Information Hiding, 2008

Residual Information of Redacted Images Hidden in the Compression Artifacts

Nicholas Zhong-Yang Ho Ee-Chien Chang

School of Computing
National University of Singapore

Background

- Many images needed to be redacted before they are released to the public.

examples from WWW

NEW HAMPSHIRE STATE POLICE
CONTINUATION OF INVESTIGATION REPORT

1. CASE NO.	2. INVESTIGATING TROOPER	3. I.D. NO.	4. TOWN OF CRIME	5. TN. CD.	6. DATE OF REPORT
MC07-4121	Sergeant Charles West	526	Franconia	*	May 11, 2007

[redacted] OK, don't go any clo-, closer to the situation, we do have multiple officers en route

[redacted] OK

[redacted] Um, so the officer is down and the, the bald headed fellow is still up

[redacted] He's still standing, he has a, has the pistol in his hand, pointing it at someone, I can't see um, the officer was in pursuit of this car uh when they came up the road, he stopped the car and then uh, they shot the guy (inaudible)

Dispatcher Henson	OK, we have multiple officers there now, sir, we'll uh, give us a call if anything escalates or changes
Chester Thompson	OK
Dispatcher Henson	OK, thank you
Female Voice	Hey sir
Lincoln Dispatch	Yah
Dispatcher Henson	Dispatcher Henson
Lincoln Dispatch	Hi there, it's Lincoln Dispatch
Dispatcher Henson	Hi
Lincoln Dispatch	Anything we can do to help you
Dispatcher Henson	Uh, we got everybody headed, I guess Bruce McKay's been shot
Lincoln Dispatch	OK
Dispatcher Henson	Um, we got other officers signing off there now
Lincoln Dispatch	Ok
Dispatcher Henson	We're gonna get a, actually if you wanted to head your Lincoln Ambulance that way, um, I'll get the Woodsville one going this way

	Page 8 of 21	SIGNED Chas West	DATE 06-20-07

DSSP 102 (Rev. 08/94)

928

United Nations Nations Unies
INTEROFFICE MEMORANDUM MEMORANDUM INTERIEUR

TO: Mr. Bruno Henn, Officer-in-Charge
A: Safety and Security Service
Office of Central Support Services

DATE: 16 March 2005

REFERENCE:

THROUGH:
S/C DE:

FROM: Ahmad Fawzi, Director
DE: News and Media Division
Department of Public Information

SUBJECT: **Revocation of Media Accreditation/Grounds Pass – Ms. [redacted]**
OBJET:

An investigation into the letters of assignment presented by Ms. [redacted] to the Media Accreditation and Liaison Unit (MALU) in order to obtain United Nations accreditation has disclosed that Ms. [redacted] has been submitting forged letters under the name of the Editor and Managing Director of an Indian newspaper.

When contacted to confirm Ms. [redacted]'s association with his news organization, the Editor of The Pioneer, Mr. Chandan Mitra, stated categorically in writing - by fax and email - that Ms. [redacted] had "never been engaged by The Pioneer for any purpose"; that his signature that is purported to be affixed to the letters of assignment was "an outright forgery"; that the letterhead used by Ms. [redacted] appeared to be "computer generated". He added: "Clearly, Ms. [redacted] obtained accreditation to the UN by fraud", he requested the UN to "cancel her press credentials immediately", and said he was sorry to find that this fraud had been going on since 2003. (Copy of Mr. Mitra's letter, as well as copies of Ms. [redacted]'s forgeries are attached for ease of reference.)

Because of the serious nature of the fraud perpetrated by Ms. [redacted] for the past three years, I hereby revoke her accreditation, and request that she be made permanently and irrevocably ineligible for any type of United Nations pass. As early as possible, MALU will arrange for a temporary pass for Ms. [redacted], will accompany her to Room S-301 to clear out her desk, and will then escort her from United Nations premises.

In light of this grave breach of the rules and regulations governing UN media accreditation, I would greatly appreciate your help in the Safety and Security Service in implementing this decision.

Thanks, as always, for your kind cooperation.

cc: Mr. Shashi Tharoor
Ms. Shirley Brownell
Mr. Nikolay Bolshakov

examples from WWW

Type of redaction studied in this talk.

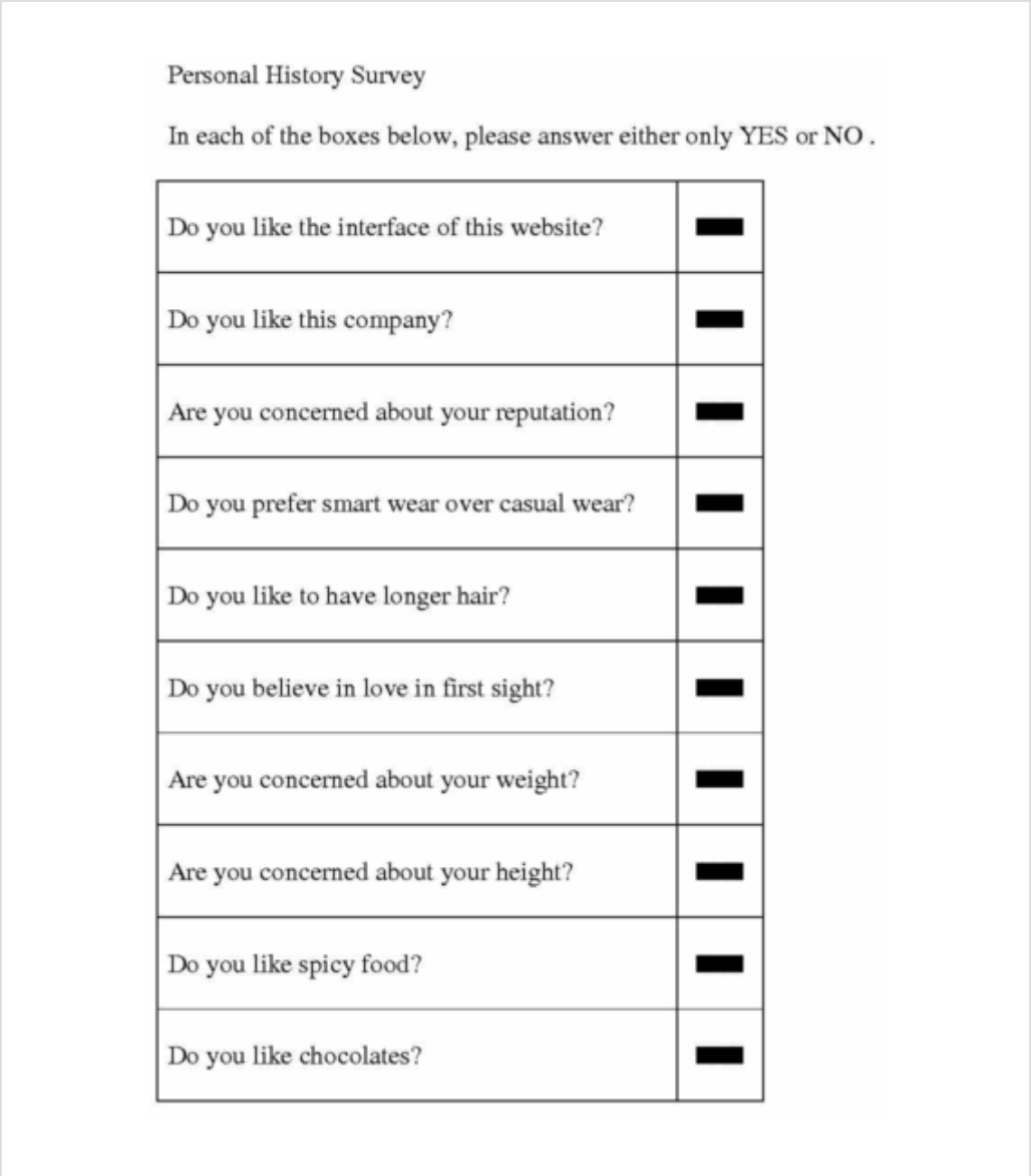
Personal History Survey

In each of the boxes below, please answer either only YES or NO .

Do you like the interface of this website?	NO
Do you like this company?	NO
Are you concerned about your reputation?	NO
Do you prefer smart wear over casual wear?	NO
Do you like to have longer hair?	NO
Do you believe in love in first sight?	YES
Are you concerned about your weight?	YES
Are you concerned about your height?	YES
Do you like spicy food?	YES
Do you like chocolates?	YES

constructed example

Type of redaction studied in this talk.



Personal History Survey

In each of the boxes below, please answer either only YES or NO .

Do you like the interface of this website?	[redacted]
Do you like this company?	[redacted]
Are you concerned about your reputation?	[redacted]
Do you prefer smart wear over casual wear?	[redacted]
Do you like to have longer hair?	[redacted]
Do you believe in love in first sight?	[redacted]
Are you concerned about your weight?	[redacted]
Are you concerned about your height?	[redacted]
Do you like spicy food?	[redacted]
Do you like chocolates?	[redacted]

constructed example

Pixels in the sensitive region are replaced by black/white pixels

Goal:

How effective is digital redaction?

- Under certain conditions, we still can extract information from the surrounding pixels.

Main Observation

- Images are lossily-compressed or processed before redaction. Information in the sensitive region may has *spread* to the non-sensitive region before redaction.

  Hence, replacement of pixels values in the sensitive region does not *completely* purge the sensitive information.

Compression Artifacts

JPEG image

Compression artifacts

JPEG image

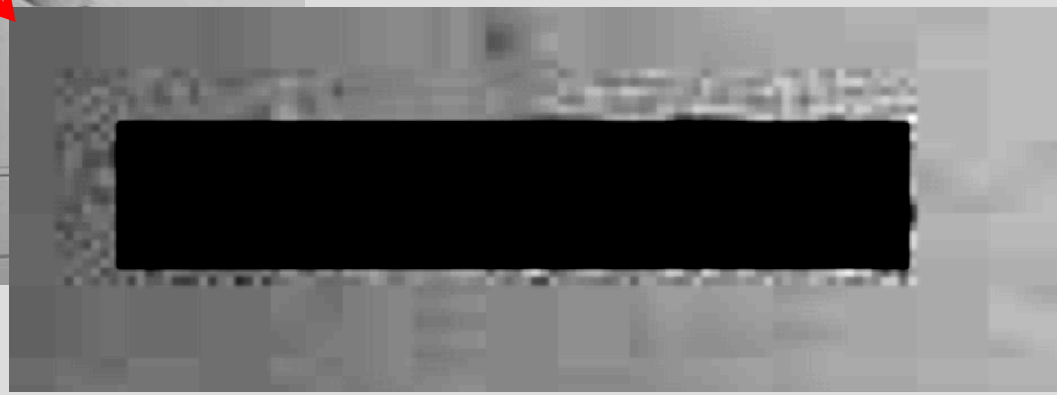


Image enhanced to illustrated the artifacts

Other types of redaction

- Physical redaction

  overwritten with marker.

  cover with tape while scanning.

  cutting out the region.

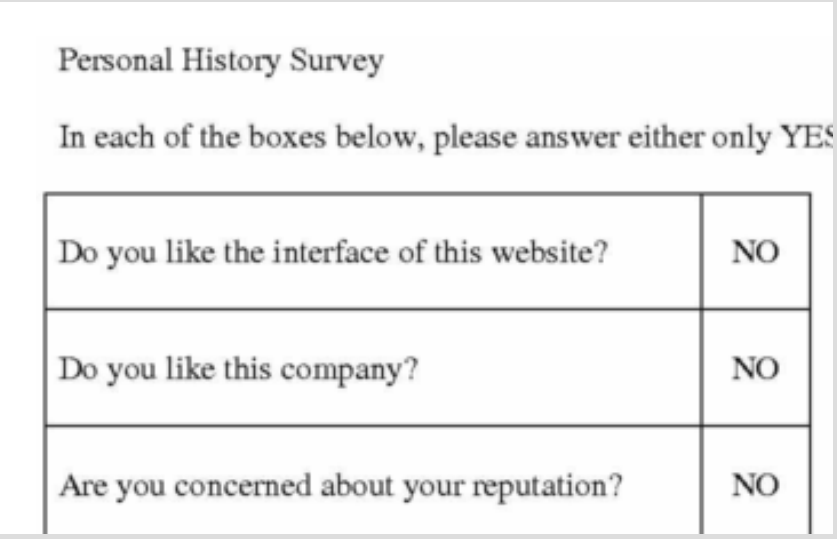
- Redaction of non-pixel representation.

  redaction of pdf file.

- Information derived from content.

  for e.g. length of words covered.

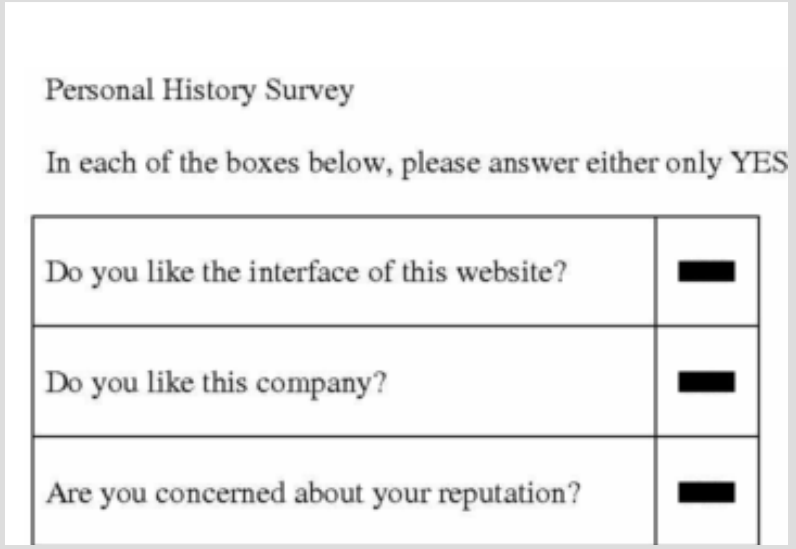
- We are concern with digital redaction.
- Derive information from image processing artifacts.



Personal History Survey

In each of the boxes below, please answer either only YES

Do you like the interface of this website?	NO
Do you like this company?	NO
Are you concerned about your reputation?	NO



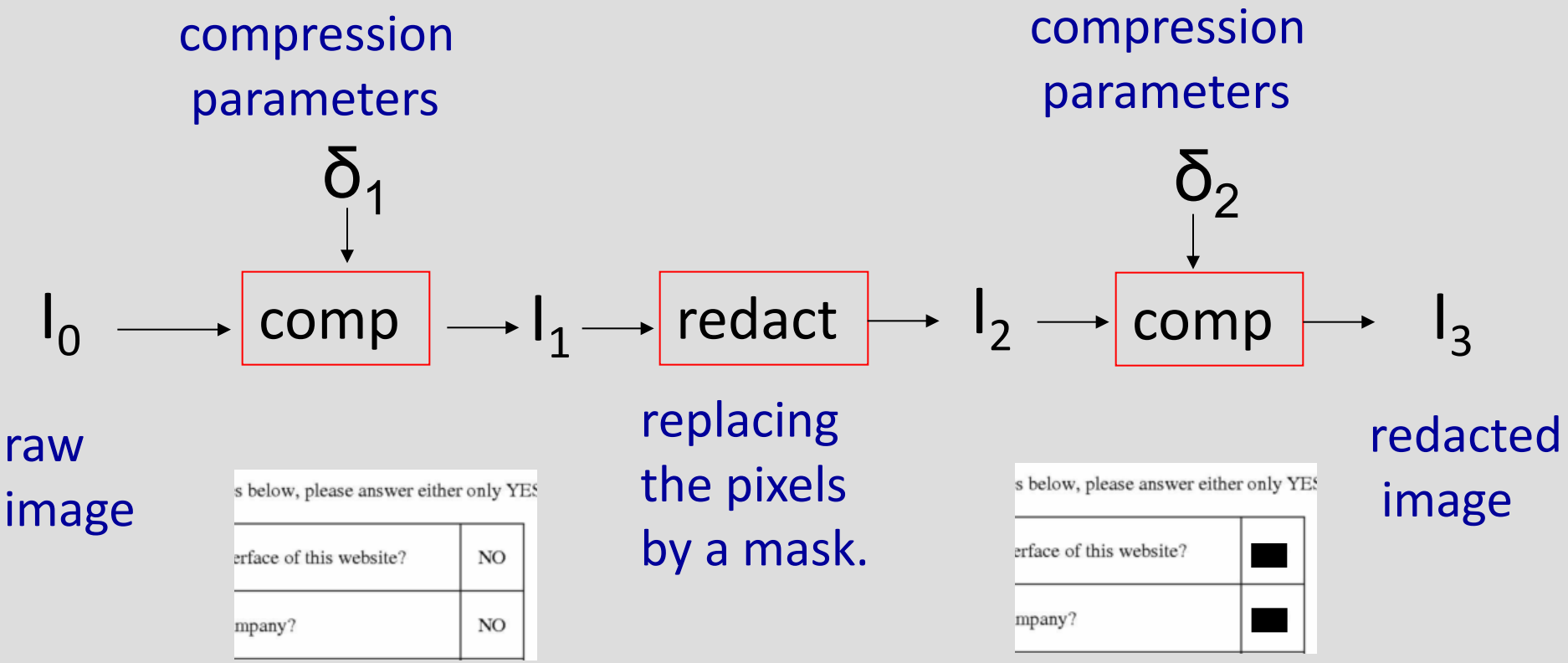
Personal History Survey

In each of the boxes below, please answer either only YES

Do you like the interface of this website?	[redacted]
Do you like this company?	[redacted]
Are you concerned about your reputation?	[redacted]

I. Formulation: Redaction

- A redacted image has been compressed at least twice.



From I_3, ***actual*** δ_2 can be obtained, and an ***estimate*** of δ_1 also can be obtained

Formulation: adversary's goal

- Given a redacted image I, where region containing a secret is removed.

  An adversary has two templates T_0, T_1 derived from two possible values of the secret 0,1.

  The adversary wants to guess which template is the original. If the chance of correct guess is

  $0.5 + \varepsilon$, then ε is the advantage of the adversary.

- If adversary achieve non-zero advantage, the redacted image must has leaked some information of the secret.

Redacted image I_3

Personal History Survey

In each of the boxes below, please answer either only YES or NO .

Do you like the interface of this website?	[redacted]
Do you like this company?	[redacted]
Are you concerned about your reputation?	[redacted]
Do you prefer smart wear over casual wear?	[redacted]
Do you like to have longer hair?	[redacted]
Do you believe in love in first sight?	[redacted]
Are you concerned about your weight?	[redacted]
Are you concerned about your height?	[redacted]
Do you like spicy food?	[redacted]
Do you like chocolates?	[redacted]

Templates

T_0

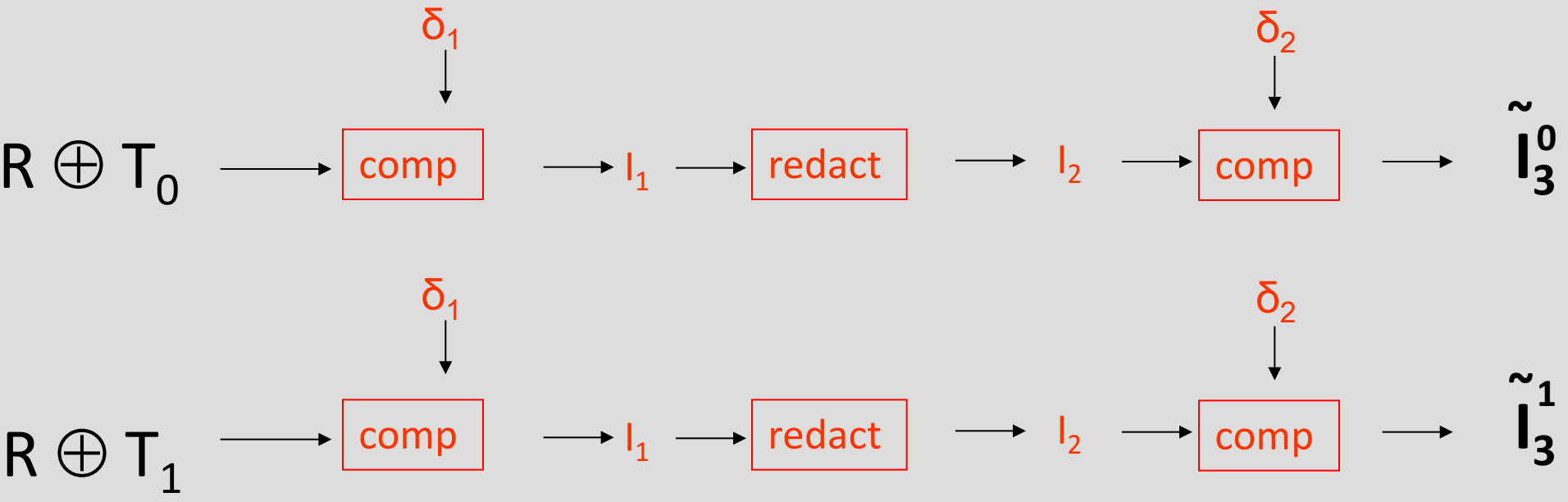
NO

T_1

YES

II. Method 1: Estimate the Raw

- Suppose a good estimate, R, of the raw image in the non-sensitive region is available, then candidates of the whole raw image can be constructed.
- Simulate the redaction process and compare the outcomes.



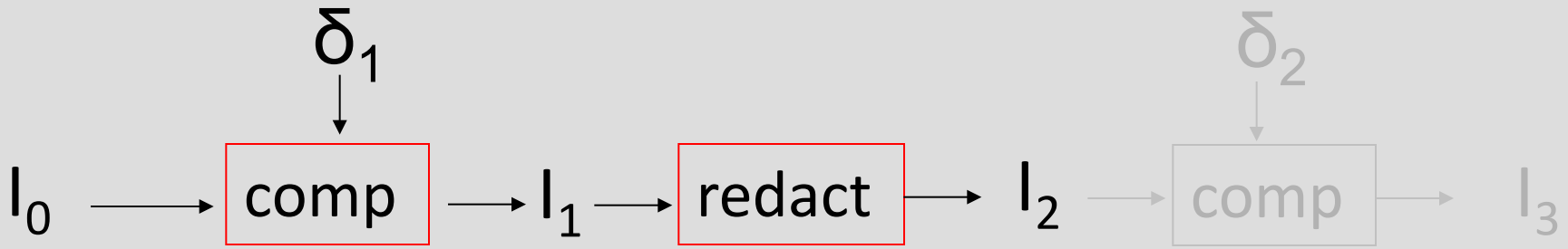
Compare distant of the actual redacted image I_3 with $\tilde{I}_3^0$ and $\tilde{I}_3^1$ respectively

- Suitable for JPEG.

- Difficult to apply to Wavelet-based compression schemes.

Method 2: Quantization error

- Ignore effect of the 2nd compression (treat it as noise).
- Has an estimate of the raw image in the sensitive region (the 2 templates).
- Simulate the first compression to get an estimate of the compressed sensitive region.
- Obtain an estimate of I_1. (the compressed original)
- I_1 should follow the statistics of images compressed with δ_1 (quantization error).



III. Noise and parameters

- δ_1: Estimation of the 1st compression parameter.
- T_0, T_1: Estimation of raw image in the sensitive region (templates)
- R: Estimation of the raw image in non-sensitive region
- Size of redacted region.
- Compression schemes and rates.

IV. Experiments

- Two compression schemes:

  JPEG:

  Quantization matrix

  Wavelet-based compression:

  CDF 9/11wavelet, and uniform quantization.

Data sets

- Random Images.
- 2 images: Document + Photo.

1034x1494 pixels

redacted region
70x28

Nokia 6125 mobile phone
640x480
“normal”
compression quality

template derive from photo captured by digital cameras.

Personal History Survey

In each of the boxes below, please answer either only YES or NO .

Do you like the interface of this website?	NO
Do you like this company?	NO
Are you concerned about your reputation?	NO
Do you prefer smart wear over casual wear?	NO
Do you like to have longer hair?	NO
Do you believe in love in first sight?	YES
Are you concerned about your weight?	YES
Are you concerned about your height?	YES
Do you like spicy food?	YES
Do you like chocolates?	YES

10-335

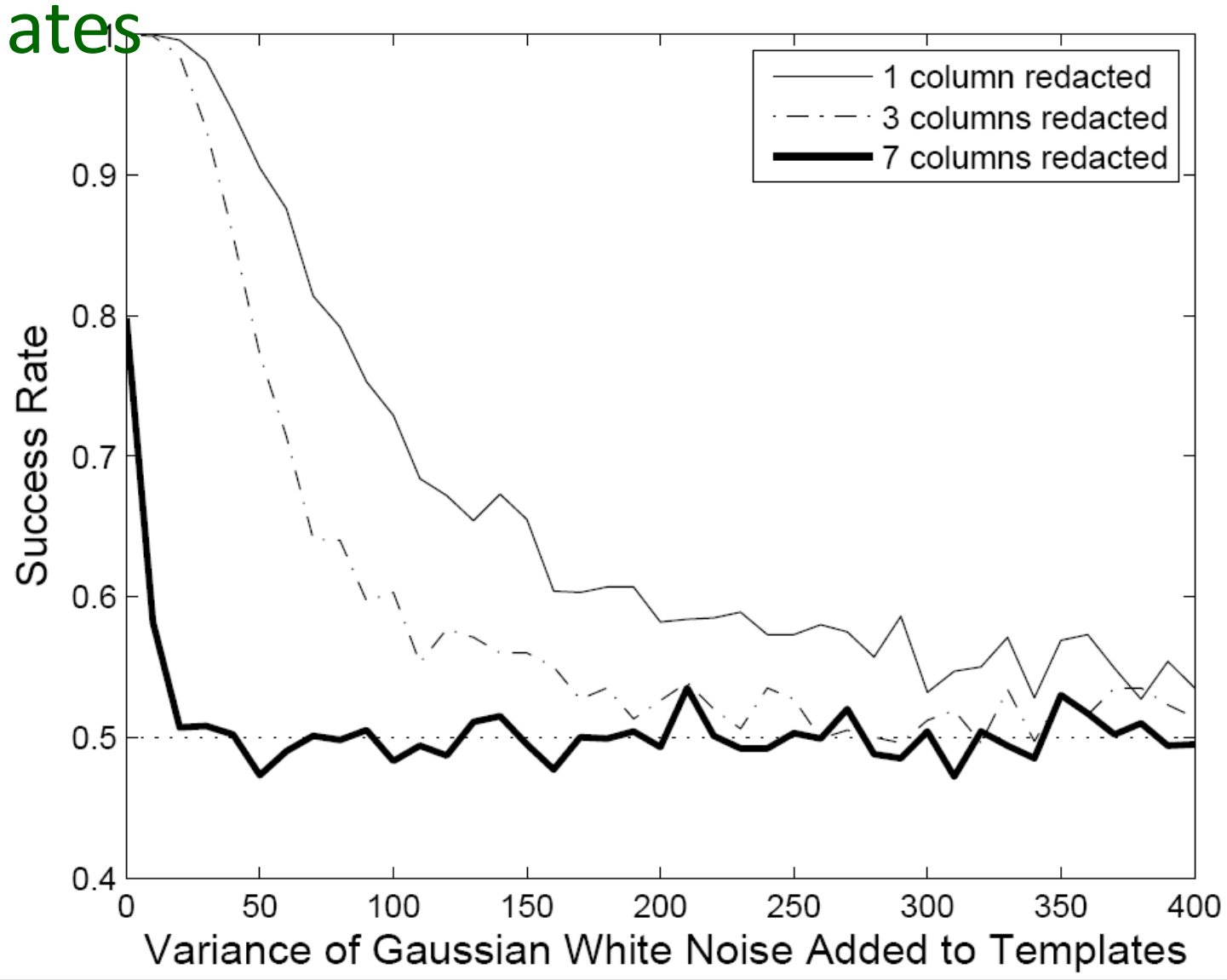
10-339

08-331

11-335

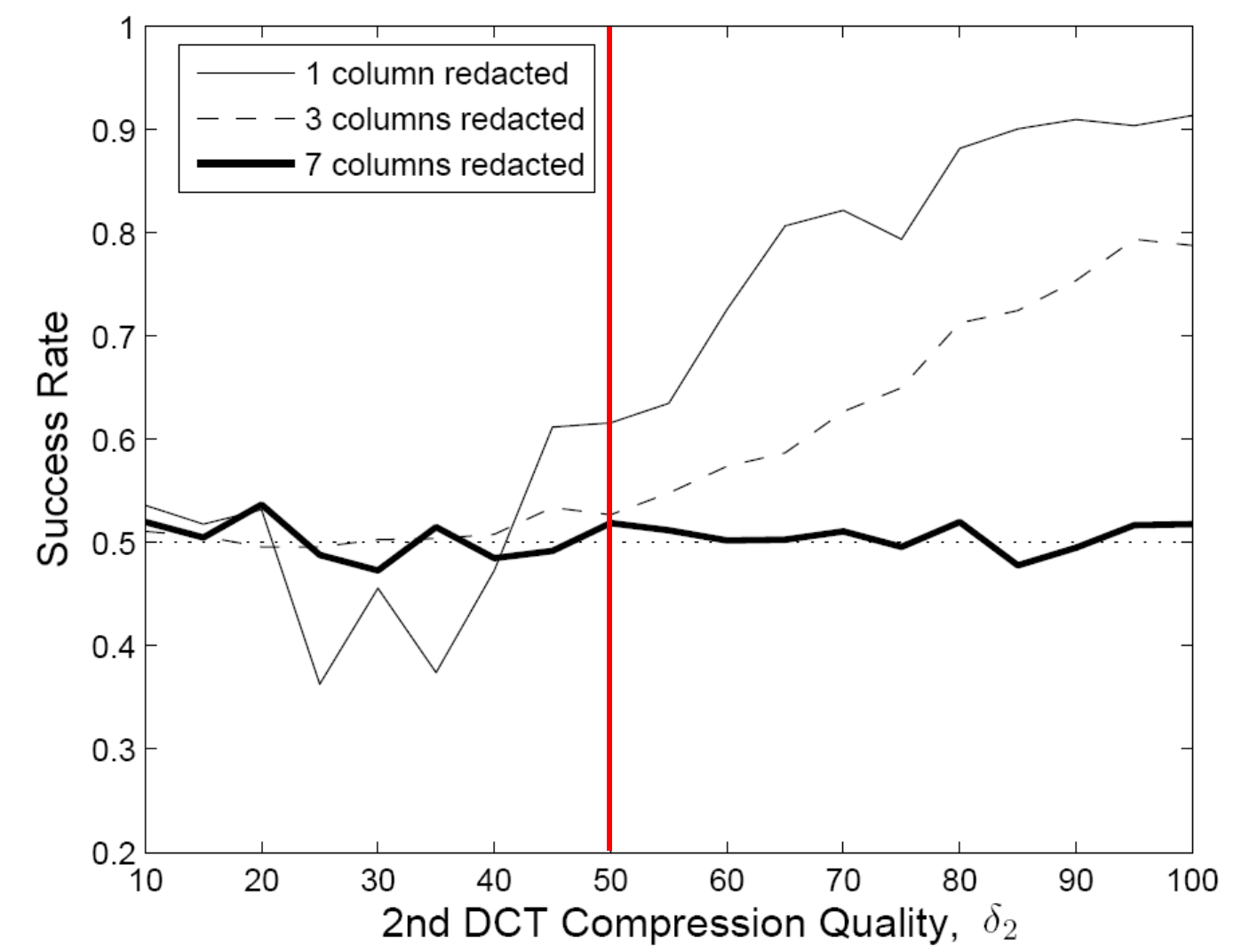
11-339

Effect of redacted region + noise on templates



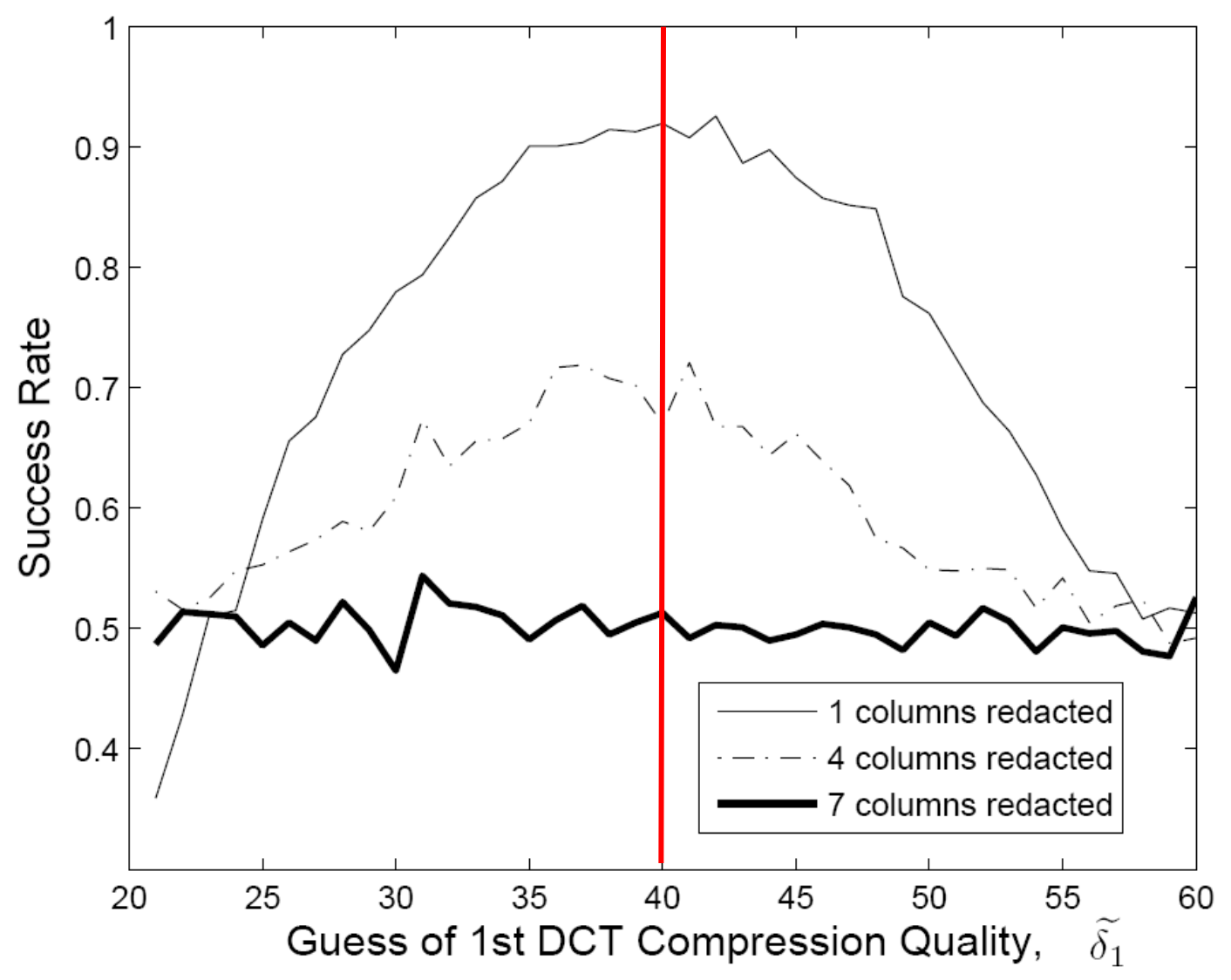
Random images, JPEG, method 2, δ_1 =50, δ_2 = 95.

The1st and 2nd compression



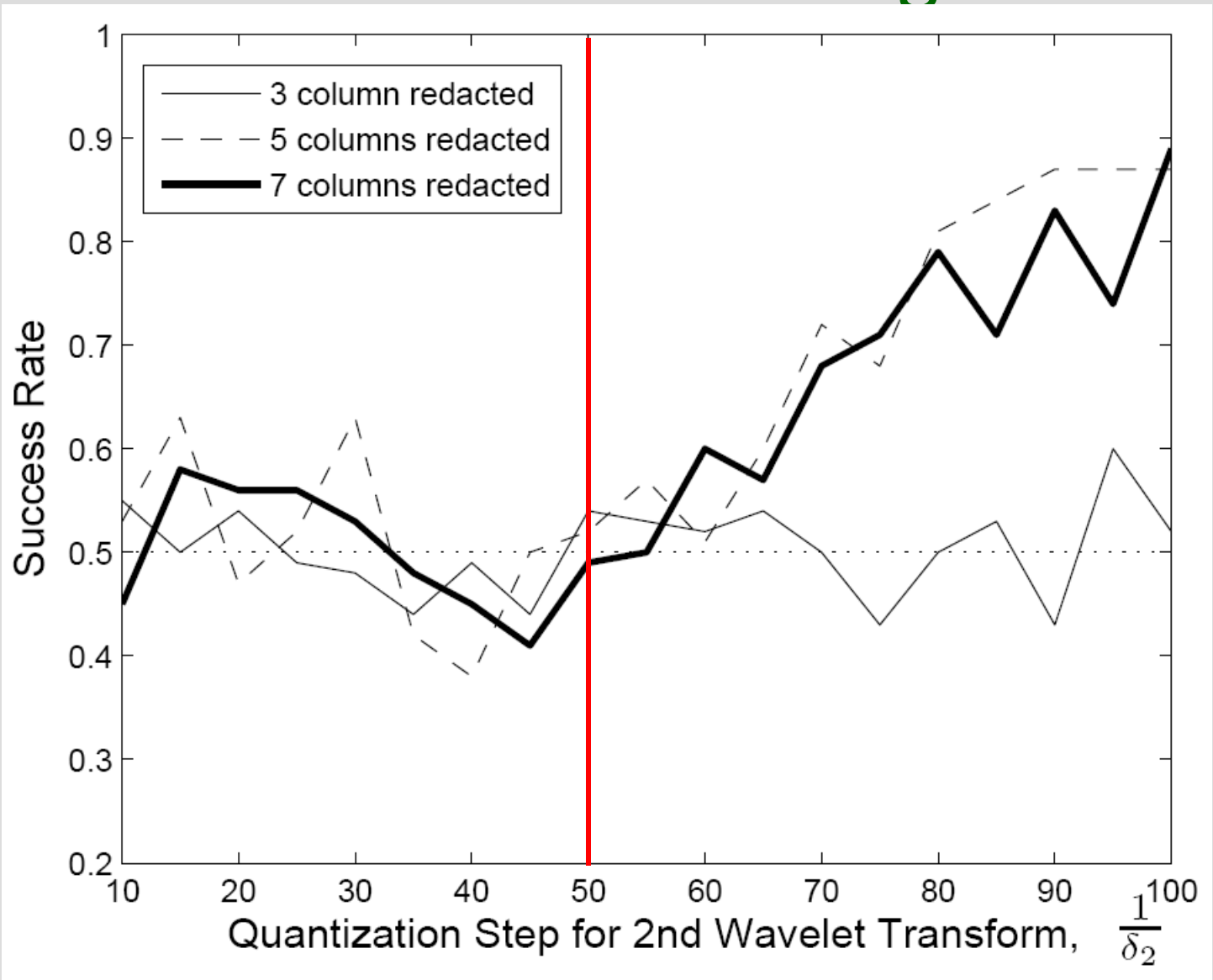
Random images, method 2, JPEG, $\delta_1 = 50$

Effect on estimation of δ_1



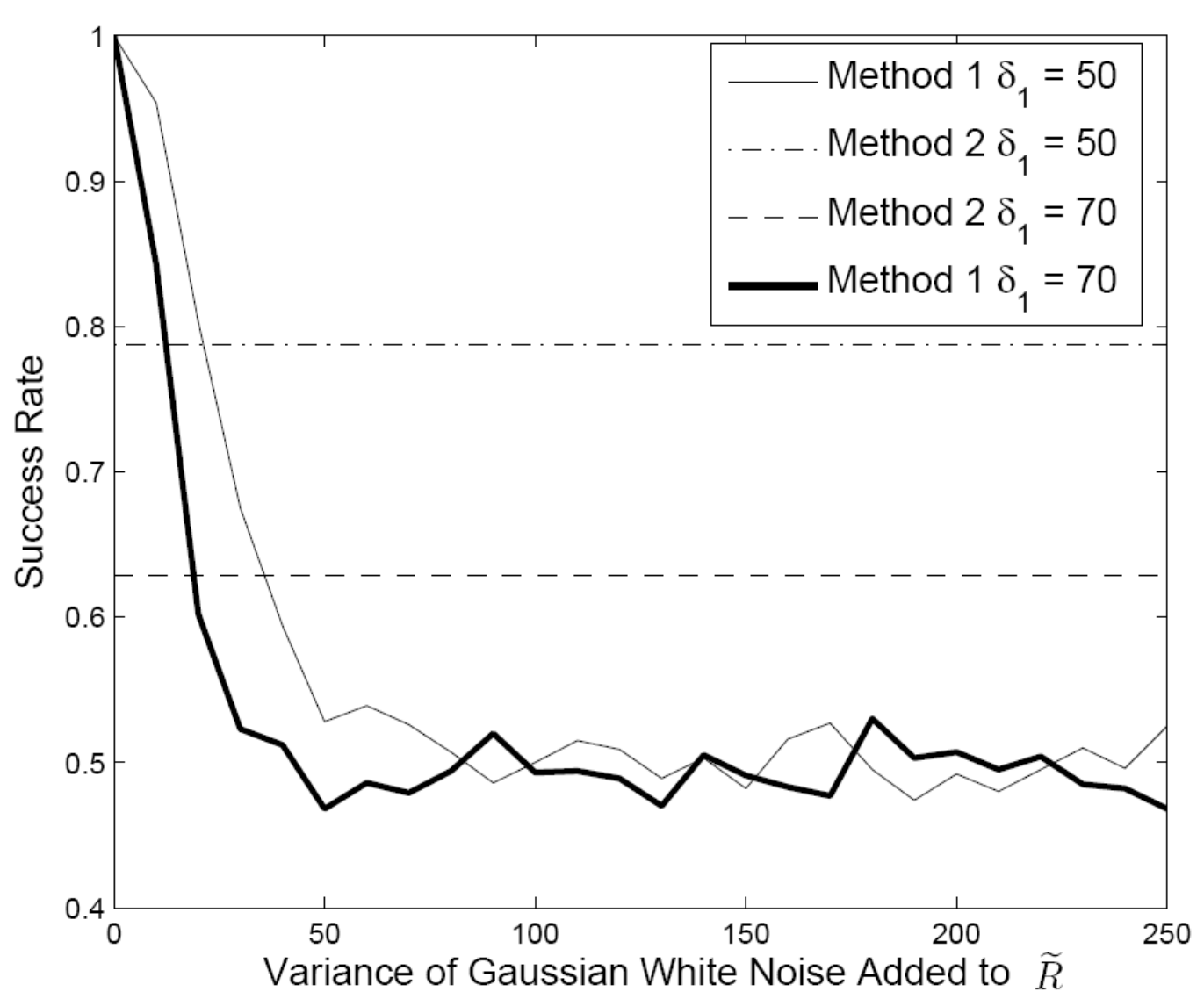
Random images, method 2, JPEG, $\delta_1 = 40$, $\delta_2 = 90$

Effect on size of redacted region

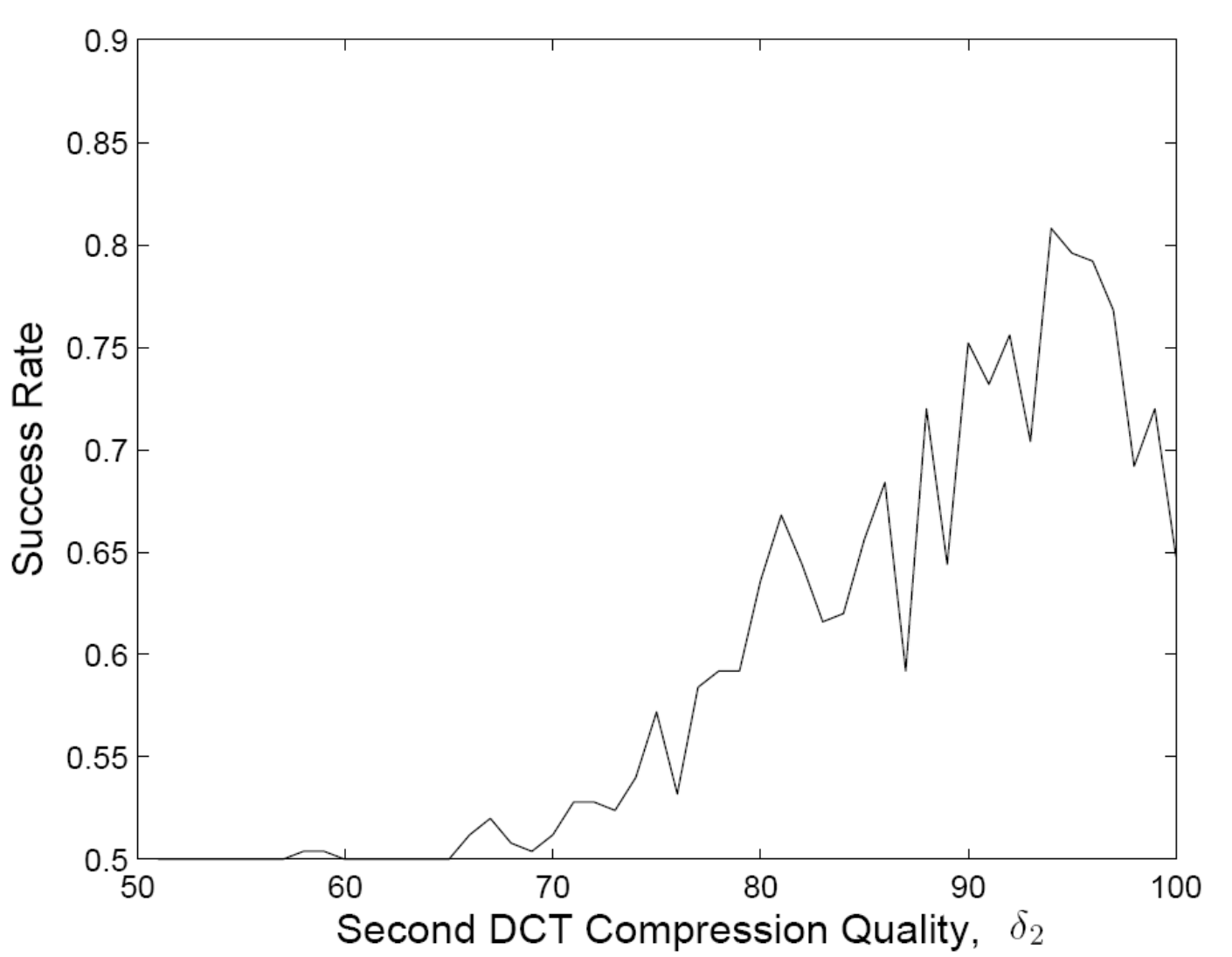


Random images, method 2, Wavelet, $\delta_1 = 50$

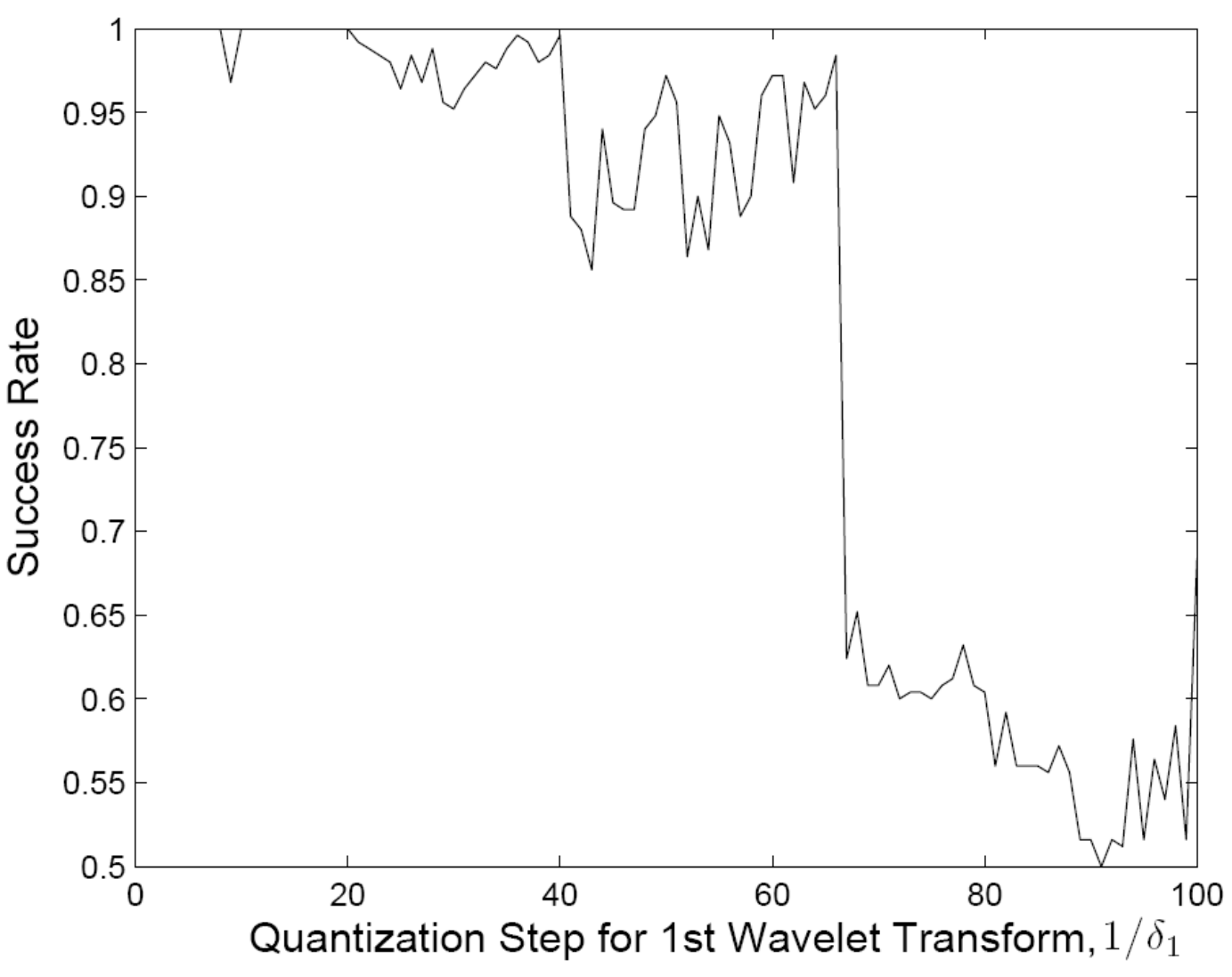
Comparison of method 1 and 2



Random images, JPEG, δ_2 = 95, 3 col's redacted



Document image, method 2, JPEG, $\delta_1 = 50$



Document image, method 2, Wavelet, $\delta_2 = 1/100$

Photo images(method 2)

	Quantization Error
Random	123.0
10-335	**92.6**
10-339	92.2
08-331	95.0
11-335	96.9
11-339	97.3

actual:10-335

	Quantization Error
Random	104.9
10-335	69.1
10-339	**67.1**
08-331	71.7
11-335	72.8
11-339	73.7

actual:10-339

Other details

- Translation and Geometric distortion.

- Many DCT blocks.

Conclusion

- When 2nd compression is of higher rate, adversary's success rate is high.

- Fortunately, typical images in public domain use lower rate for 2nd compression. (image scanned in high quality, redacted image stored in lower quality for fast downloading).

- Nevertheless, mobile phone camera is gaining popularity and images compressed in lower quality. Declassification of document images may not take the downloading speed as a consideration.

- Such subtle attack must still be taken into consideration when redacting sensitive images.

- Other similar attacks? A more accurate model and effective method.