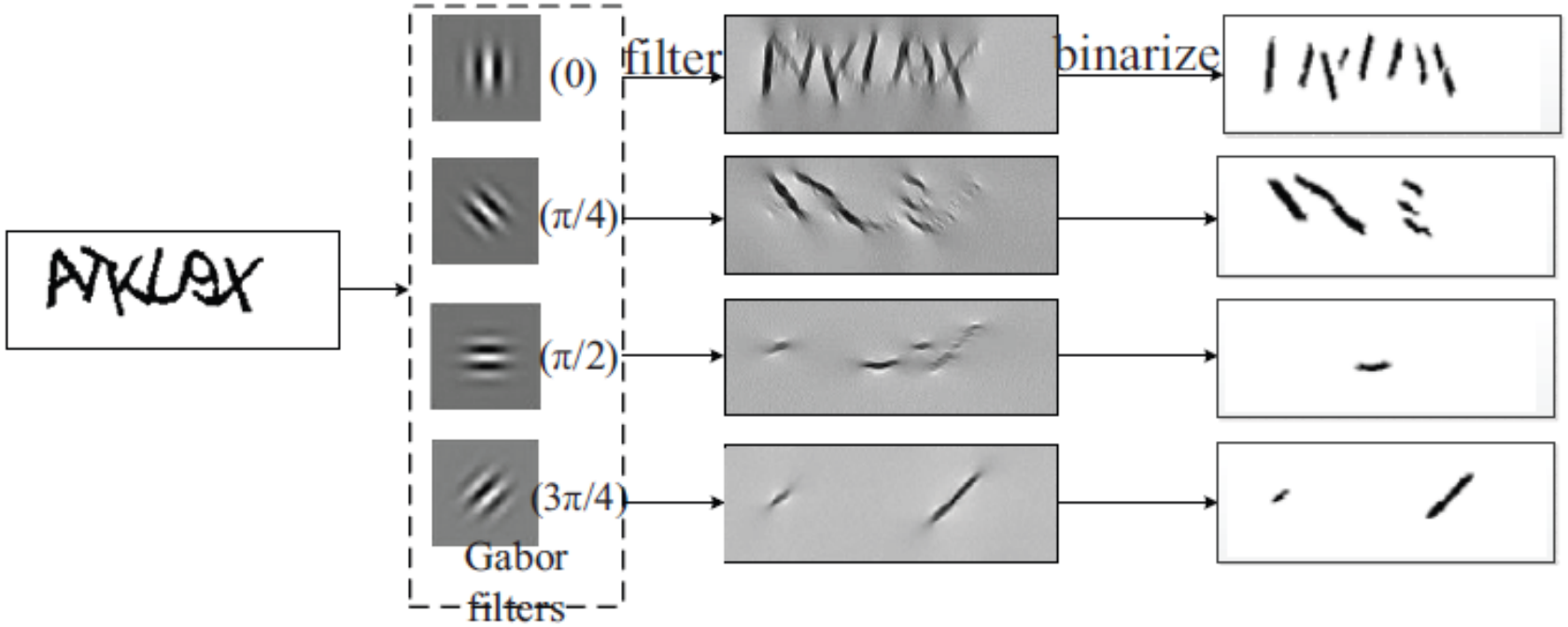
Real World Popular Captchas

Scheme	Website	Sample Captcha	Characteristics
reCAPTCHA	google, facebook, youtube, linkedin, twitter, blogspot, wordpress, google.co.in		CCT scheme, only control word tested, only digits used, rotation used, varied font size, varied Captcha length
Yahoo!	yahoo.com, yahoo.co.jp		hollow scheme, varied fonts, rotation and distortion used, varied Captcha length
Baidu	baidu.com hao123.com		CCT scheme, rotation used
Wikipedia	wikipedia.org		Character isolated scheme, varied Captcha length, no digits used
QQ	qq.com		Hollow scheme, rotation used, overlap used, varied font size
Microsoft	live.com bing.com		Character isolated scheme, varied Captcha length, varied font size, rotation used
Amazon	amazon.com		CCT scheme, constant font, rotation used
Taobao	taobao.com		CCT scheme, rotation used, large alphabet set
Sina	sina.com.cn		CCT scheme, background clutter, noise arcs used
Ebay	ebay.com		CCT scheme, varied font size, rotation used

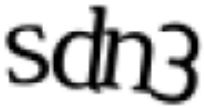
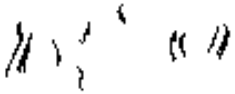
Our Attack


Captcha
Extracting components
Extracting Components
Component sorting
Graph building
Graph pruning
Recognizing component combinations
Graph search
Partition and Recognition
Result

Our Attack : *Extracting Components*

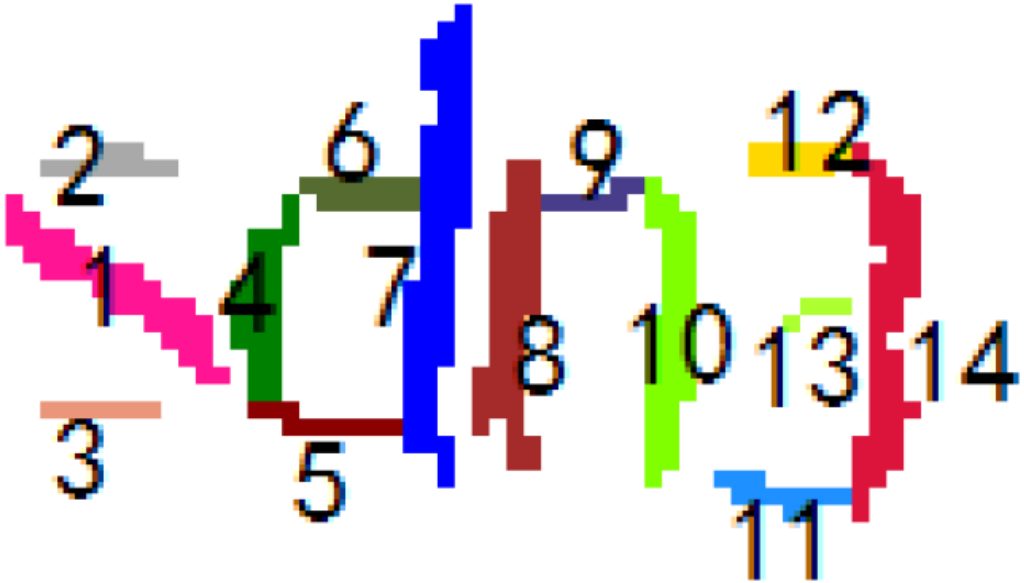


Our Attack : *Extracting Components*

	Microsoft	QQ	Baidu
Angle			
0			
$\pi/4$			
$\pi/2$			
$3\pi/4$			
+			

Our Attack : *Partition and Recognition*

Step 1. Component sorting



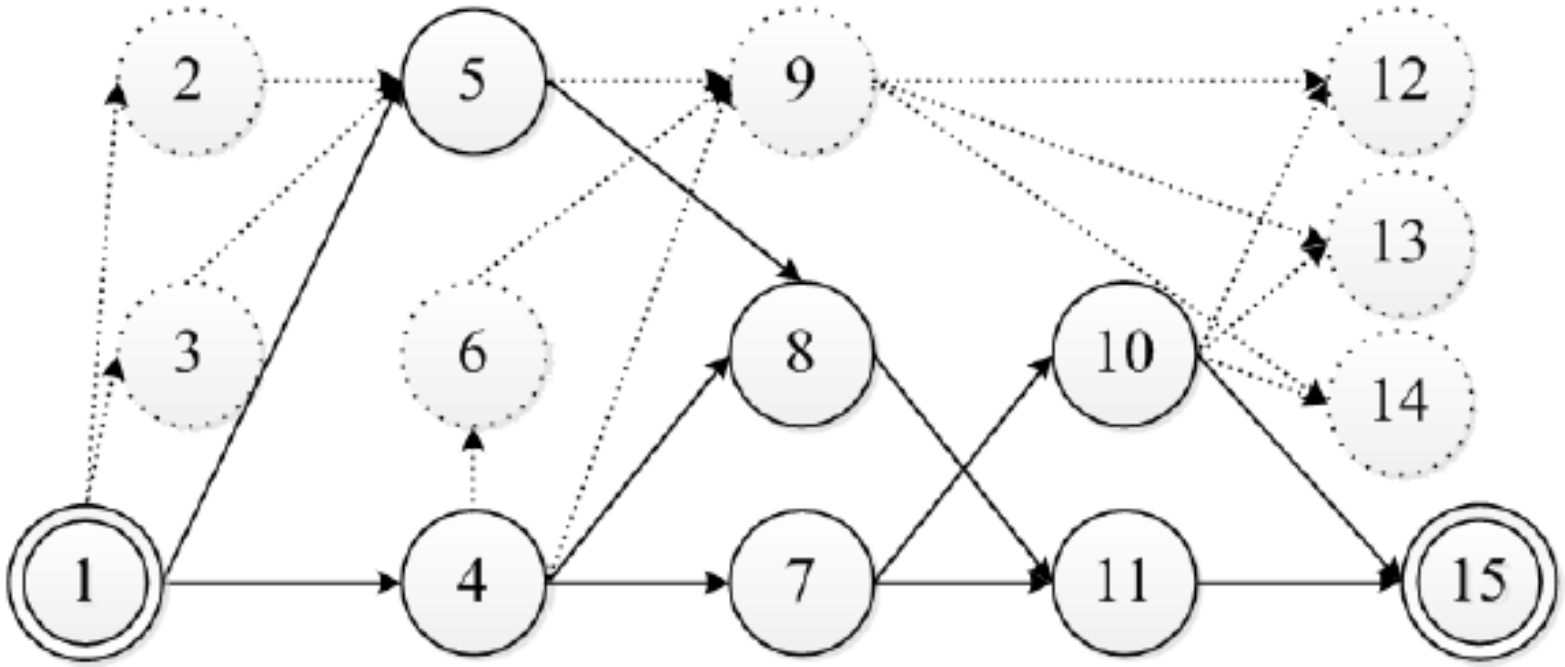
Our Attack : *Partition and Recognition*

Step 2. Graph building

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	●	●	●	●										
2				●										
3				●										
4					●	●	●	●						
5							●	●						
6								●						
7									●	●				
8										●				
9											●	●	●	
10											●	●	●	●
11														●
12...14														

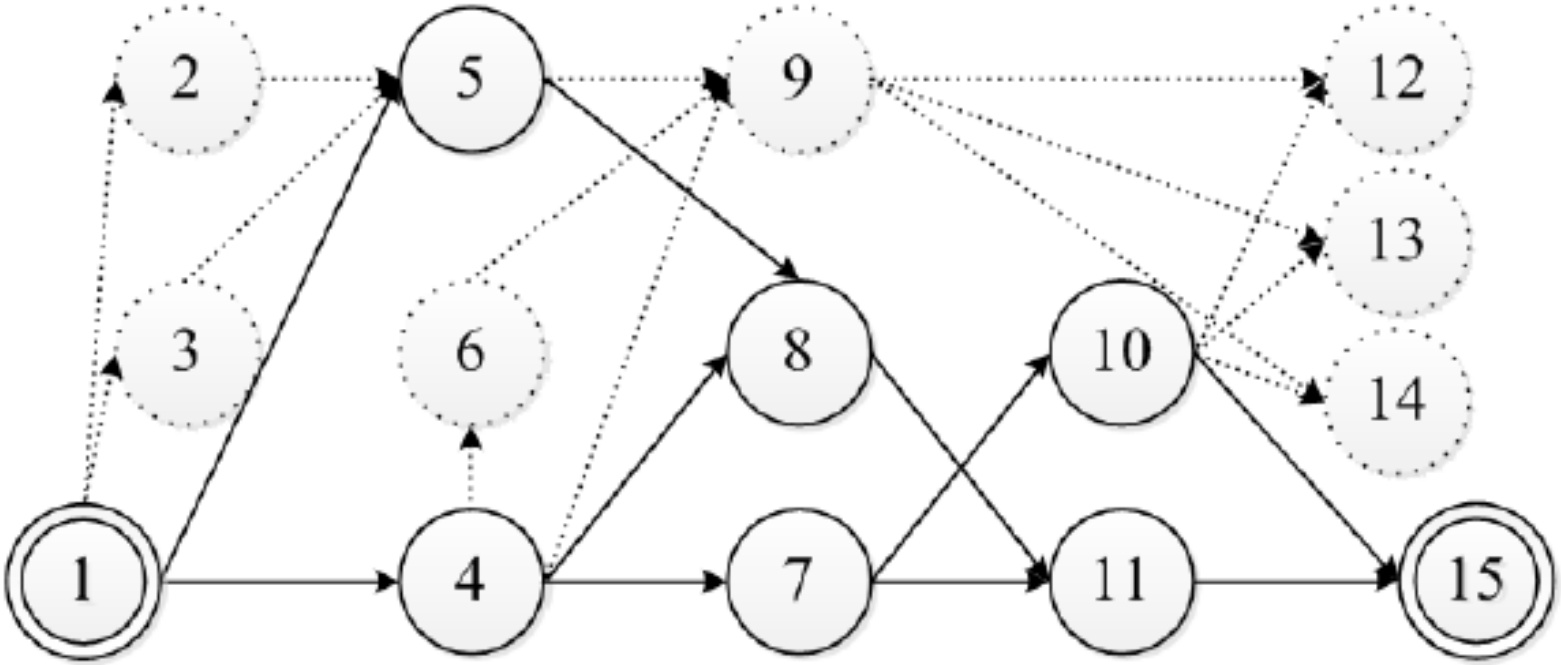
Our Attack : *Partition and Recognition*

Step 2. Graph building



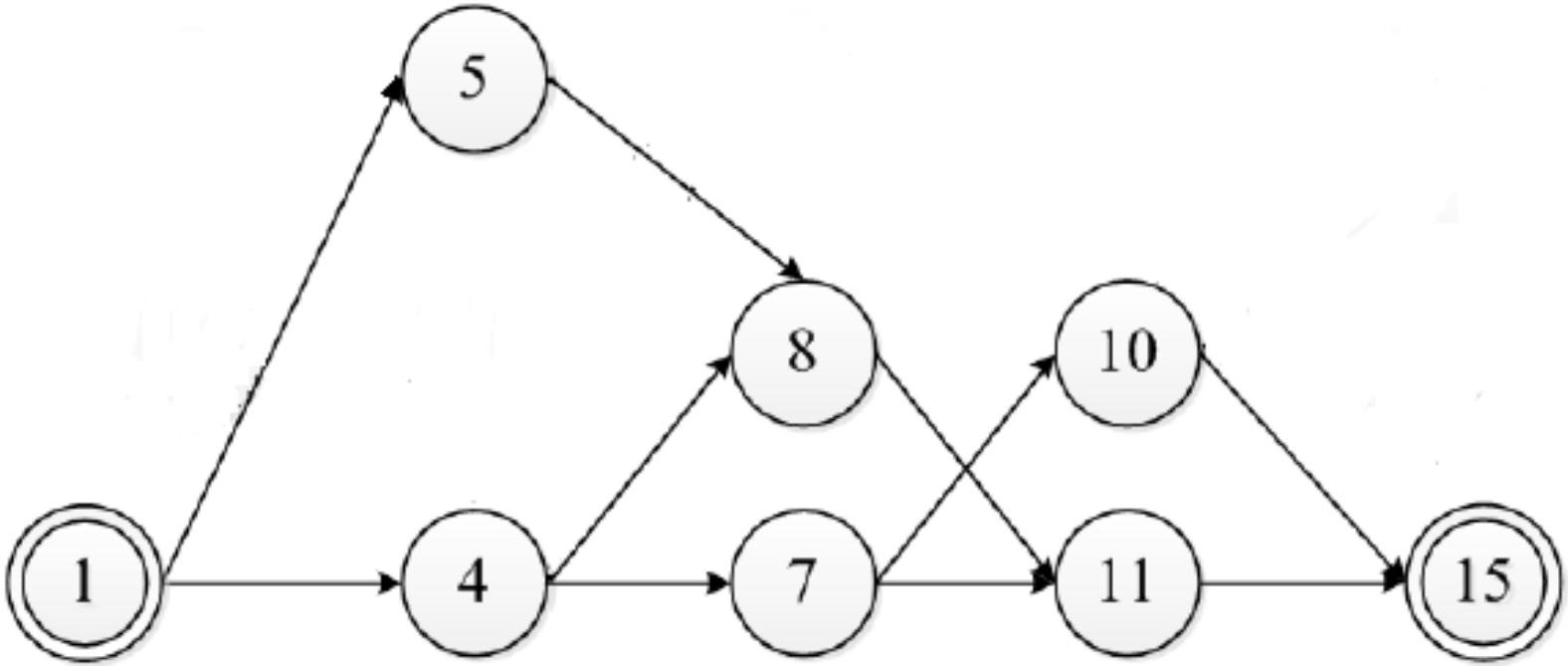
Our Attack : *Partition and Recognition*

Step 3. Graph pruning



Our Attack : *Partition and Recognition*

Step 3. Graph pruning



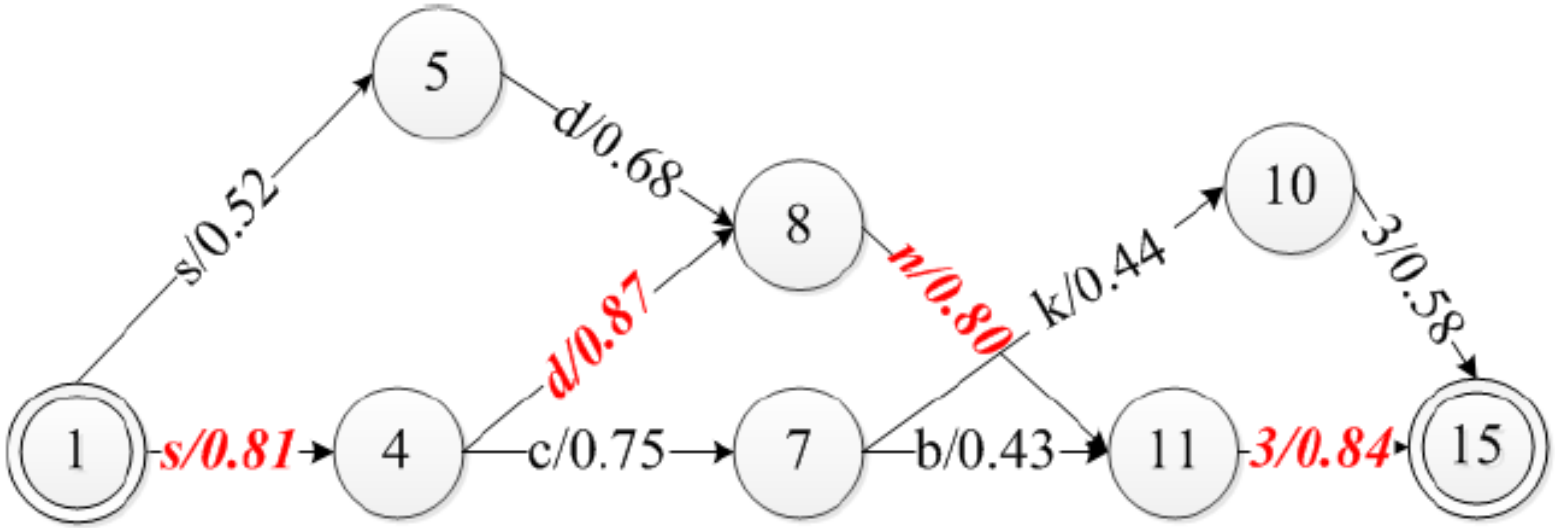
Our Attack : *Partition and Recognition*

Step 4. Recognizing component combinations

	1	2	3	4	5	6	7	8	9	10	11...13	14
1			***s/0.81***	s/0.52								
2,3												
4						c/0.75	***d/0.87***					
5							d/0.68					
6												
7									k/0.44	b/0.43		
8										***n/0.80***		
9												
10												3/0.58
11												***3/0.84***
12...14												

Our Attack : *Partition and Recognition*

Step 4. Recognizing component combinations



Our Attack : *Partition and Recognition*

Step 5. Graph search

j	$step[j]$	Path	$value[j]$	$result[j]$
4	1	1→4	0.81	s
5	1	1→5	0.52	s
7	2	1→4→7	1.56	sc
8	2	1→4→8	1.68	sd
10	3	1→4→7→10	2.00	sck
11	3	1→4→8→11	2.48	sdn
15	***4***	***1→4→8→11→15***	***3.32***	***sdn3***

Evaluation : *Attack Results*

Scheme	Success rate	Speed(s)
reCAPTCHA	77.2%	10.27
Yahoo!	5.0%	28.56
Baidu	44.2%	2.81
Wikipedia	23.8%	3.74
QQ	56.0%	4.95
Microsoft	16.2%	12.59
Amazon	25.8%	13.18
Taobao	23.4%	4.64
Sina	9.4%	4.83
Ebay	58.8%	5.98

Evaluation : *Further Applicability Test*

Scheme	Original image	Reconstruction	Success rate	Speed (s)
Early reCAPTCHA			7.8%	8.06
Yandex			2.2%	15.5

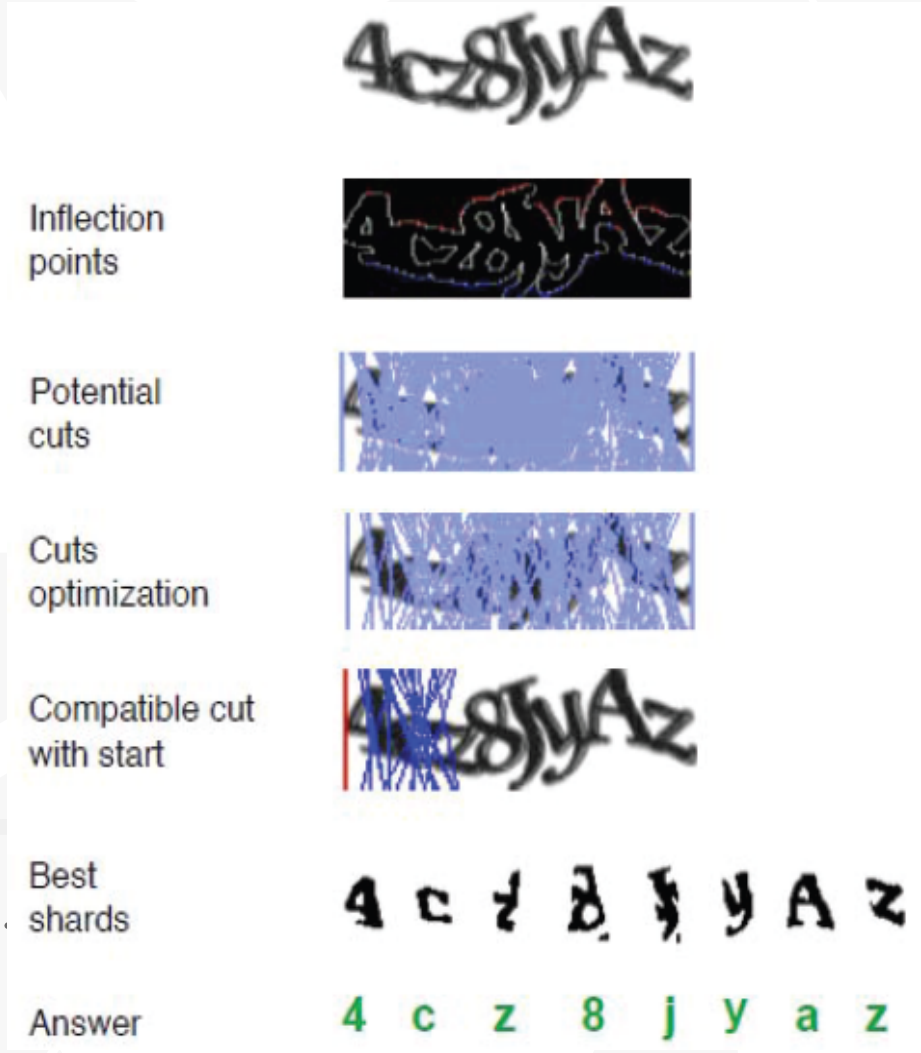
Evaluation : *A Comparison with Prior Art*

pixel counting, histogram analysis, CFS

Decaptcha: preprocessing, segmentation, post-segmentation, recognition, and post-preprocessing

CCS'13: The Robustness of Hollow Captchas.

WOOT' 14:The End is Nigh: Generic Solving of Text-based Captchas.



Design Choices : *Graph Search Algorithms*

Scheme	Average attack speed (Seconds)		
	DP search	Integer partition algorithm	DFS search
reCAPTCHA	10.27	10.31	10.87
Yahoo!	28.56	33.33	34.32
Baidu	2.81	3.00	3.14
Wikipedia	3.74	3.78	3.83
QQ	4.95	5.15	5.55
Microsoft	12.59	14.93	15.49
Amazon	13.18	14.60	15.28
Taobao	4.64	4.74	4.80
Sina	4.83	4.93	5.03
Ebay	5.98	6.01	6.06

Design Choices : *Extraction Orientations*

3 orientations: $0, \pi/3, 2\pi/3$;

4 orientations: $0, \pi/4, 2\pi/4, 3\pi/4$;

6 orientations: $0, \pi/6, 2\pi/6, 3\pi/6, 4\pi/6, 5\pi/6$;

8 orientations: $0, \pi/8, 2\pi/8, 3\pi/8, 4\pi/8, 5\pi/8, 6\pi/8, 7\pi/8$.

(a) 3 orientations

(b) 4 orientations

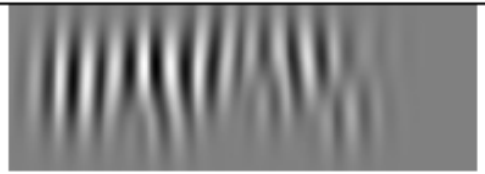
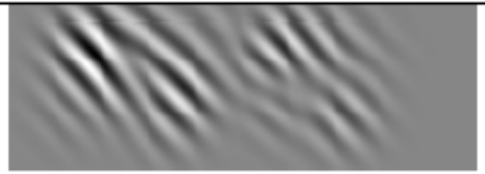
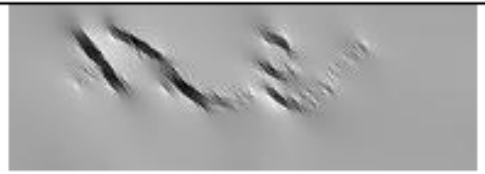
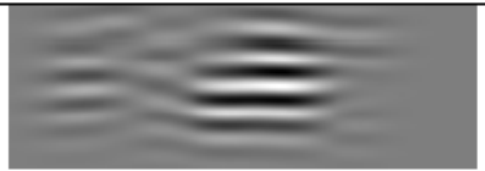
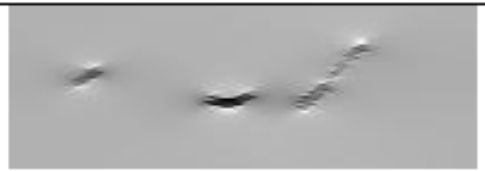
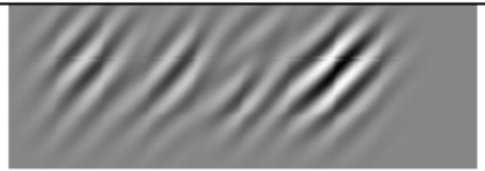
(c) 6 orientations

(d) 8 orientations

Design Choices : *Extraction Orientations*

Orientations	Success rate	Average attack speed (Seconds)
3	20.8%	12.25
4	25.8%	14.32
6	9.2%	21.55
8	7.4%	30.01

Design Choices : *Extracting Methods*

	2D Gabor	Steerable filter	Log-Gabor
0			
$\pi/4$			
$\pi/2$			
$3\pi/4$			

Design Choices : *Classifiers*

Schemes	Success rate		Speed(s)	
	KNN	CNN	KNN	CNN
reCAPTCHA	77.2%	38.4%	10.27	10.19
Yahoo!	5.0%	5.2%	28.56	23.81
Baidu	44.2%	46.6%	2.81	2.21
Wikipedia	23.8%	20.4%	3.74	2.90
QQ	56.0%	22.4%	4.95	4.61
Microsoft	16.2%	8.6%	12.59	6.64
Amazon	25.8%	20.2%	13.18	8.68
Taobao	23.4%	20.4%	4.64	5.25
Sina	9.4%	4.4%	4.83	5.21
Ebay	58.8%	32.6%	5.98	5.50

Defence

Experiments	Sample Image	Reconstruction Image	Overlapping	Rotating	Warping	Attack Success
1			✓			11.6%
2				✓		13%
3					✓	8.8%
4			✓	✓		7.6%
5			✓		✓	7.4%
6				✓	✓	6.8%
7			✓	✓	✓	1.4%

Summary and Conclusion

Thank you !