SFU SIMON FRASER UNIVERSITY
THINKING OF THE WORLD

NII National Institue of Informatics

SALIENCY-COGNIZANT ERROR CONCEALMENT IN LOSS-CORRUPTED STREAMING VIDEO

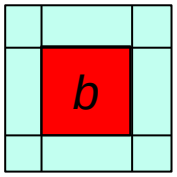
Hadi Hadizadeh and Ivan V. Bajić
Simon Fraser University
Burnaby, BC, Canada

Gene Cheung
National Institute of Informatics
Tokyo, Japan

The Idea

Goal: Reconstruct a missing block ***b*** lost during video streaming such that some fitting error cost function is minimized:

$$\min_{\boldsymbol{b}} fit\_err(\boldsymbol{b})$$



Problem: The problem is **under-determined**.

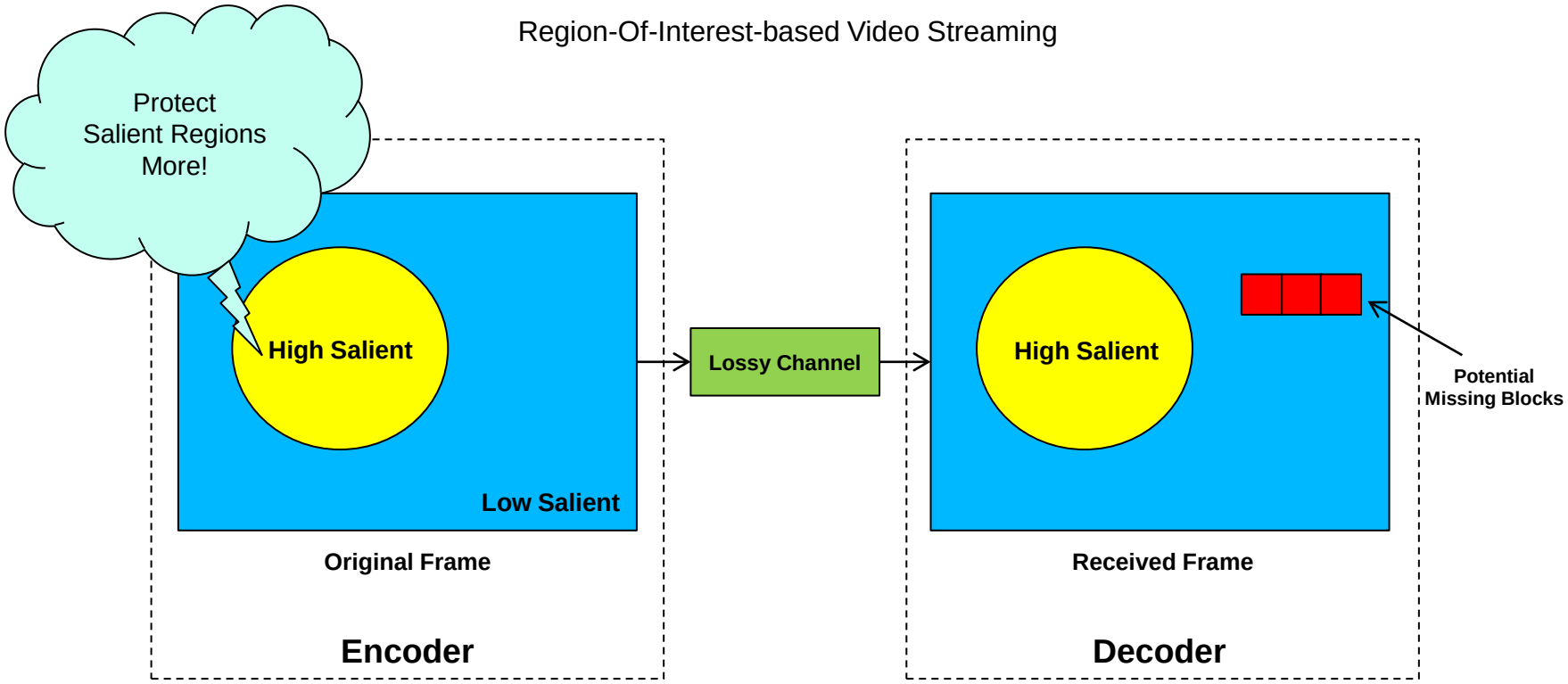
Solution: Add a **saliency term** as follows:

$$\min_{\boldsymbol{b}}\{fit\_err(\boldsymbol{b}) + \lambda\, saliency(\boldsymbol{b})\}$$

Advantages:

1. Potential wrong candidates become **less attention-grabbing**.
2. It serves as **a true prior** in an ROI-based streaming application.

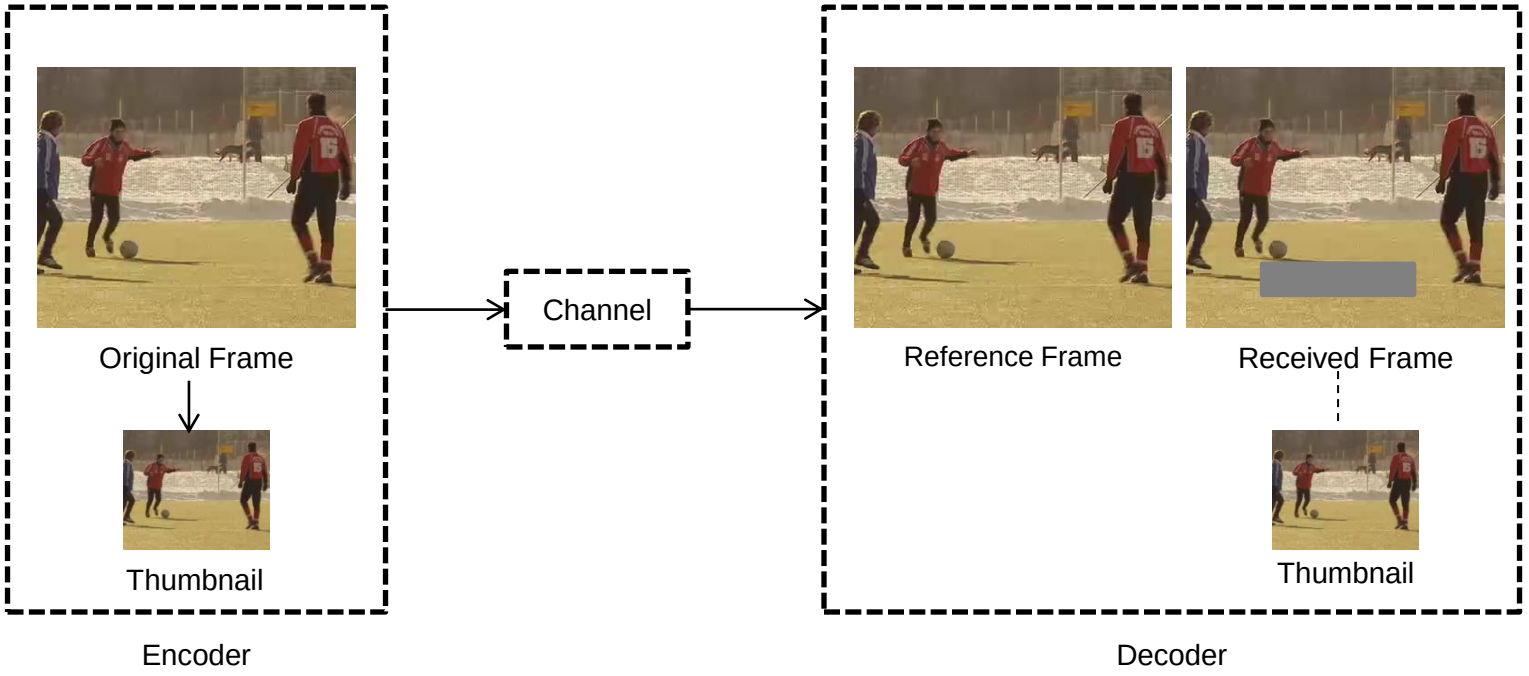
Loss Resilient Video Streaming



G.-M. Muntean, et al., “Region of interest-based adaptive multimedia streaming scheme,” IEEE Transactions on Broadcasting, 2008

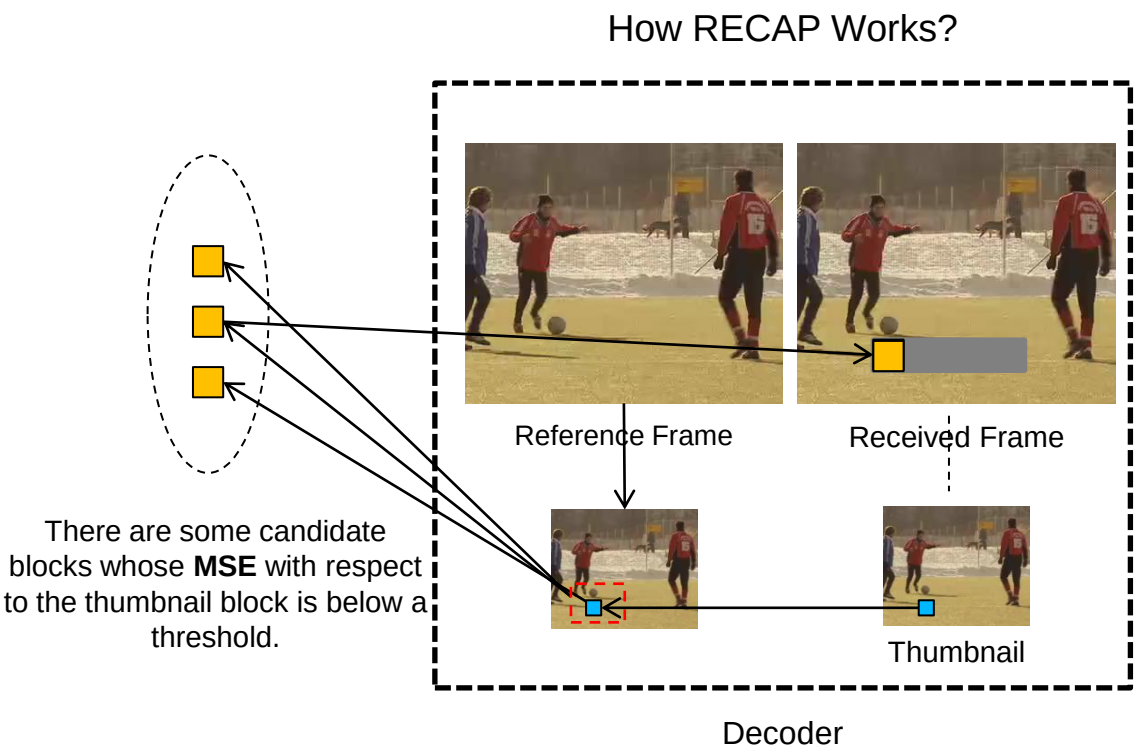
RECAP

RECAP: **R**eceiver **E**rror **C**oncealment using **A**cknowledge **P**review (Thumbnail)

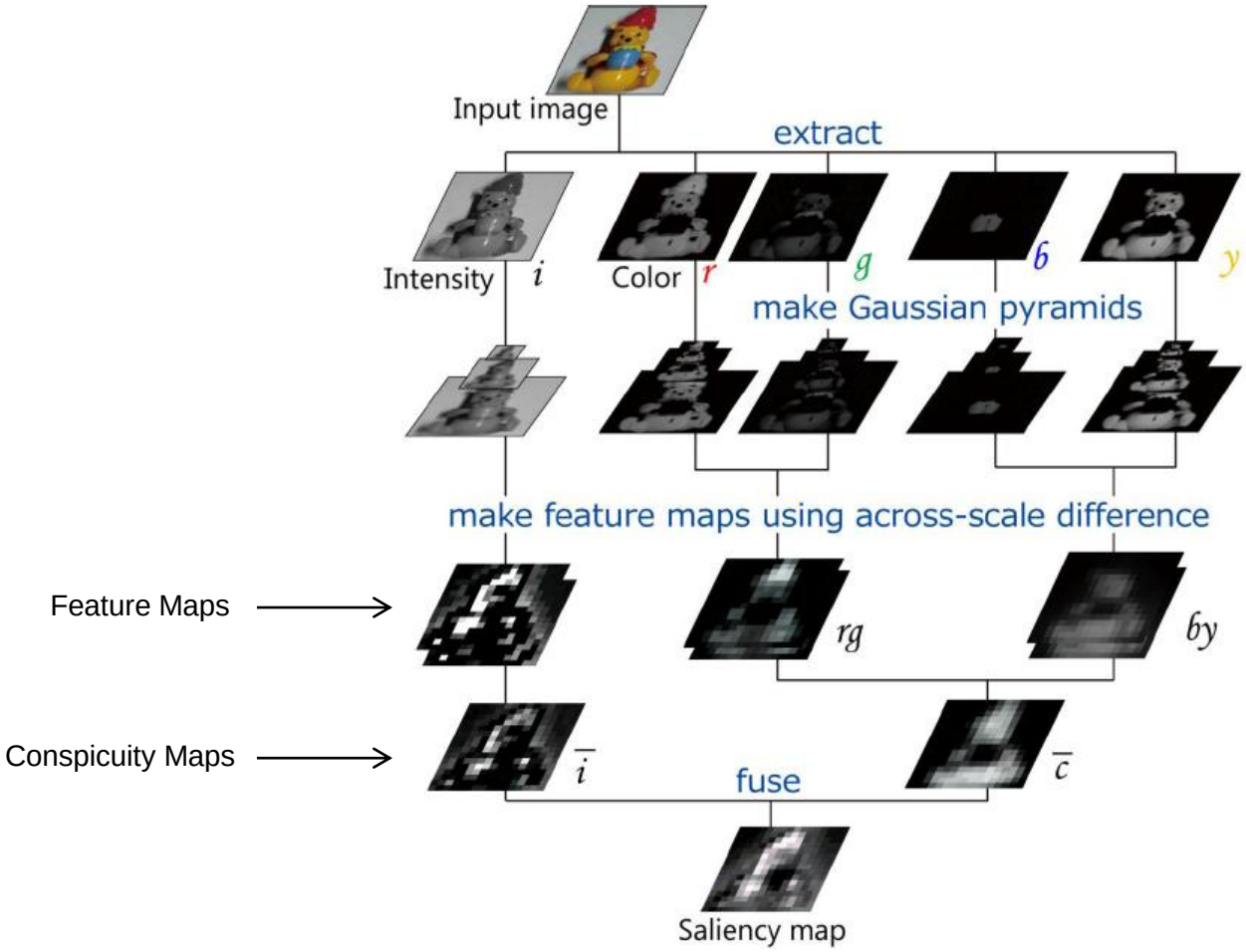


C. Yeo, et al. Receiver Error Concealment using Acknowledge Preview – An Approach to Resilient Video Streaming, ICASSP 2009

RECAP

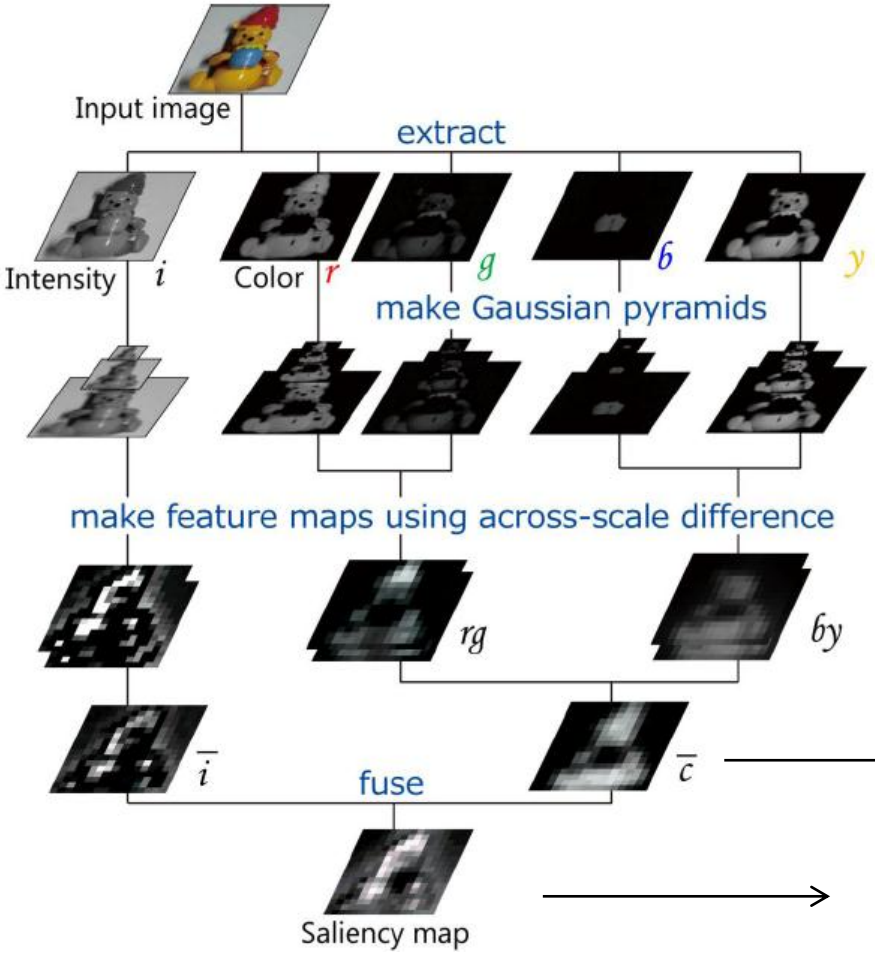

How RECAP Works?
Reference Frame
Received Frame
There are some candidate blocks whose **MSE** with respect to the thumbnail block is below a threshold.
Thumbnail
Decoder

The Itti Saliency Model



L. Itti, “Automatic Foveation for Video Compression using a Neurobiological Model of Visual Attention,” IEEE TIP 2004.

Frequency Analysis of the Itti Model



8 Levels are used in each pyramid.

The conspicuity maps are combined at Level 4.

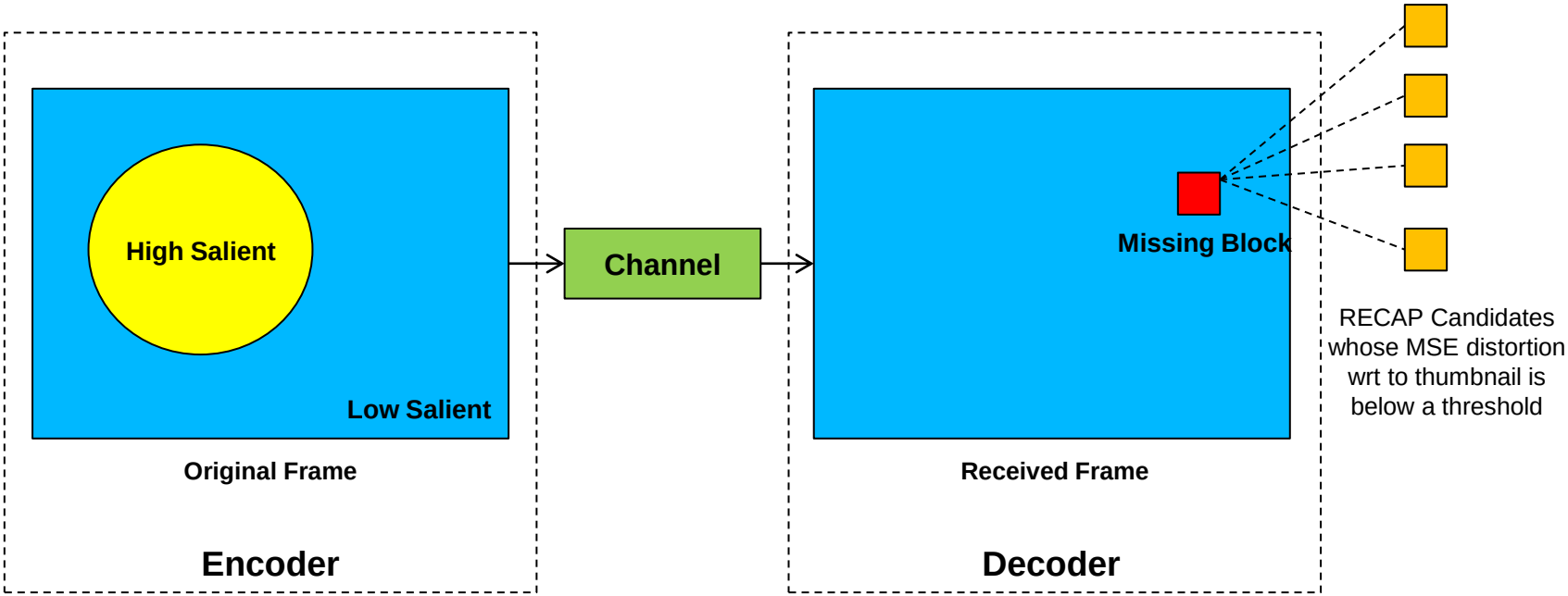
$$\left[\frac{\pi}{256} \sim \frac{\pi}{16}\right]$$

The Effective Frequency Range Used by the Itti Model

R. Achanta, et al., “Frequency-tuned salient region detection,” CVPR 2009

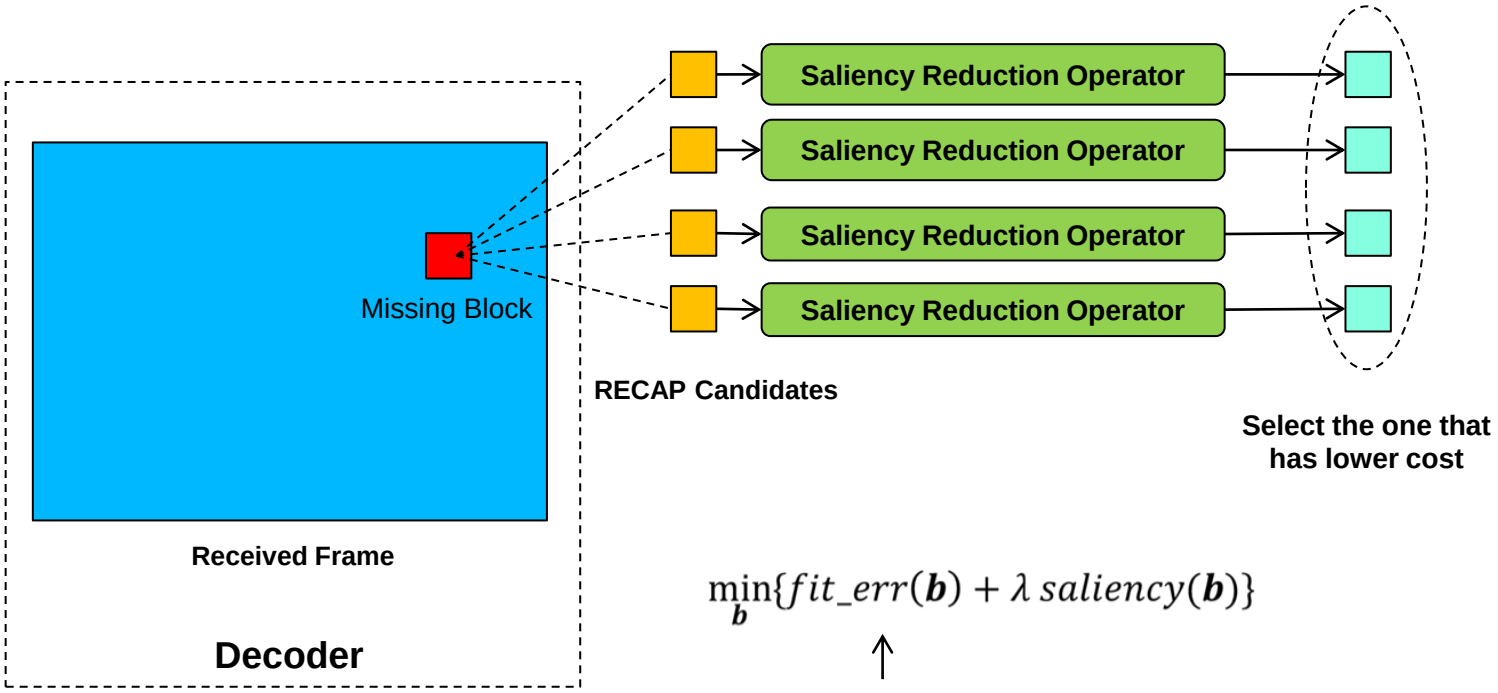
The Proposed Method

We implement our proposed method within a **ROI-based video streaming framework.**
We use **RECAP** as our error concealment method.



Problem: Which candidate block should we take?
Solution: Take the one that has lower saliency!

The Proposed Method



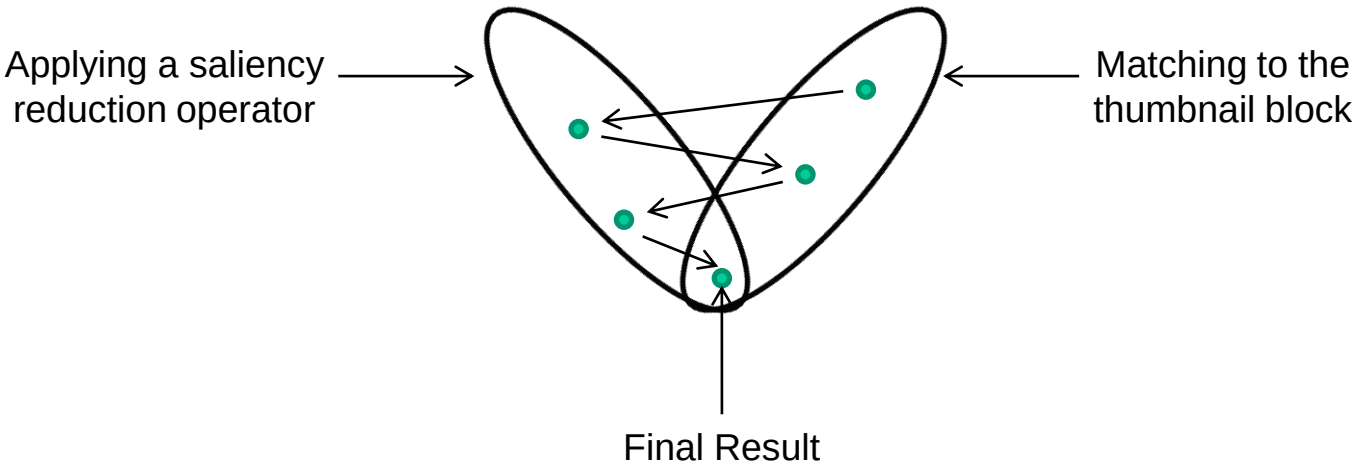
$$\min_{\boldsymbol{b}}\{fit\_err(\boldsymbol{b}) + \lambda\, saliency(\boldsymbol{b})\}$$

MSE between ***b*** and thumbnail.

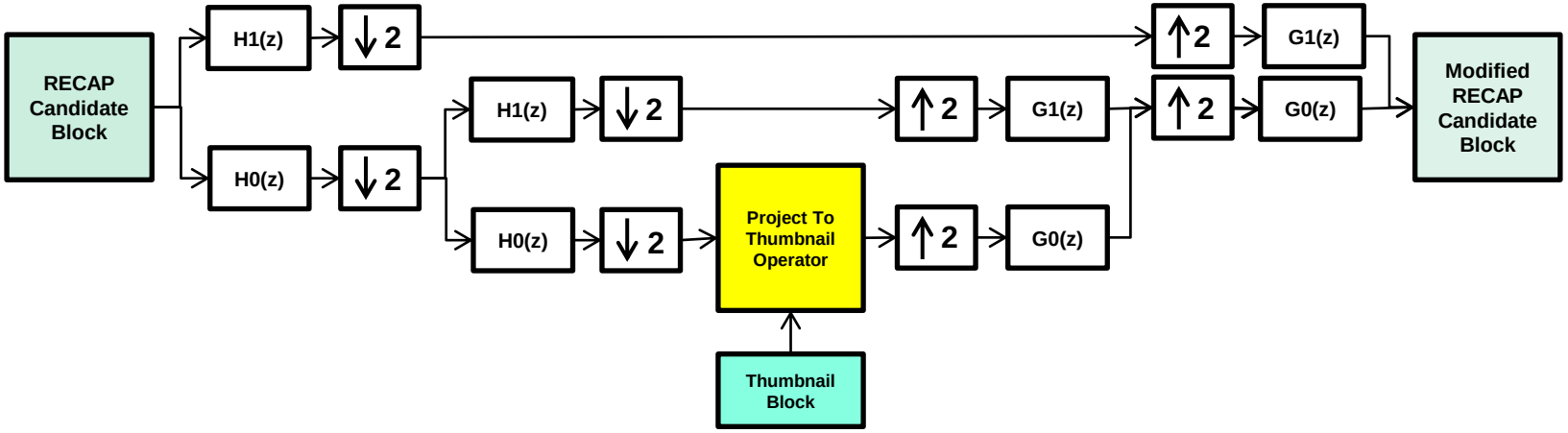
Yet a better solution: First, apply some filters on a bunch of candidates to reduce their saliency as much as possible, then select the one that has a lower saliency.

The Proposed Method

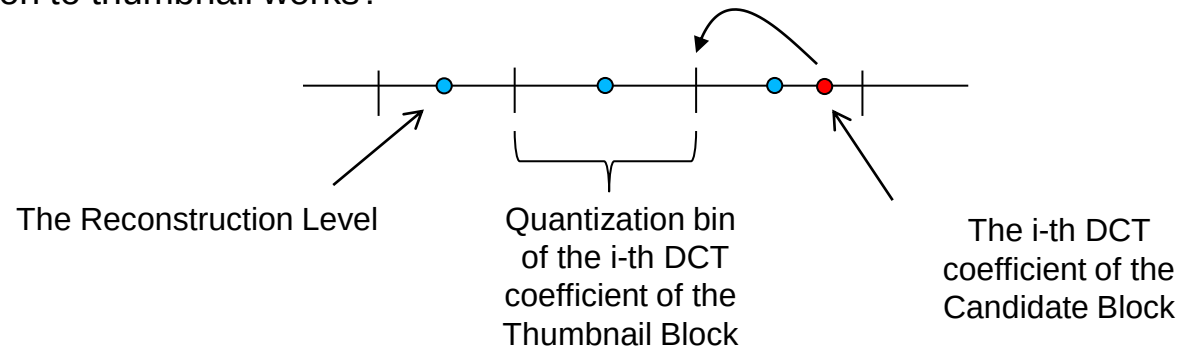
- Iteratively apply a Saliency Reduction Operator.
- Each time make sure that the solution is in good match with the thumbnail block.



Matching to Thumbnail



How projection to thumbnail works?



Saliency Reduction Operators

We propose 4 saliency reduction operators:

1. Deblocking Filter
2. Notch Filter
3. Frequency Outlier Filter
4. Attention-Guiding Method

Deblocking Filter

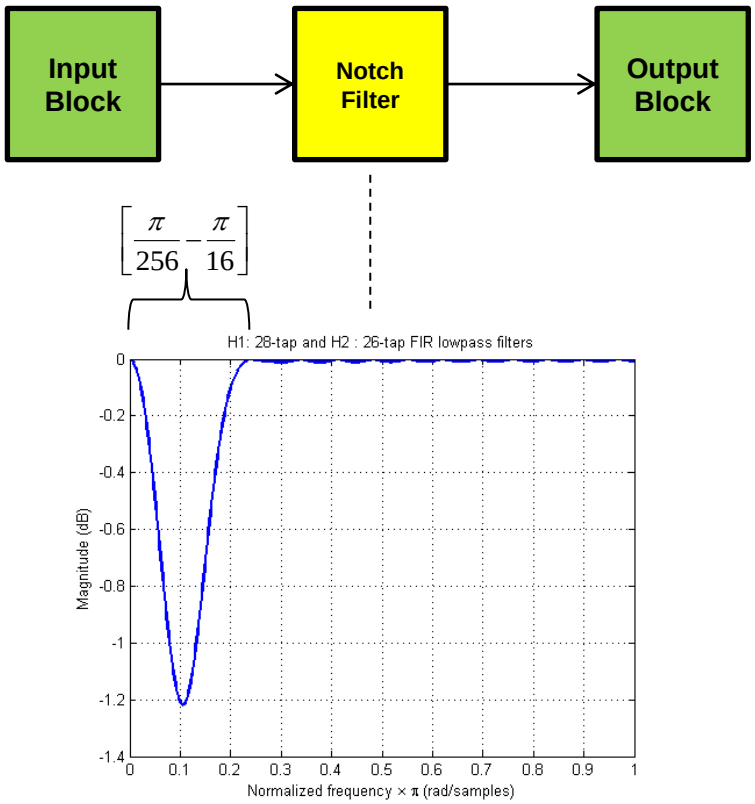
Deblocking reduces saliency!

Before Deblocking

After Deblocking

P. List, et al., “Adaptive deblocking filter”, IEEE TCSVT 2003

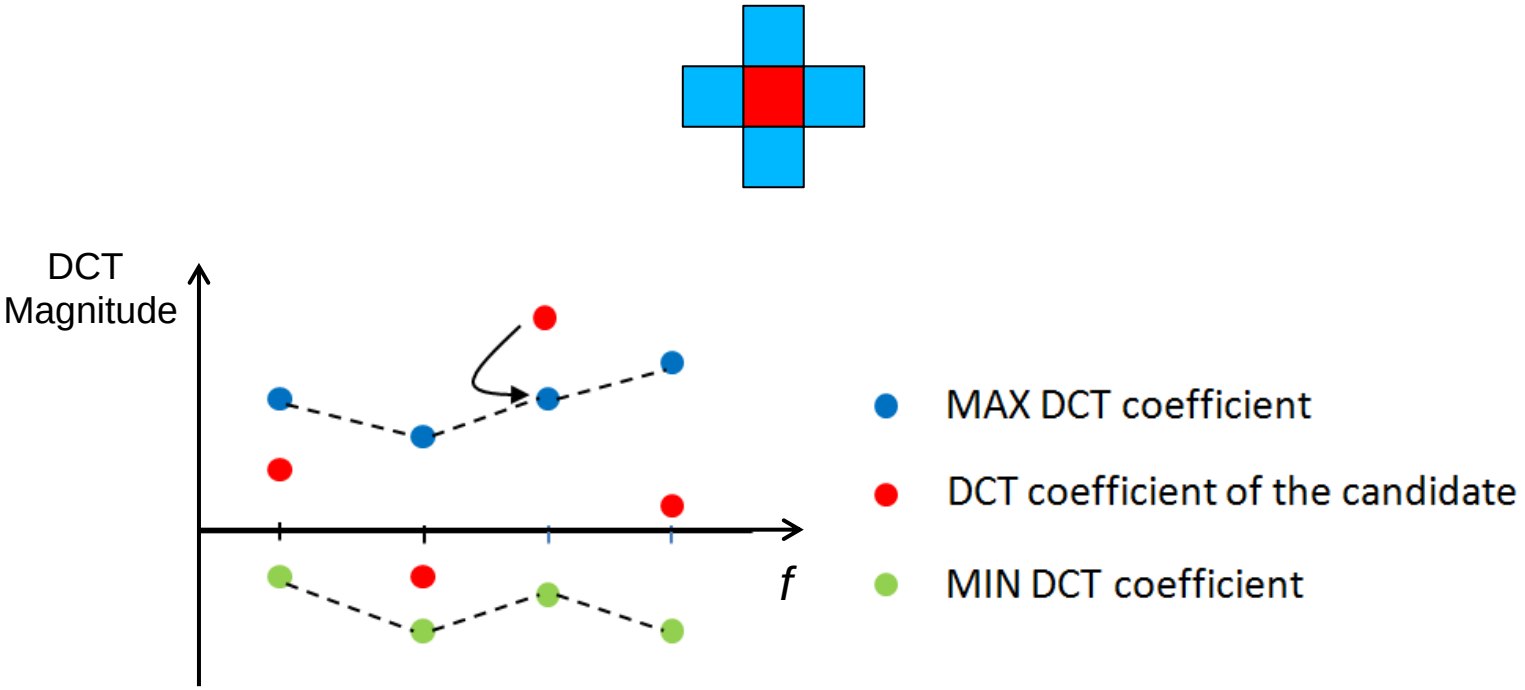
Notch Filter



R. Achanta, et al., “Frequency-tuned salient region detection,” CVPR 2009

Frequency Outlier Filter

Clip the DCT coefficients of the central block based on the DCT coefficients of the neighboring blocks.



X. Hou, et al. “Saliency detection: A spectral residual approach,” CVPR 2007

Attention-Guiding Method

Modify RGB values so that the region of interest becomes more/less salient.

Input Image | Saliency Map

Output Image | Saliency Map

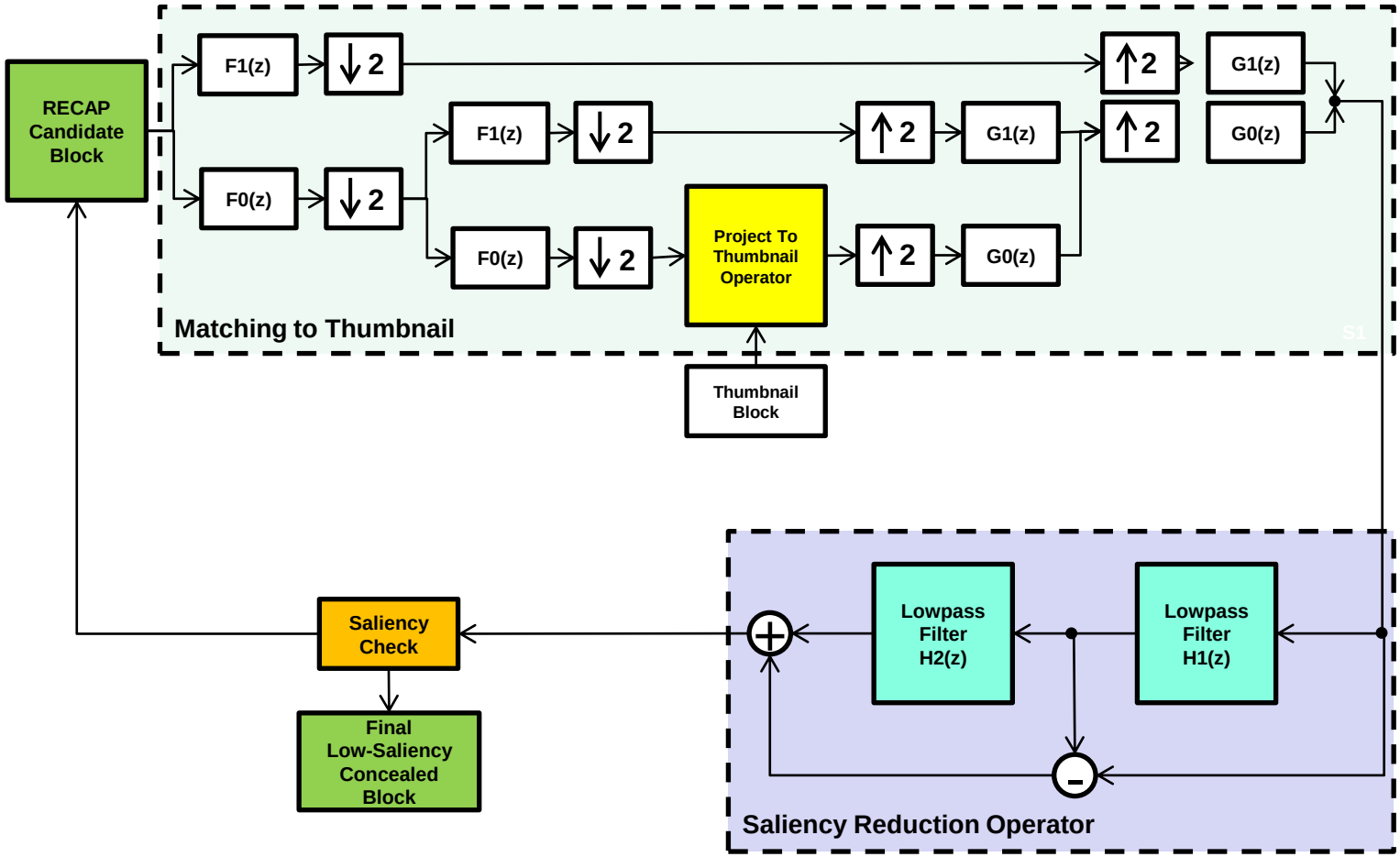
We replace this by a negative sign to reduce saliency instead.

$$\alpha^*_{pq} = \alpha_{pq} + w_{pq} V_{\alpha_{pq}}$$

- α^*_{pq}: New RGB value
- α_{pq}: Old RGB value
- w_{pq}: A Weight Factor
- $V_{\alpha_{pq}}$: A factor that reflects how much a feature influences the saliency of the pixel

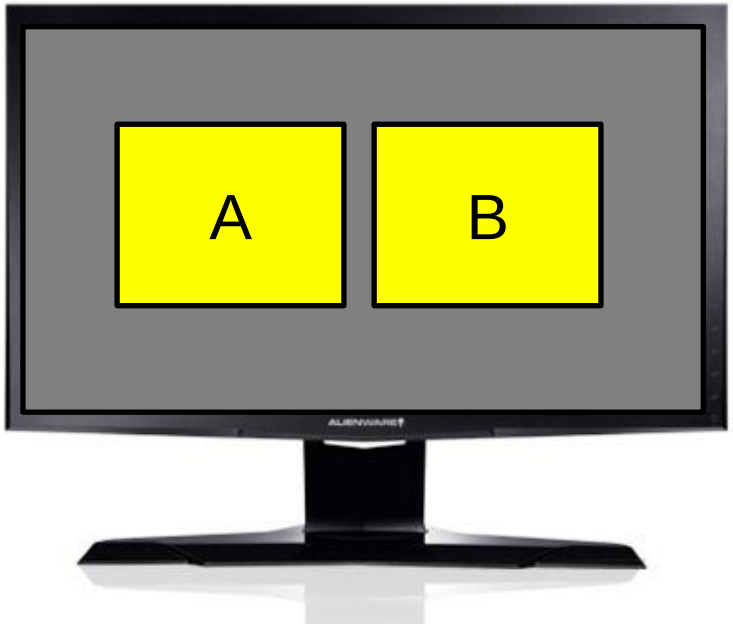
A. Hagiwara, et al., "Saliency-Based Image Editing for Guiding Visual Attention", PETMEI 2011

A Block-Diagram of the Proposed Method



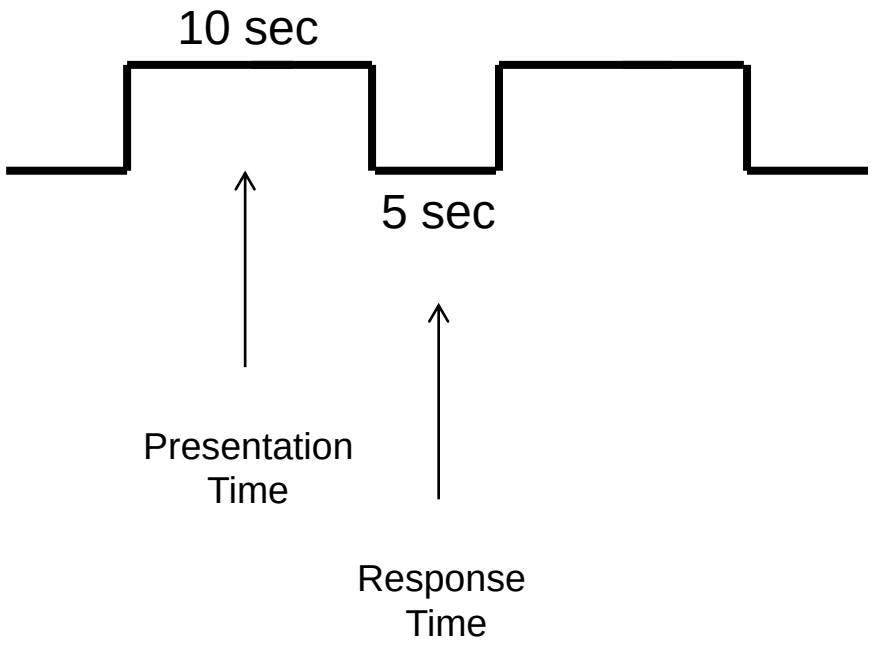
Subjective Tests

17 people



2AFC Experiment

2AFC: Two Alternative Forced Choice



Subjective Tests

Comparing the proposed method with the RECAP method based on the subjective results at 5 different average loss rates.

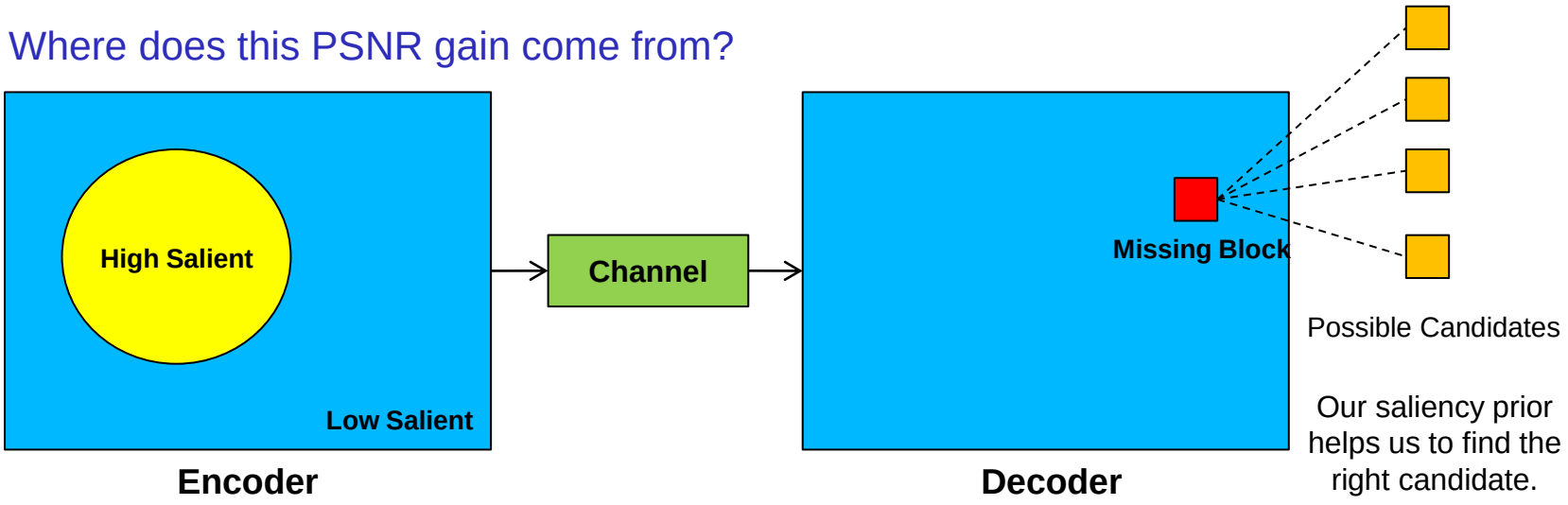
Loss Rate	Method	*Bus*	*Crew*	*Football*	*Stefan*
2%	RECAP	7	4	9	10
	Proposed Method	27	30	25	24
	p-value	0.0006	0.0001	0.0061	0.0164
5%	RECAP	4	3	7	9
	Proposed Method	30	31	27	25
	p-value	0.0001	0.0001	0.0006	0.0061
10%	RECAP	7	3	10	8
	Proposed Method	27	31	24	26
	p-value	0.0006	0.0001	0.0164	0.0020
20%	RECAP	8	8	11	7
	Proposed Method	26	26	23	27
	p-value	0.0020	0.0020	0.0396	0.0006
30%	RECAP	8	10	11	10
	Proposed Method	26	24	23	24
	p-value	0.0020	0.0164	0.0396	0.0164

Quantitative Results

Average PSNR gain and saliency reduction amount achieved by the proposed method over RECAP.

	Bus	*Crew*	*Football*	*Stefan*
PSNR Gain ($d = 10$)	1.2 dB	3.2 dB	1.8 dB	0.9 dB
PSNR Gain ($d = 5$)	0.9 dB	2.8 dB	1.5 dB	0.6 dB
Saliency Reduction	10%	19%	12%	9%

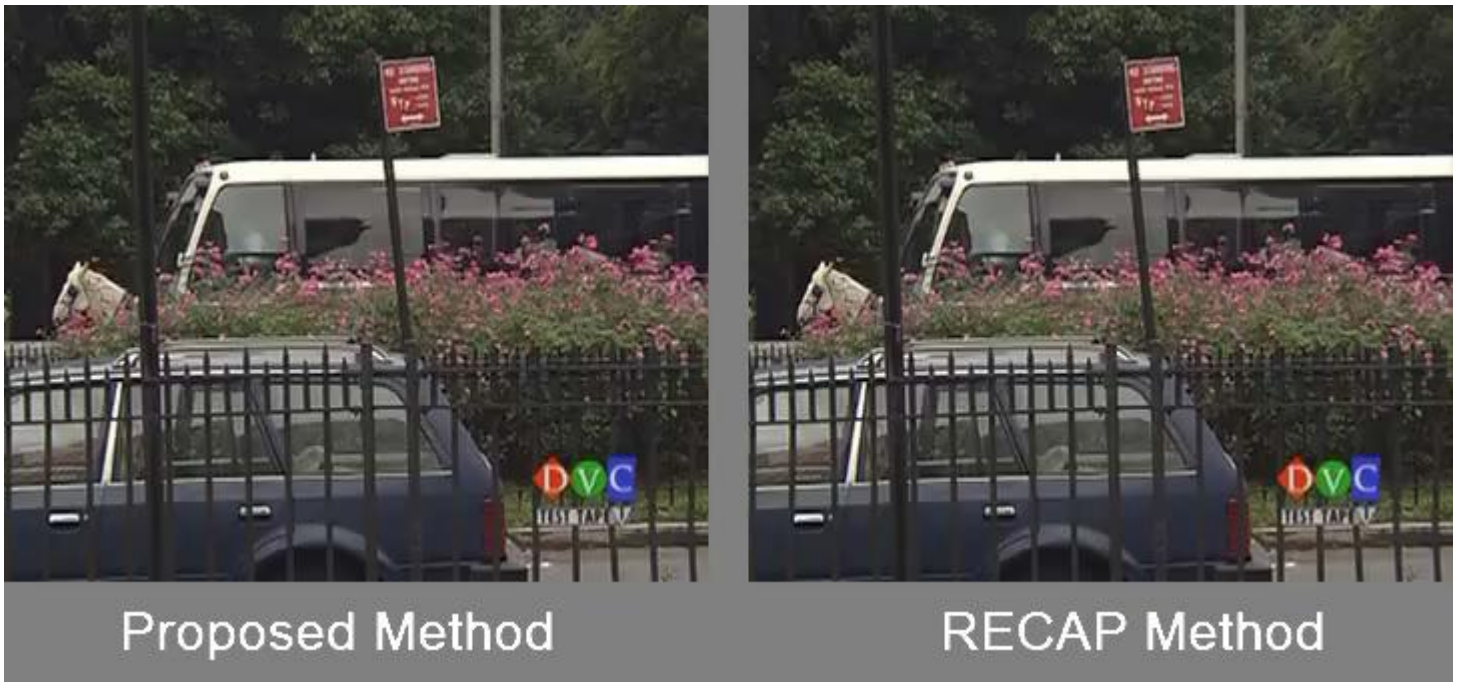
Where does this PSNR gain come from?



Our saliency prior helps us to find the right candidate.

Sample Results

bit rate=700kbps, loss Rate =10%



Conclusions

1. We introduced the concept of low-saliency prior for error concealment in a ROI-based video streaming application.

2. Low-saliency prior can increase the PSNR of reconstructed regions.

3. Reconstructed regions becomes less attention-grabbing.

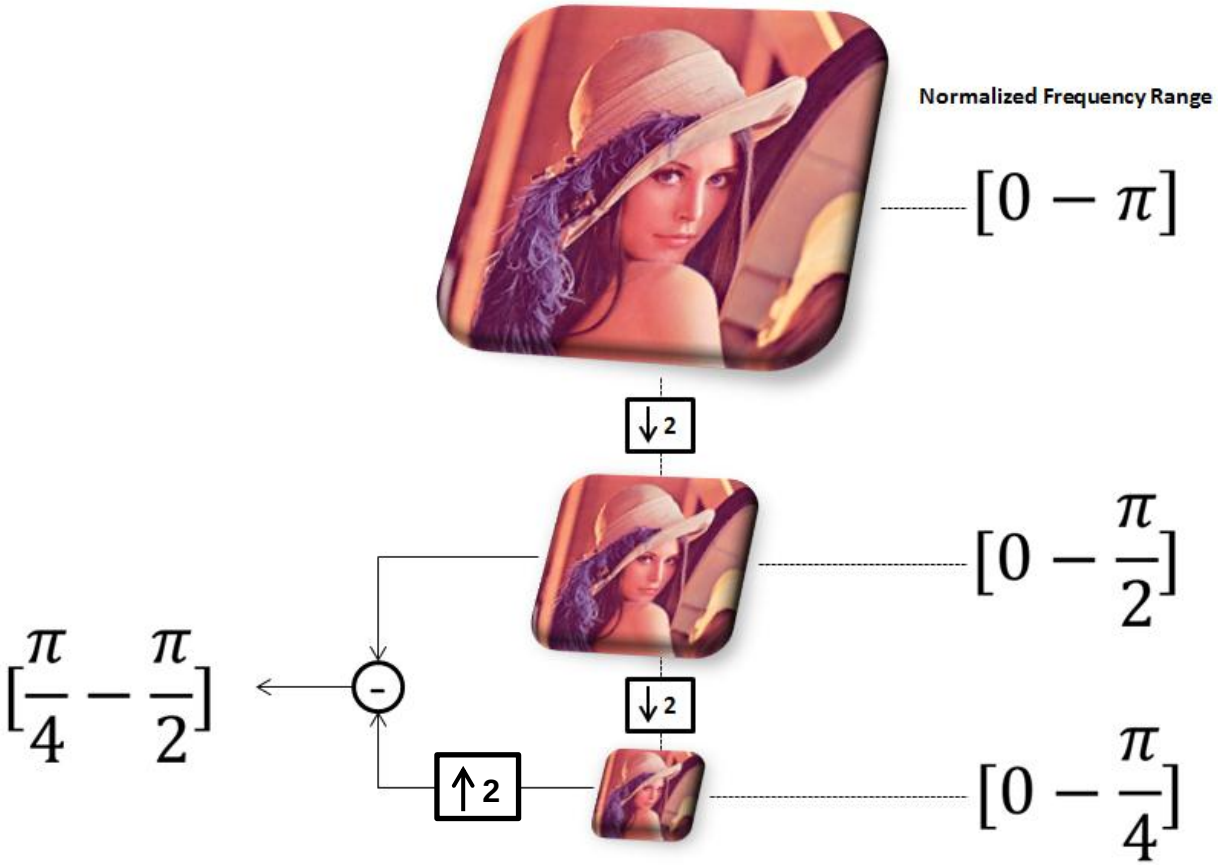
$$\min_{\boldsymbol{b}}\{fit\_err(\boldsymbol{b}) + \lambda\, saliency(\boldsymbol{b})\}$$

Thank You!
Any Question?

Presenter: Hadi Hadizadeh
hha54@sfu.ca

Supplementary Slides

Frequency Range



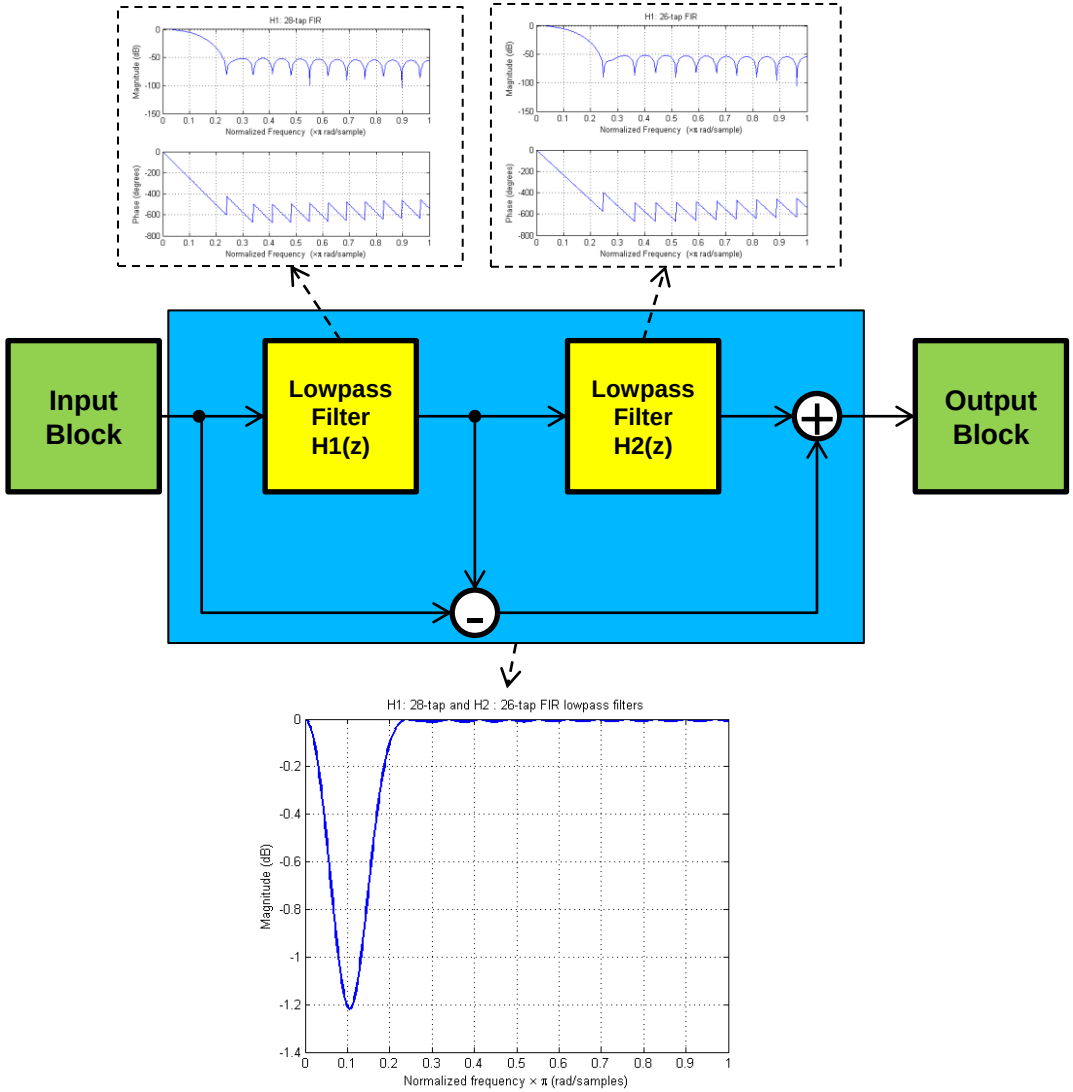
Subjective Tests

- Number of Participants: **17**
- A two-sided **chi-square test** was used to examine the statistical significance of the results.

  - Risk Level : 95% (**p-value = 0.05**)
  - Null-hypothesis: the two methods are the same.

Video	RECAP	Proposed	*p*-value
Bus	7 votes	27 votes	0.006

Notch Filter



Adaptive Window for Saliency Computation