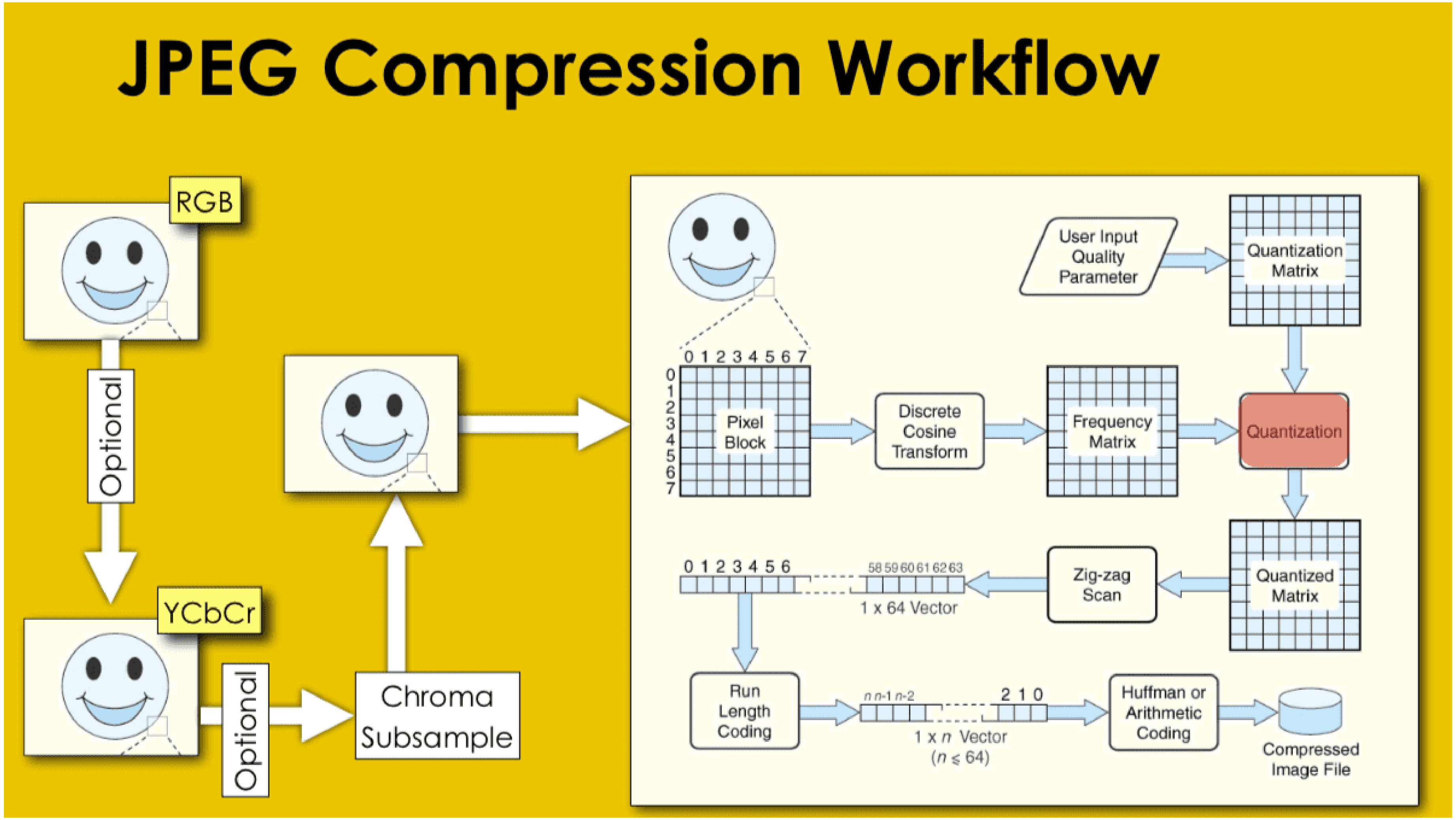
Steganography With Multiple JPEG Images of the Same Scene

TomÆsˇ Denemark, Student Member, IEEE, and Jessica Fridrich, Fellow, IEEE

IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 12, NO. 10, OCTOBER 2017

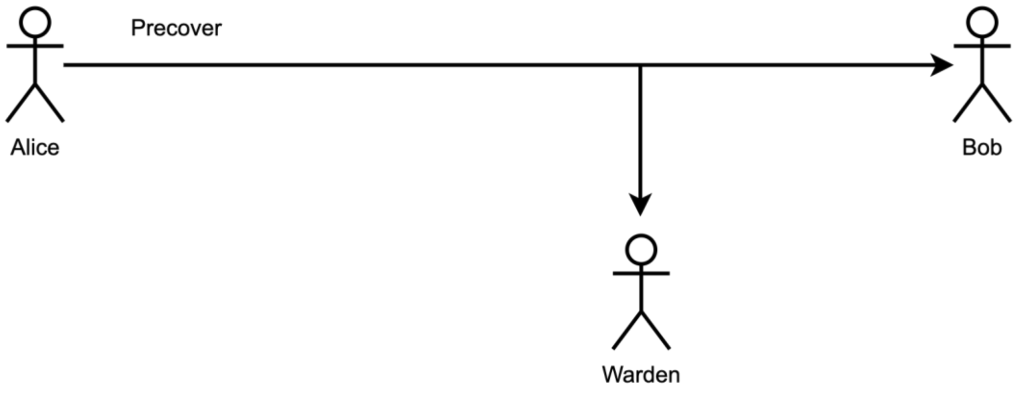
Prep: how to make a JPEG



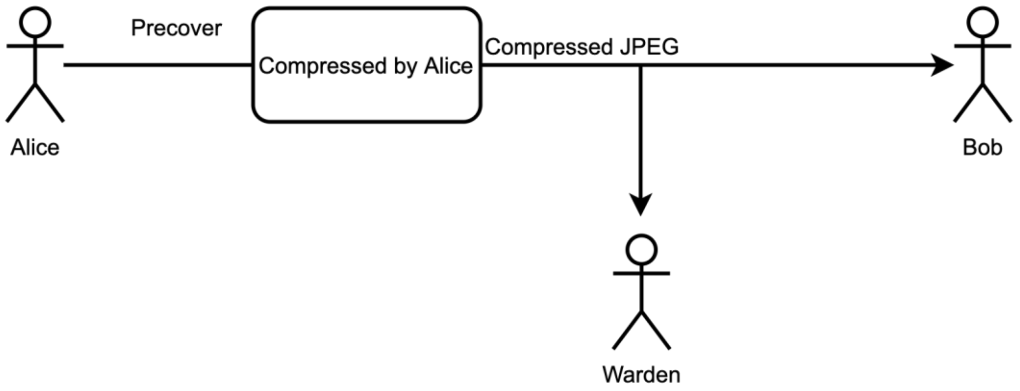
https://asvrada.github.io/blog/jpeg-compression/

Steg and delivery

Ideal model



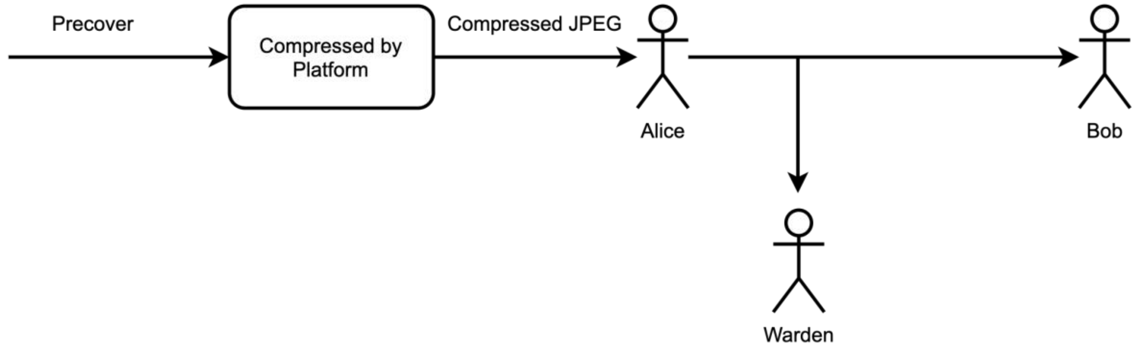
Real-life model



Under the real-life model

- Alice holds the precover edition
- But not for Bob and Warden
- Precover means…
  - Un-rounded DCT-> Embedding while compressing
  - True color info -> Embedding while converting color

A worse case



- Device or platform only saves compressed JPEG
- Neither did Alice have the precover

Previous attempts

- Side information: steganography with the help of multi scans
- Hard to model **noise** -inevitably
- Require too much scans
- Difficult to track the movement of scanner

Contributions

- A new method to embed with multi photos in the same scene
- Without noise modeling
- New way to calculate cost – to guide embedding
- More efficient – more bits can be embedded in per non-0 AC

Preliminaries

- $z_{i,j}^{(u,v)}$: original grayscale
- $d_{i,j}^{(u,v)}$: DCT coefficient
- $q_{i,j}^{(u,v)}$: quantization matrix depends on quality coefficient
- $c_{i,j} = \frac{d_{i,j}}{q_{i,j}}, x_{i,j} = [c_{i,j}]$: quantized DCT coefficient
- Then $x_{i,j}$ is losslessly encoded

a. Low compression

1	1	1	1	1	2	2	4
1	1	1	1	1	2	2	4
1	1	1	1	2	2	2	4
1	1	1	1	2	2	4	8
1	1	2	2	2	2	4	8
2	2	2	2	2	4	8	8
2	2	2	4	4	8	8	16
4	4	4	4	8	8	16	16

b. High compression

1	2	4	8	16	32	64	128
2	4	4	8	16	32	64	128
4	4	8	16	32	64	128	128
8	8	16	32	64	128	128	256
16	16	32	64	128	128	256	256
32	32	64	128	128	256	256	256
64	64	128	128	256	256	256	256
128	128	128	256	256	256	256	256

FIGURE 27-13
JPEG quantization tables. These are two example quantization tables that might be used during compression. Each value in the DCT spectrum is divided by the corresponding value in the quantization table, and the result rounded to the nearest integer.

Previous works: Steg based on the cost

- Cost=Distortion

- To minimize the cost of changing quantized DCT
- To embed: $y_{i,j} \in \{x_{i,j}-1, x_{i,j}, x_{i,j}+1\}$
- $\rho_{i,j}^{u,v}(\pm 1)$
- Total cost: $D(x,y) = \sum_{x_{i,j} \neq \mathrm{y}_{i,j}} \rho_{i,j}\left(y_{i,j} - x_{i,j}\right)$

To determine the cost: J-UNIWARD

- Symmetric cost
- $\rho_{i,j}^{u,v}(+1)$=$\rho_{i,j}^{u,v}(-1)$
- To prohibit embedding: set $\rho_{i,j}^{u,v}(+1)$=$\rho_{i,j}^{u,v}(-1) = \infty = C_{wet}$

To determine the cost: SI-UNIWARD

- How : leave DCT coefficient unmodified or round it to other side
- Use non-rounded DCT coefficients as side-information
- Not suitable for the worse model
- Round cost: $e_{i,j} = c_{i,j} - x_{i,j} \in (-0.5, +0.5)$
- When $|e_{i,j}|$ closes to ½, a small perturbation could cause $c_{i,j}$ to be rounded to the other side
- E.g. 4.4→5 cost~0.1 4.1→5 cost~0.4
- $\rho_{i,j}^{u,v}(+1) = \left(1 - 2\left|e_{i,j}\right|\right)\rho_{i,j}^{J}$
- $\rho_{i,j}^{u,v}(-1) = \infty = C_{wet}$
- Such coefficients are thus assigned a proportionally smaller cost

Steg with multiple JPEGs

- Exposures are relevant but also different
- Use other exposures as side-info to improve security
- Cost-based
- Composition of noise: $z = r + \varepsilon$
- Expandable

Evaluation

- A simple linear classifier: GFR
- GFR can be highly effective against J-UNIWARD
- $P_E = (P_{MD} + P_{FA})/2$

Two exposures: J2-UNIWARD

- Cost calculation based on J-UNIWARD
- If $x_{i,j}^1 = x_{i,j}^2$, do nothing
- If $x_{i,j}^1 \neq x_{i,j}^2$,Set$\rho_{i,j}(s_{i,j}) = m(Q) \times \rho_{i,j}^0(s_{i,j})$
- $m(\cdot)$ relies practice, based on quality factor Q
- Using STCS to embed based on the cost calculated above

Algorithm 1 Pseudo-Code for Side-Informed Embedding With Two JPEGs

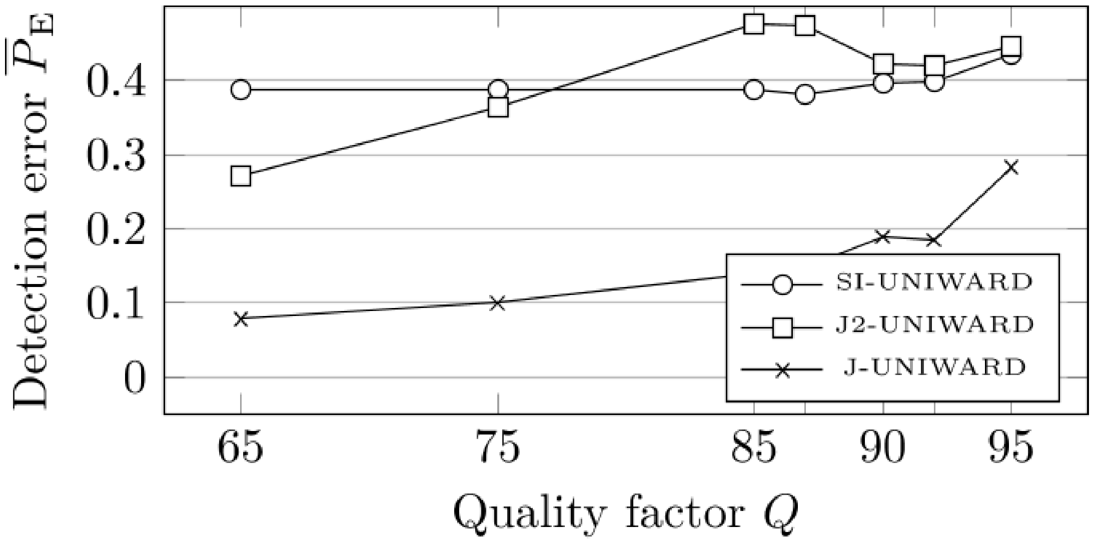
Input: Two quality factor Q JPEG images with quantized DCT coefficients $x_{ij}^{(1)}$ and $x_{ij}^{(1)}$, $1 \le i \le M$, $1 \le j \le N$
Output: Stego JPEG image with DCT coefficients $y_{ij}^{(1)}$
Compute costs $\rho_{ij}^{(0)}(-1), \rho_{ij}^{(0)}(+1)$ of DCT coefficients from JPEG $x_{ij}^{(1)}$ (the cover)
for $i = 1, \ldots, M$ **do**
  for $j = 1, \ldots, N$ **do**
    $\rho_{ij}(\pm 1) = \rho_{ij}^{(0)}(\pm 1)$
    $s_{ij} = \text{sign}(x_{ij}^{(2)} - x_{ij}^{(1)})$
    IF $x_{ij}^{(1)} \neq x_{ij}^{(2)}$ **THEN** $\rho_{ij}(s_{ij}) = m(Q)\rho_{ij}^{(0)}(s_{ij})$
  end for
end for
Embed message in $x_{ij}^{(1)}$ using costs ρ_{ij} using STCs to obtain stego JPEG file with DCT coefficients y_{ij}
Recipient reads the secret message using STCs from the stego JPEG file y_{ij}

Solution of multi exposures

- $x_{i,j}^1, x_{i,j}^2, x_{i,j}^3, \cdots, x_{i,j}^k$
- More accurate estimation of noise-free scene(precover): $\widehat{r_{i,j}} = average(x_{i,j})$
- But spatial misalignment will also accumulate with k increase
- $e = \widehat{r_{i,j}} - x_{i,j}^1$ may be out of $(-0.5, +0.5)$, cause calculation errors
- The best practice is to select two closest exposures

Test1: simulated acquisition noise

- Gaussian noises $z_{i,j}^1, z_{i,j}^2$ are added
- 0.4 bits payload per non-zero AC DCT coefficient
- Evaluation metrics: $\overline{P_E}$
- Better than J-UNIWARD
- Better than SIUNIWARD when Q>80
- Side info $x_{i,j}^2$ can only be useful when $x_{i,j}^1 \neq x_{i,j}^2$, the possibility increase with the quality of JPEG
- $x_{i,j}^2$ can tell Alice the direction along which the costs should be modulated, rather than simple $e_{i,j} = c_{i,j} - x_{i,j}$



Real-life datasets

BURSTBase

- Using tripod
- 133×7 pics
- 9310 smaller 512×512 grey scales

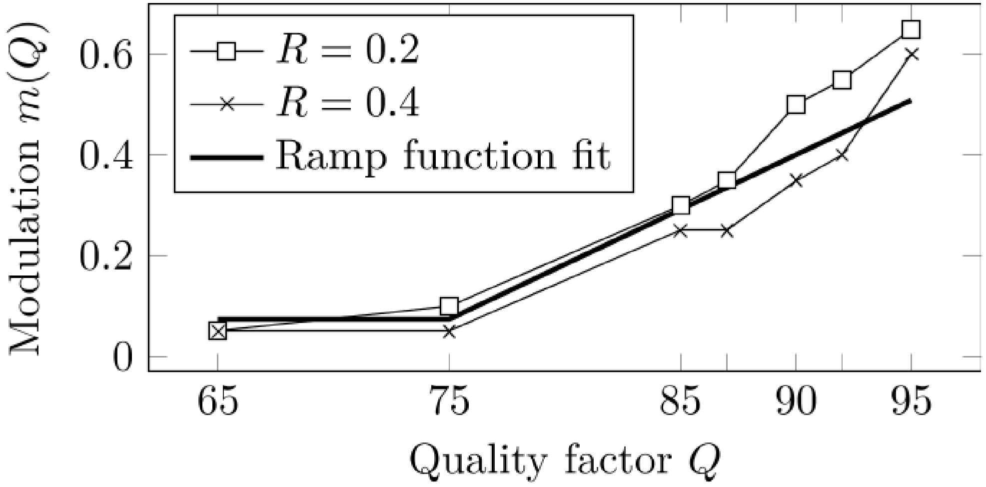
BURSTBaseH

- Using hands
- 154×(7~13) pics
- 17480 smaller 512×512 grey scales

Mean square error(MSE): to evaluate the difference between exposures
Exposures are sorted by MSE in ascending order

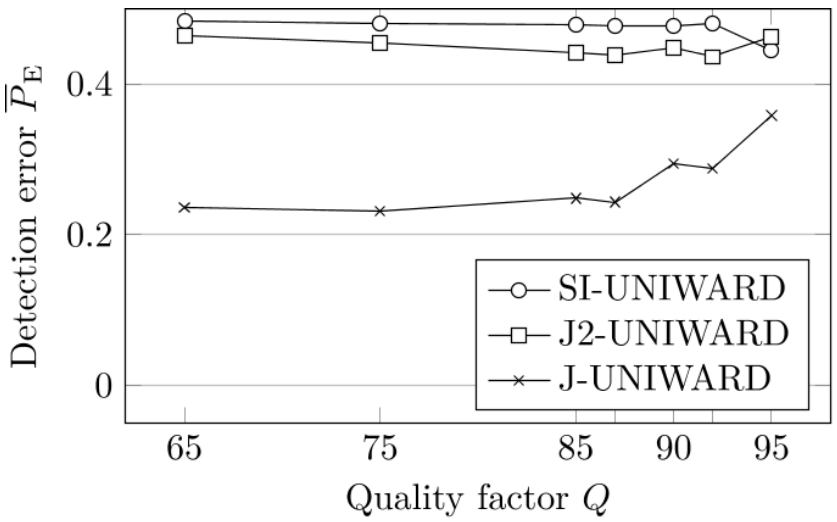
Test2:BURSTBase

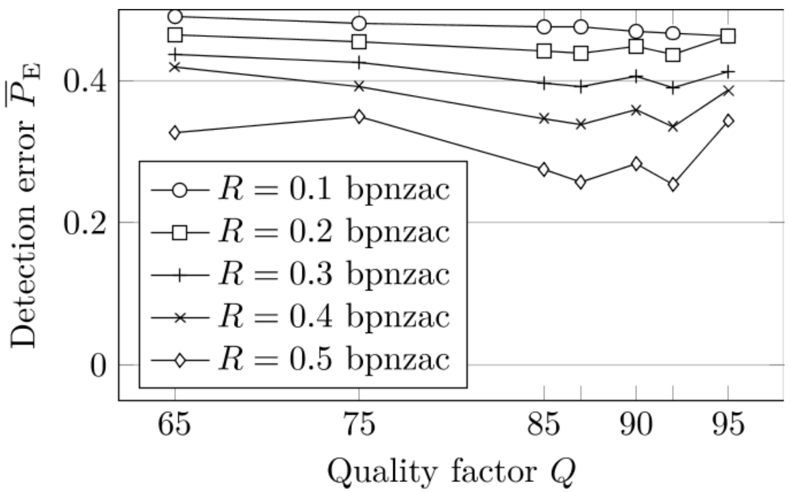
- $m(Q) = Max(0.075, 0.02167 \times Q - 1.55)$



Test2:BURSTBase

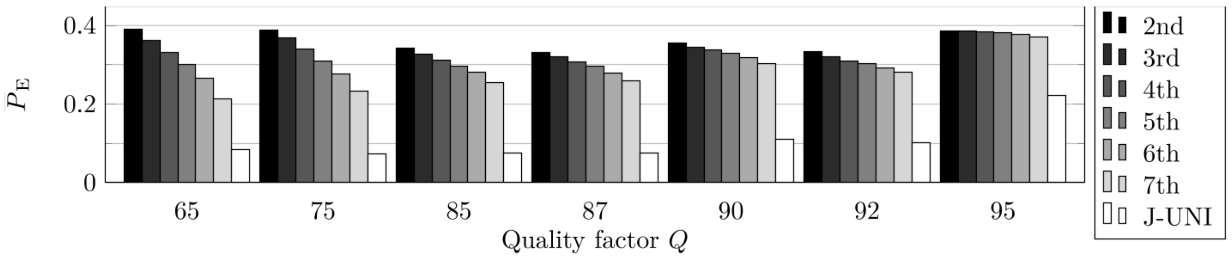
- The empirical security is better than J-UNIWARD
- Not better than non-rounded DCT coefficients are used as side-information (SI-UNIWARD).
- Less efficiency, better security





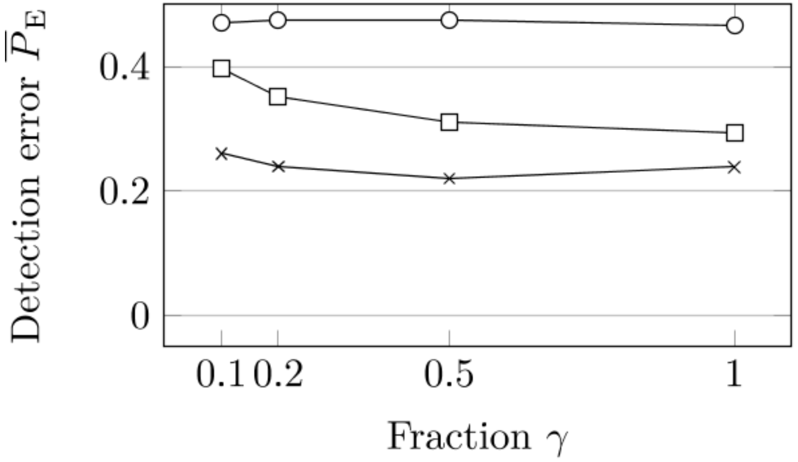
Test2:BURSTBase

- Exposures with more difference have lower rank
- Too much differences leads to lower security performance
- The trend is less pronounced when Q increases

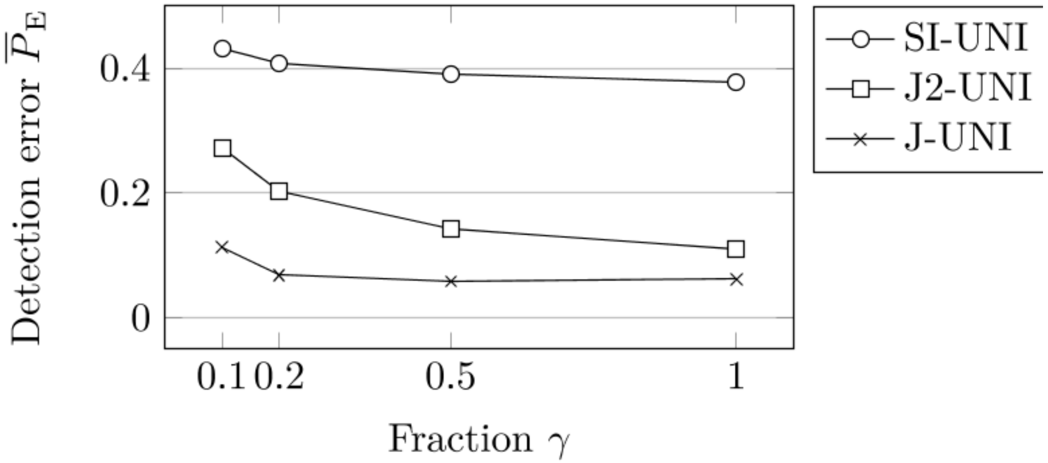


Test 3: BURSTBaseH

- $\gamma \in (0,1]$: the top γ of the images with the least MSE in the dataset
- Set $m(Q) = \mathbf{0.01}$ and $Q = \mathbf{75}$
- Payload=0.2/0.4 pnzac
- Substantial gain compared with J-UNIWARD
- More rejected burst, better security performance

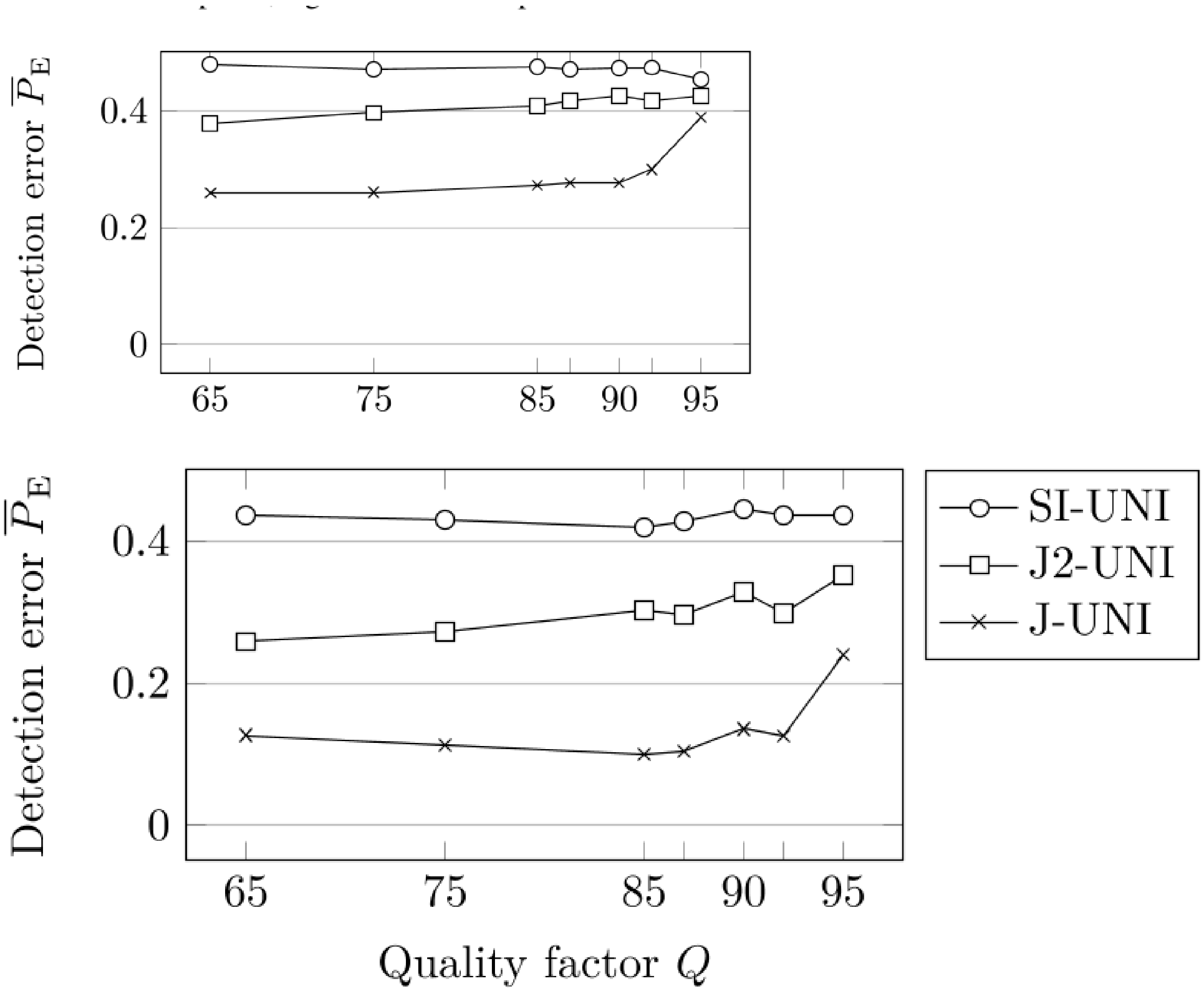


–



Test 3: BURSTBaseH

- Set γ=0.1
- Payload=0.2/0.4 pnzac
- Gain with Q is smaller than BURSTBase



Conclusion

- Using another same exposure to infer the preferred direction of steganographic embedding changes
- Can significantly increase security, much better than J-UNIWARD
- In ideal environment(regular noise or larger quality factors), even better than SI-UNIWARD
- Adapt to real-life communication model
- Rejecting “bad bursts” can be a good idea
- Future work: get more stable identical exposures from consecutive frames of a video clip

Thanks!