Enhancement and Artifact Removal for Transform Coded Document Images

Tak Shing Wong

School of Electrical and Computer Engineering
Purdue University

April, 2010

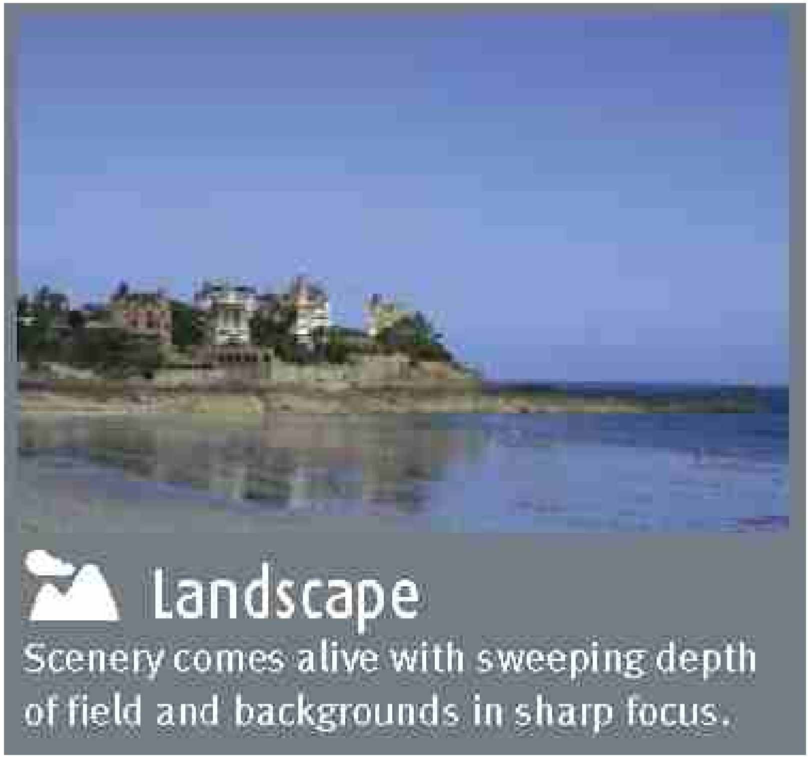
Outline

- **Overview**
  - The problems with JPEG decoding.
  - Rational and related works.
- **A document image model and estimation algorithm for optimized JPEG decompression[1].**
  - A Bayesian reconstruction approach.
  - A document image is partitioned into 3 classes: background, text, and picture.
  - Each class is characterized by a specific prior model for improved decoding quality.
- **Optimized JPEG decoding of document images using soft classification.**
  - A purely post-processing approach.
  - No segmentation is needed.
  - Introduced the Hypothesis Selection Filter (HSF) as a new approach for image enhancement.

[1] Tak-Shing Wong, Charles A. Bouman, Ilya Pollak, and Zhigang Fan, "A Document Image Model and Estimation Algorithm for Optimized JPEG Decompression," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2518-2535, Nov. 2009.

The Problems with JPEG Decoding

Original



JPEG encoded and decoded

- JPEG encoding is lossy and leads to severe image artifacts.
- Loss of high frequency content causes ringing artifacts, as apparent in the text region.
- Loss of low frequency content causes blocking artifacts, as apparent in the sky region.
- These artifacts are inconsistent to human interpretation, thus they are easily noticeable.

Rational and Related Works

- Rational:
  - In practice, JPEG is also commonly used with document images.
  - JPEG is still superior in simplicity over the more advanced schemes, e.g. JPEG 2000, DjVu, MRC.
- Objective:
  - Assume document images are compressed by the traditional JPEG encoder.
  - Improve the decoding quality of any such JPEG-compressed document images.
- Important, related approaches which aim at natural images:
  - Projection onto convex sets[1,2]
  - Bayesian reconstruction[3]
  - Adaptive post-processing[4]
- Relatively few works focus on document images.

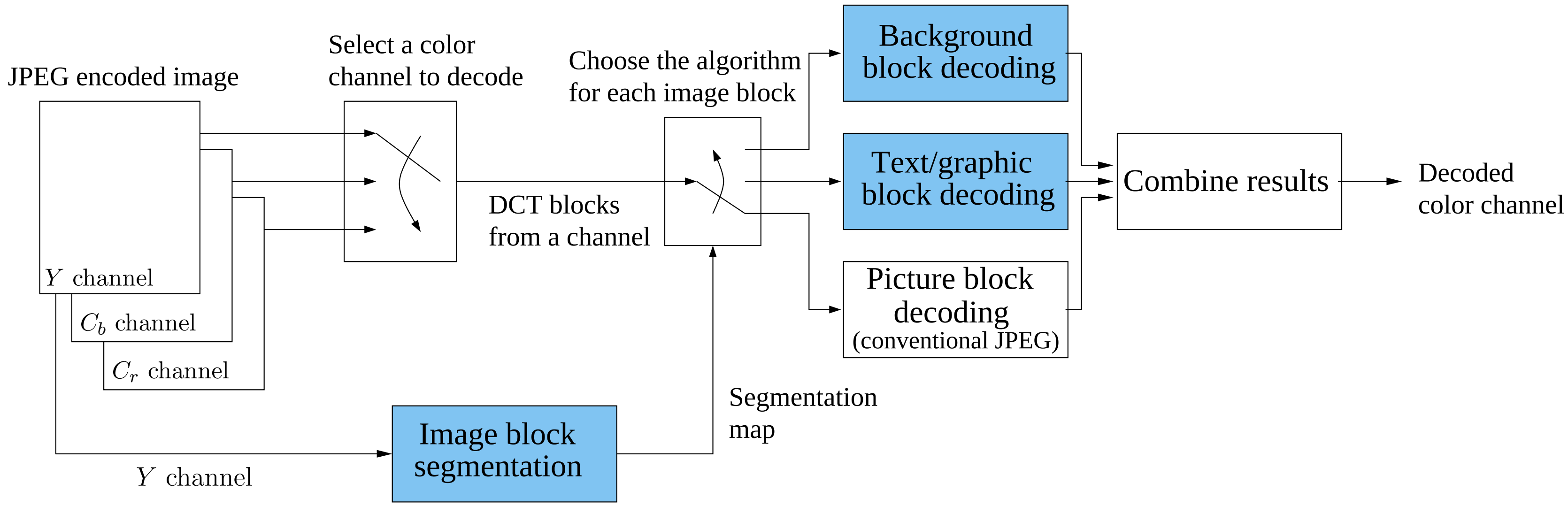
[1] A. Zakhor, “Iterative procedures for reduction of blocking effects in transform image coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 2, no. 2, pp. 91-95, Mar. 1992.

[2] Y. Yang, N. Galatsanos, and A. Katsaggelos, “Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, no. 6, pp. 421-432, Dec. 1993.

[3] T. O’Rourke and R. Stevenson, “Improved image decompression for reduced transform coding artifacts,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, vol. 6, pp. 490-499, Dec. 1995.

[4] A. Averbuch, A. Schclar, and D. Donoho, “Deblocking of block-transform compressed images using weighted sums of symmetrically aligned pixels,” *IEEE Transactions on Image Processing*, vol. 14, no. 2, pp. 200-212, Feb. 2005.

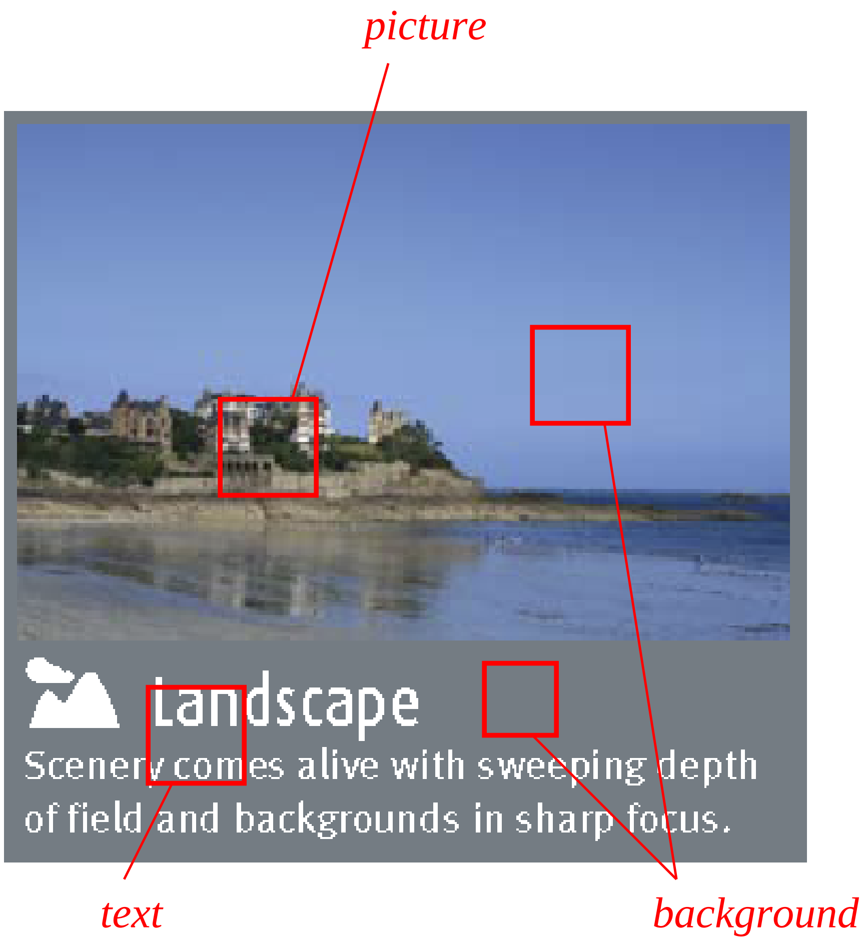
First Approach – A Document Image Model and Estimation Algorithm for Optimized JPEG Decompression



- Segment document into regions with distinct statistical attributes
  - Background
  - Text/graphics
  - Picture
- For each region, design a decoding strategy optimized for that content

Classes of Content

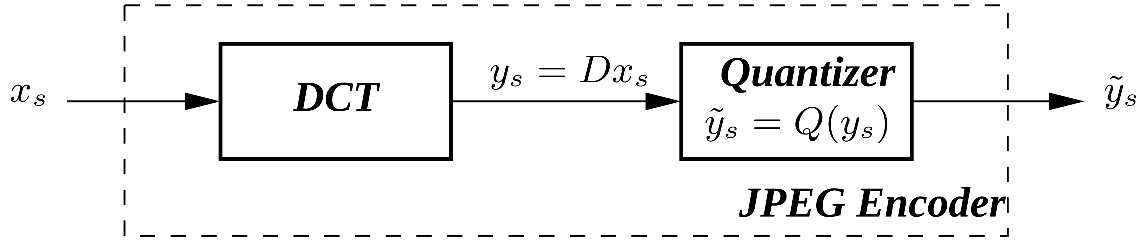
- Decompose image blocks into three classes:
  - Background
    - Background and smooth regions
    - Low AC energy
    - Vulnerable to blocking artifacts
  - Text/graphic
    - High density of sharp edges
    - Susceptible to ringing artifacts
  - Picture
    - Non-smooth part of natural images
    - Conventional JPEG decoding



JPEG Decoding as an Inverse Problem

- JPEG decoding as an inverse problem
  - Solution is non-unique $\Rightarrow$ inverse is ill-posed.
    - JPEG quantization is a many-to-one mapping.
    - Many image blocks would be quantized to the same set of DCT coefficients.
  - Prior model should be different for background, text, and picture regions
  - Maximum a posteriori (MAP) estimate can be computed using optimization

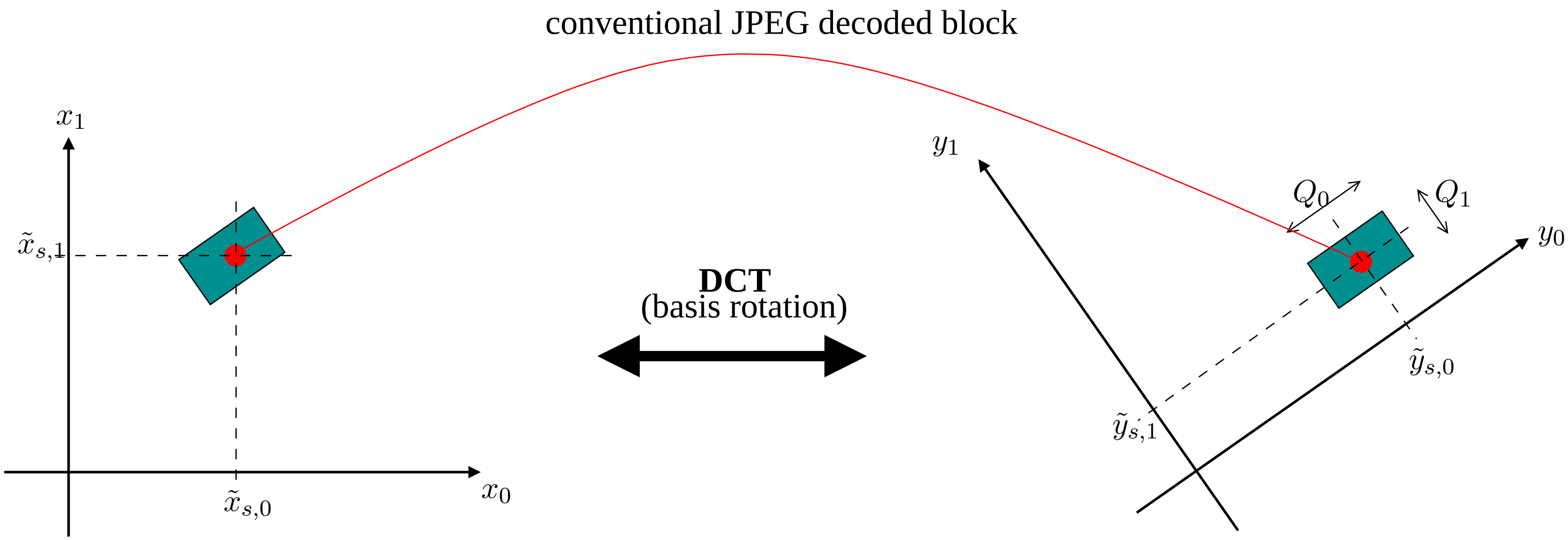
- Forward model for JPEG encoding



  - Probability of data is either 1 or 0

$$p(\tilde{y}_s | x_s) = \begin{cases} 1, & \text{if } Q(Dx_s) = \tilde{y}_s \\ 0, & \text{otherwise} \end{cases}$$

MAP Reconstruction



Spatial domain

$\tilde{x}_{s,i}$: i-th pixel of conventional JPEG decoding

■ : possible estimate of x_s

Frequency domain

$\tilde{y}_{s,i}$: i-th quantized DCT coefficient

Q_i : i-th quantization step size

- The MAP reconstruction of the document is given by

$$\hat{x}_s = \arg\min_{x_s} \{ -\log p(\tilde{y}_s | x_s) - \log p(x_s) \} \tag{1}$$

$$\hat{x}_s = \arg\min_{x_s : Q(Dx_s) = \tilde{y}_s} -\log p(x_s) \tag{2}$$

  - The forward model is a binary constraint for the estimate of x_s

Prior Model for Luminance Text Block

- Model for text
  - Each JPEG block consists of two predominant colors

    c_1 - background color

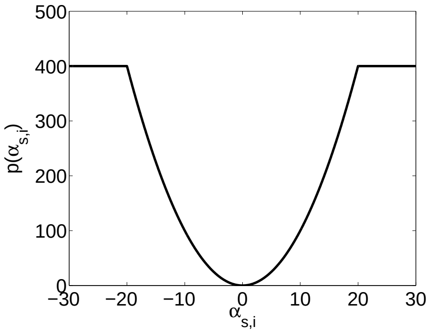
    c_2 - foreground color

  - Colors are mixed together for each pixel using an alpha channel α_i

    $$x_i = \alpha_i c_1 + (1-\alpha_i)c_2 + \text{noise}, \qquad 0 \le \alpha_i \le 1$$

- Distribution of pixels

$$p(x|c_1, c_2, \alpha) = \frac{1}{\text{const}} \exp\left\{ -\frac{1}{2\sigma_W^2} \sum_{i \text{ in block}} |x_i - \alpha_i c_1 - (1-\alpha_i)c_2|^2 \right\}$$



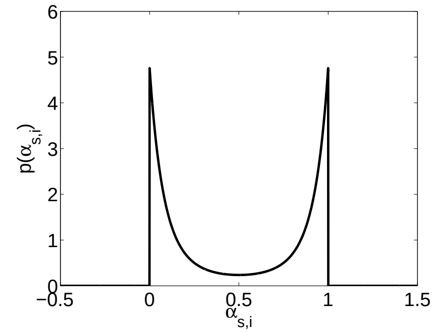
- The foreground and background colors of image blocks form a MRF

$$p(c_1, c_2) = \frac{1}{\text{const}} \exp\left\{ -\frac{1}{2\sigma_C^2} \sum_{\substack{(r,s) \text{ are} \\ \text{neighbor text blocks}}} \left[\rho(c_{1,r} - c_{1,s}) + \rho(c_{2,r} - c_{2,s}) \right] \right\}$$

where $\rho(t) = \min(t^2, \tau^2)$

- The alpha values are i.i.d. on interval $[0, 1]$

$$p(\alpha_i) = \begin{cases} \frac{1}{\text{const}} \exp\left\{ \nu |\alpha_i - .5|^2 \right\}, & \text{if } \alpha_i \in [0,1] \\ 0, & \text{otherwise} \end{cases}$$



Prior Model for Luminance Background Blocks

- Model for background blocks
  - Each background block is characterized by its DC coefficient
  - DC coefficients are modeled as a Gaussian MRF
- Distribution of DC coefficients

$$p(x) = \frac{1}{\text{const}} \exp\Big(-\frac{1}{2\sigma_B^2} \sum_{\{r,s\} \in K_{bb}} |y_{s,0} - y_{r,0}|^2 \Big)$$

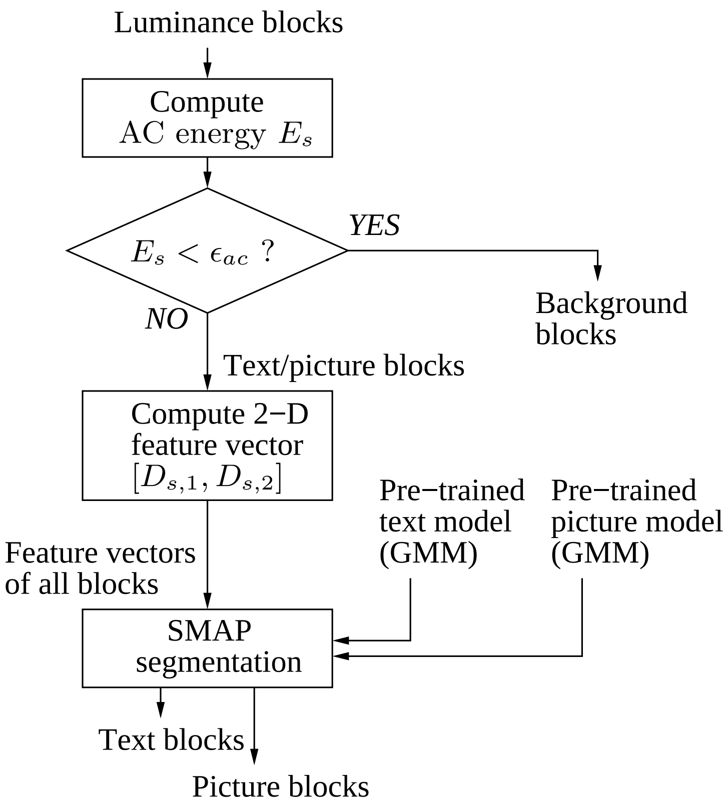
$y_{s,0}$ - DC coefficient for block s

Prior Model for Luminance Picture Blocks

- Model for Picture Blocks
  - Each block contains irregular natural image content
- MAP reconstruction
  - Decode by the conventional JPEG decoding
  - Less vulnerable to artifacts than text and background areas

Image Block Segmentation

- Identify background blocks using AC energy
- Classify the remaining blocks into text and picture blocks
    - Compute 2-D feature vector for each block (describe in next slide)
    - Feature vectors of each class is characterized by a pre-trained Gaussian mixture model (GMM).
    - Segment text and picture blocks using SMAP[1] and the pre-trained Gaussian mixture models.



[1] C. A. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," *IEEE Transactions on Image Processing*, March 1994

Feature Vector for Classification of Text and Picture Blocks

- Feature vector
  - Feature 1: Encoding length[1,2]

$$D_{s,1} = \lambda^{.5} \times \text{number of bits to encode the block } s$$

where
$$\lambda = \frac{\sum_{i=0}^{63} Q_i Q_i^*}{\sum_{i=0}^{63} Q_i^* Q_i^*} \qquad (3)$$

Q_i – quantization step sizes encoding the image (for the 64 DCT coefficients)

Q_i^* – default quantization step sizes defined by the JPEG standard

    - λ is a measure of the coarseness of the quantization step sizes Q_i.
    - If Q_i increases, bit-rate decreases and λ increases.
  - Feature 2: 2-color normalized block variance

$$D_{s,2} = \frac{1}{|\theta_{s,1} - \theta_{s,2}|^2} \sum_{i \text{ in block}} \min\left(|\tilde{x}_{s,i} - \theta_{s,1}|^2, |\tilde{x}_{s,i} - \theta_{s,2}|^2\right)$$

    - Pixels in block are clustered into two groups
    - $\tilde{x}_{s,i}$ the i-th pixel of the block s
    - $\theta_{s,1}$ and $\theta_{s,2}$ the means of the two clusters
- Text blocks tend to have large feature 1 and small feature 2.

[1] R. L. de Queiroz, "Processing JPEG-compressed images and documents," *IEEE Transactions on Image Processing,* December 1998.

[2] K. Konstantinides and D. Tretter, "A JPEG variable quantization method for compound documents," *IEEE Transactions on Image Processing*, July 2000.

Enhancing the Decoding Quality of Chrominance Text Blocks

- Decode the chrominance text blocks with the luminance alpha channel:
  - Recall that in our text model:

    foreground, background colors

    $$x_i = \alpha_i c_1 + (1 - \alpha_i) c_2 + \text{noise}$$

    alpha channel

    i.e. each pixel is a convex combination of two colors

  - α_i tells us if the pixel i is in the foreground or the background.
  - For the chrominance components, α_i are no longer random.
  - Simply re-use/copy the luminance alpha channel for the chrominance components.
    - Decimate the luminance alpha channel for subsampled chrominance component.

- Interpolate the chrominance text blocks by combining
  1. decoded chrominance text blocks
  2. high resolution luminance alpha channel

Experiments

- Compare to two other decoding schemes
- Algorithm I[1]
  - A Bayesian reconstruction scheme
  - Prior model using a MRF
  - No segmentation is performed.
  - The same prior model for the whole image.
- Algorithm II[2]
  - Content dependent post processing
  - Two classes of contents: text and picture
  - Emphasis on ringing artifact removal in text region
  - Use the global histogram of the text blocks to
    - estimate the text background color
    - segment the foreground pixels from the background pixels
  - Problematic for text regions with different foreground/background colors

[1] T. O'Rourke and R. Stevenson, "Improved image decompression for reduced transform coding artifacts," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, vol. 6, pp. 490-499, Dec. 1995.

[2] B. Oztan, A. Malik, Z. Fan, and R. Eschbach, "Removal of artifacts from JPEG compressed document images," *Proc. SPIE Color Imaging XII: Processing, Hardcopy, and Applications*, vol. 6439, Jan. 2007.

Segmentation Results: Image 1

Region 3

Region 1

Region 2

Biomedical Engineering Highlights

BME's graduate program has doubled in size to more than 80 students.

BME's first undergraduate class will graduate this academic year.

15 primary BME faculty members will be joined by 10 more full-time faculty members in the next two years.

Cook Group Inc. has provided $750,000 to endow the Leslie A. Geddes Chair in Biomedical Engineering.

Vincent Walter

The building design encourages student and faculty interactions with informal and formal gathering spaces and strategically aligns research facilities to maximize the sharing of resources. The design "sheds the traditional teaching and spatial models of the past," says Kalevi Huotilainen of the architectural firm BSA LifeStructures, which designed the building.

The interconnected central instructional laboratory complex contains a wet-bench laboratory (cell and tissue biology), an instrumentation laboratory (mechanical and electrical testing), a tissue culture facility, and a microscope darkroom (light and fluorescence). A central prep room and instructional coordinator's office link the learning activities scheduled for all levels of undergraduate laboratories.

The "Flex Lab" instructional laboratory space—a centralized space for engineering design courses—features a "dance floor" arrangement, with services like electricity, gases, and water provided from the ceiling, that allows benches and other mobile equipment to be reconfigured as needed for prototype design and testing.

Optics laboratories are built on their own individual concrete slabs in the basement, isolating highly sensitive instrumentation from vibrations that could affect measurements.

Building the Biomed Knowledge Base

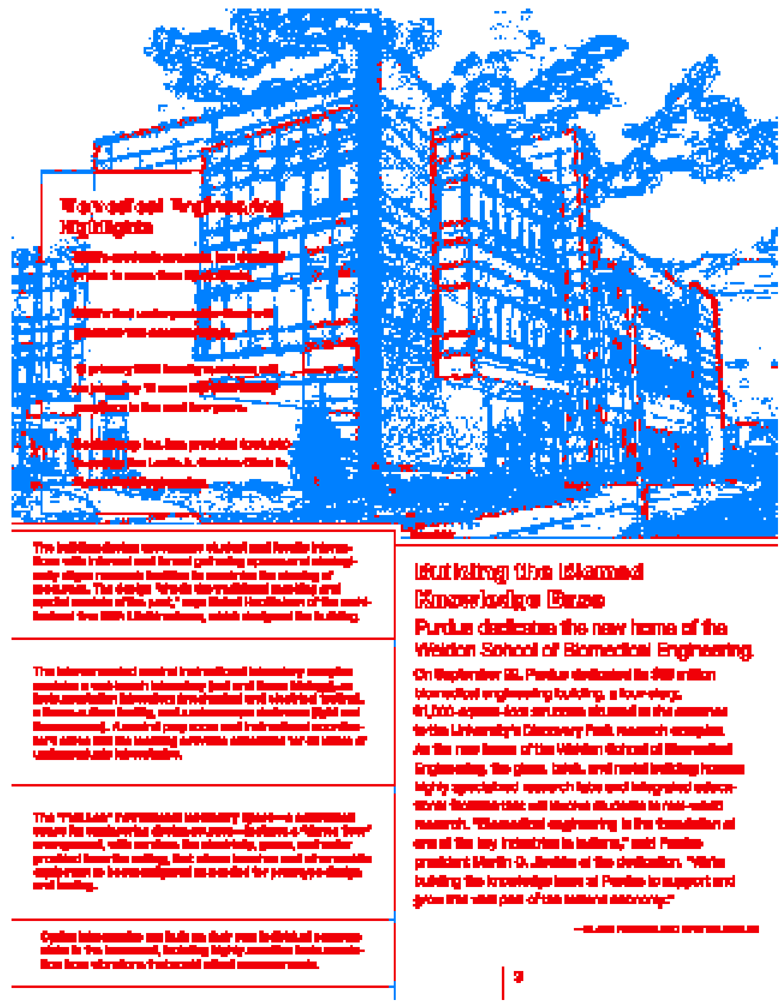
Purdue dedicates the new home of the Weldon School of Biomedical Engineering.

On September 22, Purdue dedicated its $25 million biomedical engineering building, a four-story, 91,000-square-foot structure situated at the entrance to the University's Discovery Park research complex. As the new home of the Weldon School of Biomedical Engineering, the glass, brick, and metal building houses highly specialized research labs and integrated educational facilities that will involve students in real-world research. "Biomedical engineering is the foundation of one of the key industries in Indiana," said Purdue president Martin C. Jischke at the dedication. "We're building the knowledge base at Purdue to support and grow this vital part of the Indiana economy."

—BLAKE POWERS AND CYNTHIA SEQUIN

3

Original of Image 1



Block Segmentation

: Background

: Text

: Picture

Text Decoding Results: Luminance Component

Luminance component of Region 1, Image 1

Conventional JPEG decoding

The proposed scheme

Text Decoding Results: Chrominance Component

Chrominance component (C_r) of Region 1, Image 1

Conventional JPEG decoding with pixel replication

The proposed scheme

Text Decoding Results: Region 1

Original of Region 1, Image 1

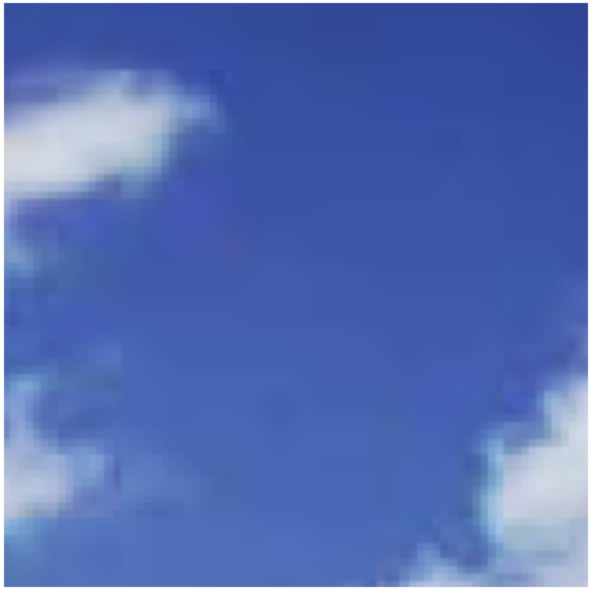
Conventional JPEG decoding

The proposed scheme

Algorithm I

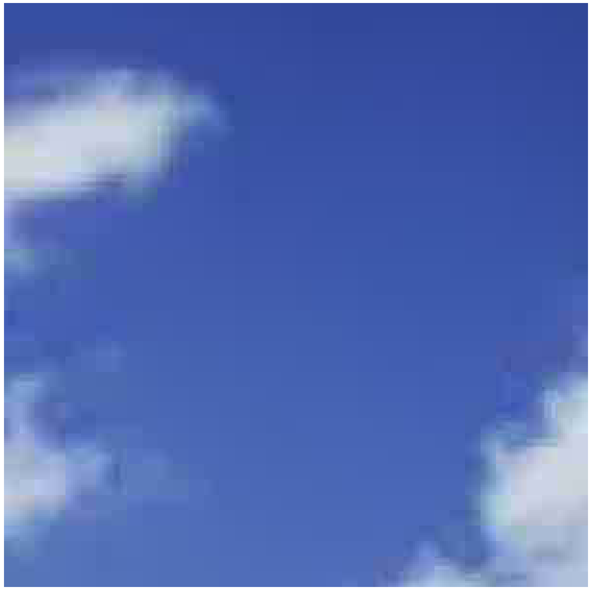
Algorithm II

Background Decoding Results: Region 3

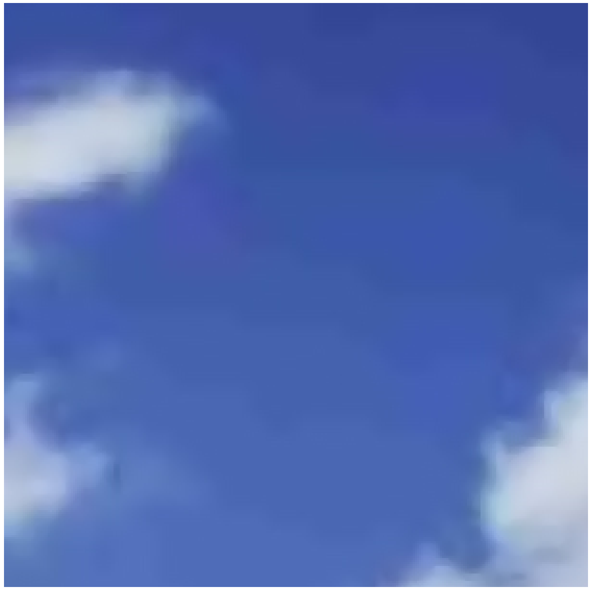


Original of Region 4, Image 1

Conventional JPEG decoding



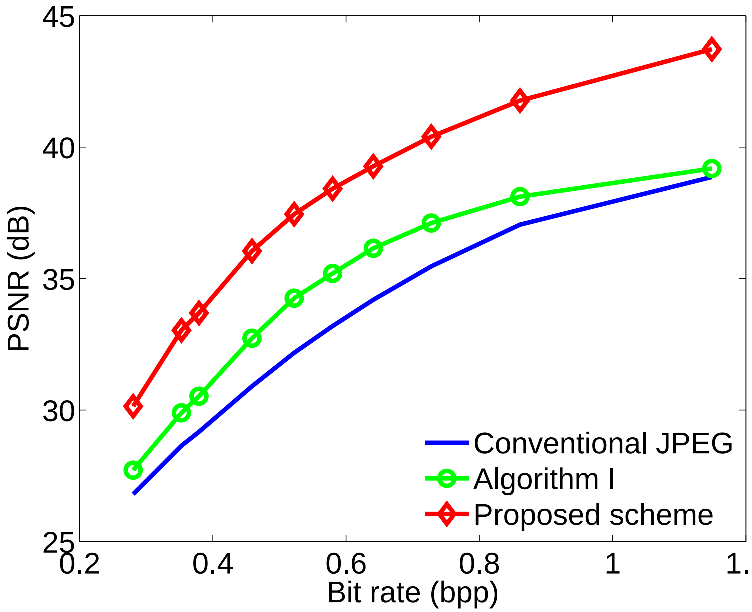
The proposed scheme



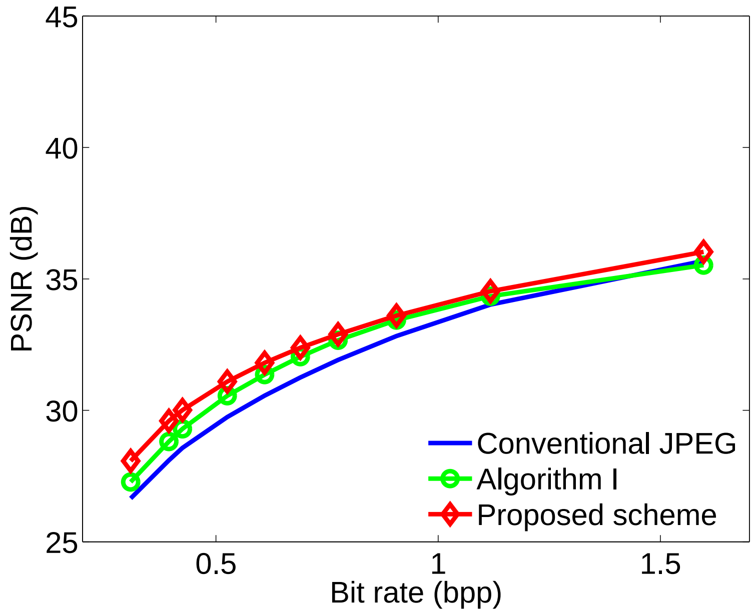
Algorithm I

Algorithm II

PSNR Comparison



PSNR vs bit-rate for 30 digital images



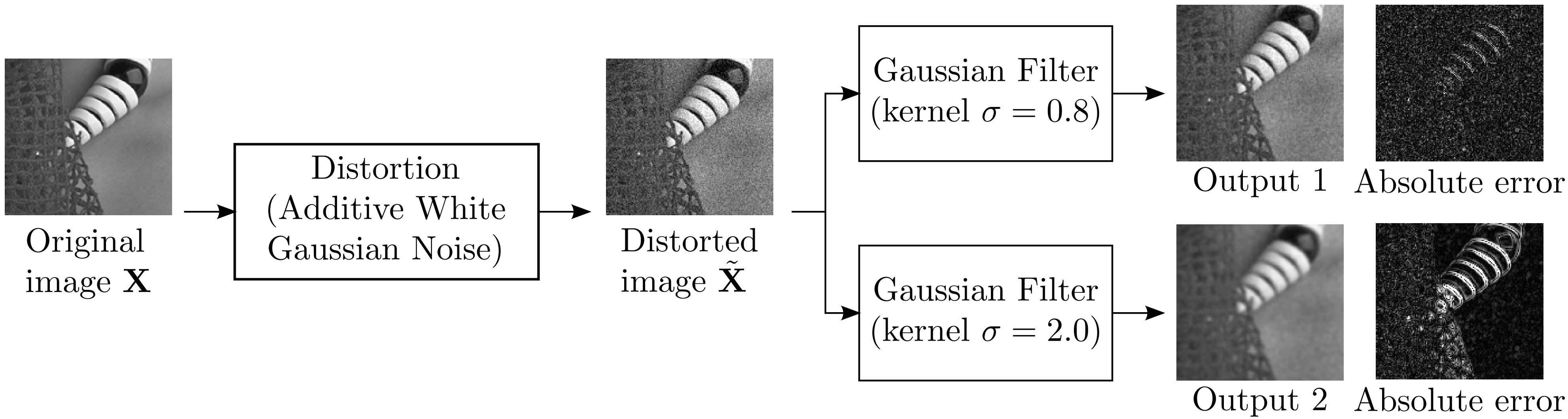
PSNR vs bit-rate for 30 scanned images

- Comparison using two set of images (30 digital images and 30 scanned images).
  - JPEG compressed at 10 different settings.
  - Compute average bit-rate at each setting.
  - Compute average PSNR for images decoded by each scheme at each setting.
- Lower PSNR improvement for scanned images despite of improved visual quality.
  - Removal of screening noise.
  - Sharpening of text.

Second Approach – Optimized JPEG Decoding of Document Images Using Soft Classification

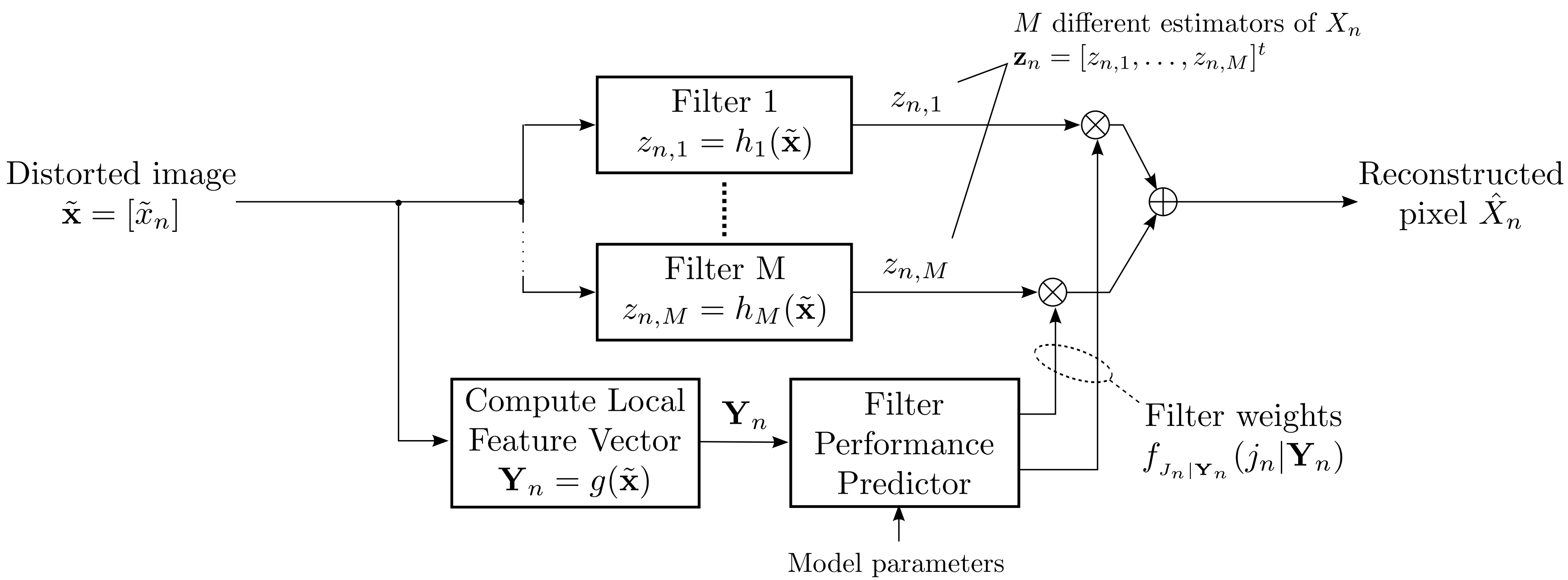
- Hypothesis Selection Filter (HSF) as a new approach for image enhancement.
  - Process the distorted image by several image filters.
  - Assume performance of image filters to be content dependent.
  - Use a local feature vector to determine the relative performance of the filters.
  - HSF provides a systematic approach to combine filtering results optimally.
- Apply the HSF as a post-processing step for JPEG decoding.
  - Decode the image using a conventional JPEG decoder.
  - Apply the HSF to eliminate the JPEG artifacts.

Motivation: A Simple Example



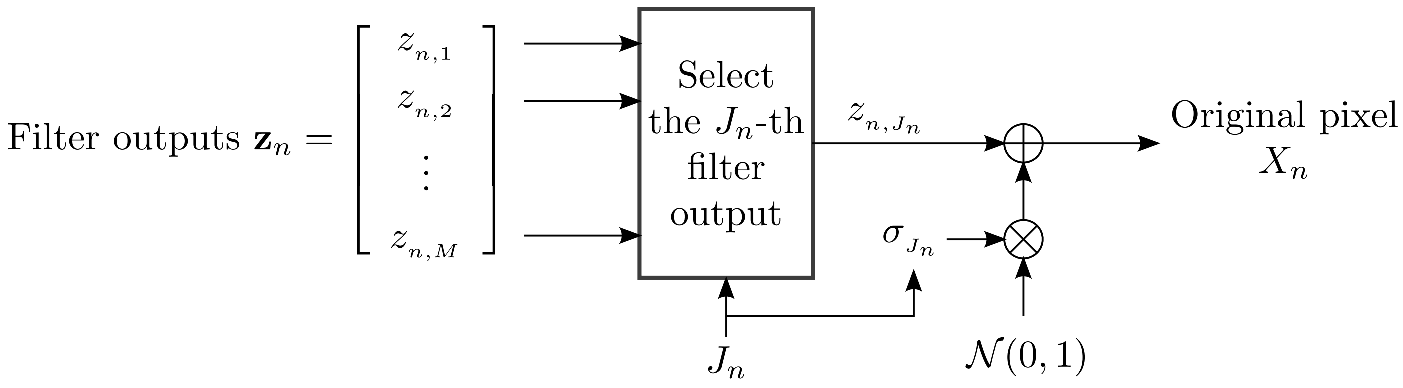
- Consider each filter as an estimator of the original image.
  - Filter performance is measured by the estimation error.
- Filter performance is usually content dependent.
  - Filter 1: Smaller error around edges.
  - Filter 2: Smaller error in smooth regions.
- Question: How do we combine the results of the adopted filters optimally?
  - A local feature vector for predicting filter performance.
  - A training procedure for designing an accurate predictor.

Structure of the Hypothesis Selection Filter



- $\tilde{\mathbf{x}} = [\tilde{x}_n]$ is a distorted version of the original image $\mathbf{X} = [X_n]$.
- Process the distorted image $\tilde{\mathbf{x}}$ using M different image filters.
  - Image filters as different estimators for the original pixel X_n.
- Compute a local feature vector $\mathbf{Y}_n$ to predict which filter produces the best estimate for X_n.
  - Accurate filters are assigned with larger weights.
- Combine filter outputs as a convex combination to form the enhanced pixel $\hat{X}_n$.

Probabilistic Model for Filter Error (Error Model)



- Introduce a hidden, discrete random variable $J_n \in \{1, \ldots, M\}$.
- Assume a generative model for the original pixel X_n.
  1. Sample J_n with prior probability

$$\text{Prob}(J_n = j) = \pi_j, \quad \text{for } j = 1, \ldots, M \tag{4}$$

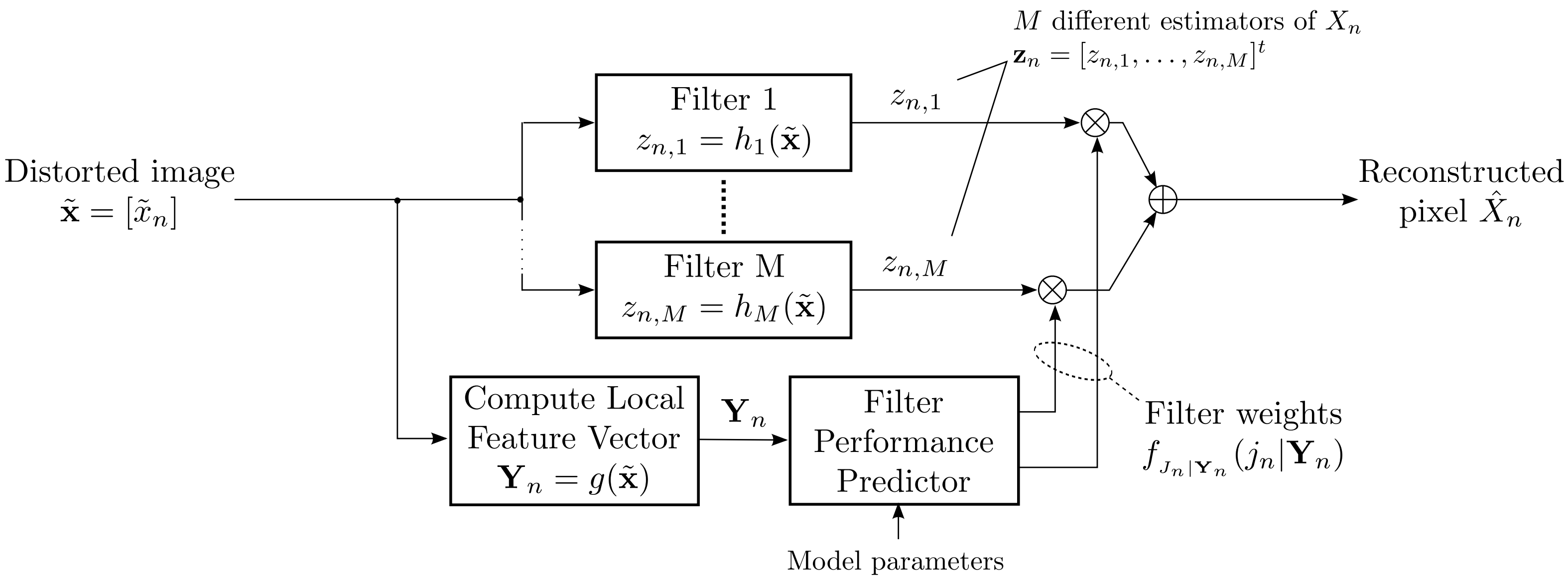
  2. Given $J_n = j$, model X_n as a sum of $z_{n,j}$ and an error component $\sim \mathrm{Normal}(0, \sigma_j^2)$.

$$f_{X_n|J_n}(x|j) = \mathrm{Normal}(x; z_{n,j}, \sigma_j^2) \tag{5}$$

- X_n follows a Gaussian mixture model, whose component means are the filter outputs:

$$f_{X_n}(x) = \sum_{j=1}^{M} \pi_j \ \mathrm{Normal}(x; z_{n,j}, \sigma_j^2) \tag{6}$$

Applying HSF to Images



- The HSF computes the MMSE estimate of X_n:

$$\hat{X}_n = E\left[X_n|\mathbf{Y}_n\right] \tag{7}$$

$$= \sum_{j=1}^{M} E\left[X_n|J_n = j, \mathbf{Y}_n\right] \; f_{J_n|\mathbf{Y}_n}(j|\mathbf{Y}_n) \tag{8}$$

$$= \sum_{j=1}^{M} z_{n,j} \; f_{J_n|\mathbf{Y}_n}(j|\mathbf{Y}_n) \tag{9}$$

- How do we compute $f_{J_n|\mathbf{Y}_n}(j|\mathbf{Y}_n)$?

How Do We Compute the Filter Weights $f_{J_n|\mathbf{Y}_n}(j|\mathbf{Y}_n)$?

- Model the conditional distribution of $\mathbf{Y}_n$ given $J_n = j$ by a Gaussian mixture model (GMM)

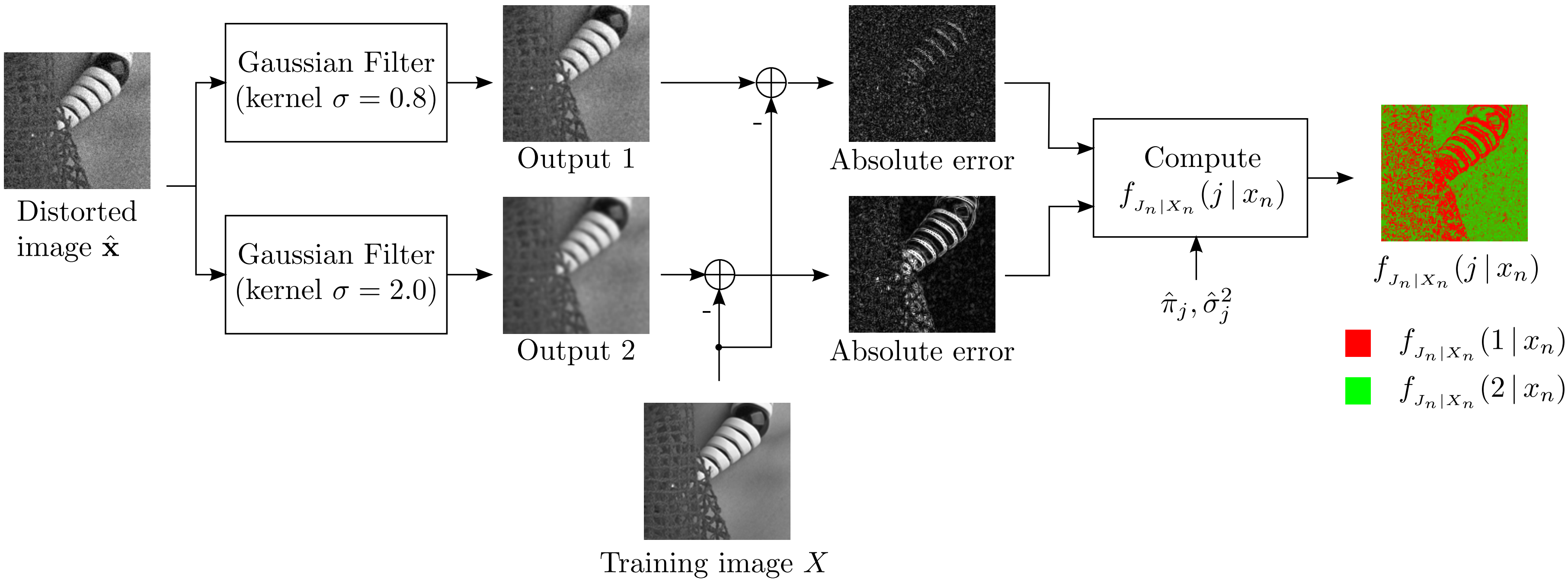
$$f_{\mathbf{Y}_n|J_n}(\mathbf{y}|j) = \sum_{l=1}^{K_j} \zeta_{j,l} \ \mathrm{Normal}(\mathbf{y}; \mathbf{m}_{j,l}, R_{j,l}), \quad \text{for } j = 1, \ldots, M \tag{10}$$

- Compute the filter weights by the Bayes' rule

$$f_{J_n|\mathbf{Y}_n}(j|\mathbf{Y}_n) = \frac{\pi_j \, f_{\mathbf{Y}_n|J_n}(\mathbf{Y}_n|j)}{\sum_{j'=1}^{M} \pi_{j'} \, f_{\mathbf{Y}_n|J_n}(\mathbf{Y}_n|j')} \tag{11}$$

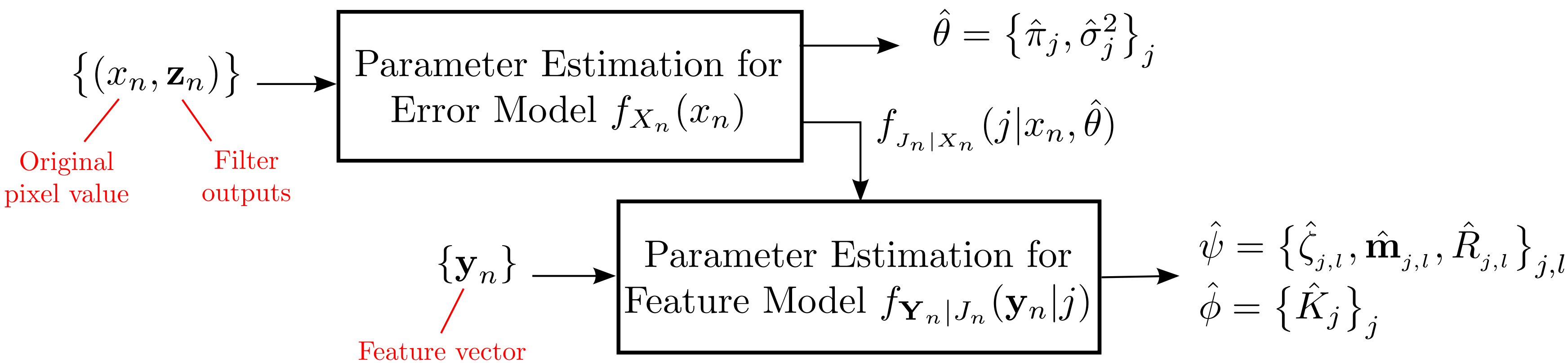
- Estimate the parameters of the Gaussian mixture models from training data.

How Do We Train?



- Notice that J_n is not available in training data.
- Solution (Soft-Training):
  - Compute the ML estimates of the parameters in the error model, $\hat{\pi}_j, \hat{\sigma}_j^2$.
  - Calculate $f_{J_n|X_n}(j|x_n)$ for each training sample x_n.
  - For each $j = 1, \ldots, M$, use $\mathbf{y}_n$ and $f_{J_n|X_n}(j|x_n)$ to estimate $f_{\mathbf{Y}_n|J_n}(\mathbf{y}|j)$

Parameter Estimation



- Estimates the following parameters:
  - Error model: $\theta = \{\pi_j, \sigma_j^2\}_j$
  - Feature model: $\psi = \{\zeta_{j,l}, \mathbf{m}_{j,l}, R_{j,l}\}_{j,l}$

    $\phi = \{K_j\}_j$ (orders of the GMM's)
- Training samples are realizations of $(x_n, \mathbf{z}_n, \mathbf{y}_n)$
- Use the EM algorithm to compute the ML estimates of θ and ψ.
- Determine K_j by the minimum description length criterion.
- Assume $f_{J_n|X_n,\mathbf{Y}_n}(j|x_n, \mathbf{y}_n) = f_{J_n|X_n}(j|x_n)$ to decomposes the EM algorithm into two parts.

Estimation of θ

1. Initialize π_j, σ_j^2 for $j = 1, \ldots, M$.

2. Update the class membership probability for each sample n, and for $j = 1, \ldots, M$:

$$f_{J_n|X_n,}(j|x_n, \theta) \leftarrow \frac{\pi_j \ \mathrm{Normal}(x_n; z_{n,j}, \sigma_j^2)}{\sum_{j'} \pi_{j'} \ \mathrm{Normal}(x_n; z_{n,j'}, \sigma_{j'}^2)}$$

3. Update the sufficient statistics for $j = 1, \ldots, M$:

$$\begin{aligned} N_j &\leftarrow \sum_n f_{J_n|X_n}(j|x_n, \theta) \\ t_j &\leftarrow \sum_n (x_n - z_{n,j})^2 f_{J_n|X_n,}(j|x_n, \theta) \end{aligned}$$

4. Re-estimate π_j and σ_j^2 for $j = 1, \ldots, M$

$$(\pi_j, \sigma_j^2) \leftarrow \left(\frac{N_j}{N}, \frac{t_j}{N} \right)$$

   where N is the number of training samples.

5. Iterate Step 2–4 until convergence.

Estimation of ψ

- Execute once for each value of j:

1. Initialize $\zeta_{j,l}$, $\mathbf{m}_{j,l}$, and $R_{j,l}$ for $l = 1, \ldots, K_j$.
2. Update the probability for each sample n and for $l = 1, \ldots, K_j$:

$$f_{J_n,L_n|X_n,\mathbf{Y}_n}(j,l|x_n,\mathbf{y}_n,\hat{\theta},\psi) \leftarrow f_{J_n|X_n}(j|x_n,\hat{\theta}) \frac{\zeta_{j,l} \ \mathrm{Normal}(\mathbf{y}_n; \mathbf{m}_{j,l}, R_{j,l})}{\sum_{l'} \zeta_{j,l'} \ \mathrm{Normal}(\mathbf{y}_n; \mathbf{m}_{j,l'}, R_{j,l'})}$$

3. Update the sufficient statistics for $l = 1, \ldots, K_j$:

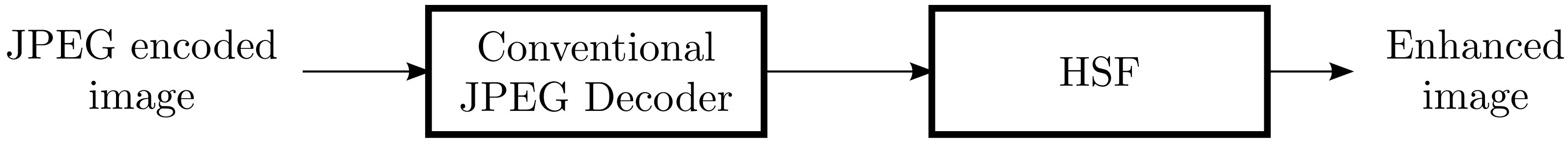
$$\begin{aligned} N_{j,l} &\leftarrow \sum_n f_{J_n,L_n|X_n,\mathbf{Y}_n}(j,l|x_n,\mathbf{y}_n,\hat{\theta},\psi) \\ \mathbf{u}_{j,l} &\leftarrow \sum_n \mathbf{y}_n f_{J_n,L_n|X_n,\mathbf{Y}_n}(j,l|x_n,\mathbf{y}_n,\hat{\theta},\psi) \\ \mathbf{v}_{j,l} &\leftarrow \sum_n \mathbf{y}_n \mathbf{y}_n^t f_{J_n,L_n|X_n,\mathbf{Y}_n}(j,l|x_n,\mathbf{y}_n,\hat{\theta},\psi) \end{aligned}$$

4. Re-estimate $\zeta_{j,l}$, $\mathbf{m}_{j,l}$, and $R_{j,l}$ for $l = 1, \ldots, K_j$:

$$(\zeta_{j,l}, \mathbf{m}_{j,l}, R_{j,l}) \leftarrow \left(\frac{N_{j,l}}{N}, \frac{\mathbf{u}_{j,l}}{N_{j,l}}, \frac{\mathbf{v}_{j,l}}{N_{j,l}} - \frac{\mathbf{u}_{j,l}\mathbf{u}_{j,l}^t}{N_{j,l}^2} \right)$$

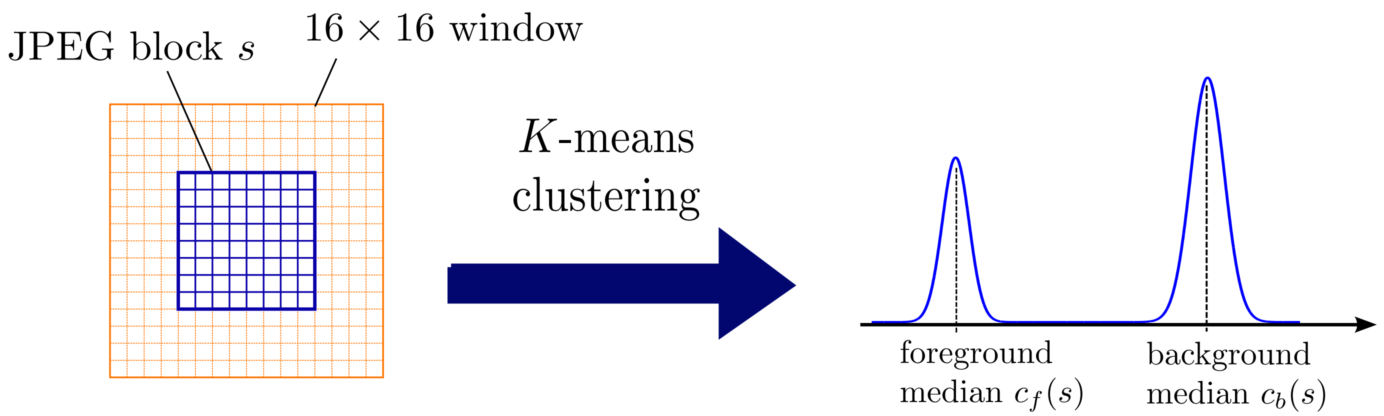
5. Iterate Step 2–4 until convergence.

HSF for JPEG Artifact Removal



- HSF Filters:
  - Filter 1 – Bilateral filter ($\sigma_{\text{spatial}} = .5, \sigma_{\text{range}} = 10$).
  - Filter 2 – Gaussian filter (kernel s.d. $\sigma = 1$).
  - Filter 3 – Local foreground gray-level of JPEG block.
  - Filter 4 – Local background gray-level of JPEG block.

Computing Local Foreground and Background Gray-Levels



- For each JPEG block s:
  - Center a 16×16 window around the JPEG block.
  - Cluster the 256 pixels in the window into two groups by K-means clustering.
  - The two cluster medians, $c_f(s)$ and $c_b(s)$, are the foreground and background gray-levels of the JPEG block s.

HSF Feature Vector, $\mathbf{Y}_n$

- Feature 1: Variance of the associated JPEG block.
- Feature 2: Two-color normalized variance of the associated JPEG block.

$$b_s = \frac{\sum_i \min(|u_{s,i} - c_b(s)|^2, |u_{s,i} - c_f(s)|^2)}{|c_f(s) - c_b(s)|^2}$$

$u_{s,i}$ – the i-th pixel of the block s.

$c_f(s)$ – median foreground gray-level of the block s

$c_b(s)$ – median background gray-level of the block s

- Feature 3: Gradient magnitude computed by Sobel operators.
- Feature 4: Difference between pixel and foreground gray-level.

$$y_{n,4} = u_n - c_f(s),$$

- Feature 5: Difference between pixel and background gray-level.

$$y_{n,4} = u_n - c_b(s).$$

Experiment and Comparison

- We compare our results with the following schemes:
    1. Bilateral filter[1] (BF), $\sigma_{\text{spatial}} = .5, \sigma_{\text{range}} = 10$.
    2. Bayesian reconstruction[2] (BR).
        - JPEG blocks segmented into 3 classes for text, picture, and smooth regions.
        - Specific prior models designed for text and smooth blocks.
    3. Resolution Synthesis[3] (RS).
        - Set scaling factor to 1.
        - Use JPEG encoded image to perform training.
    4. Improved Resolution Synthesis[4] (IRS).
        - Median filter to remove ringing artifacts.
        - Clip pixels to within local minimum and maximum to avoid overshooting and halo.
    5. HSF with Hard-Training (HSF-HT)
        - Use hard-classification during training.

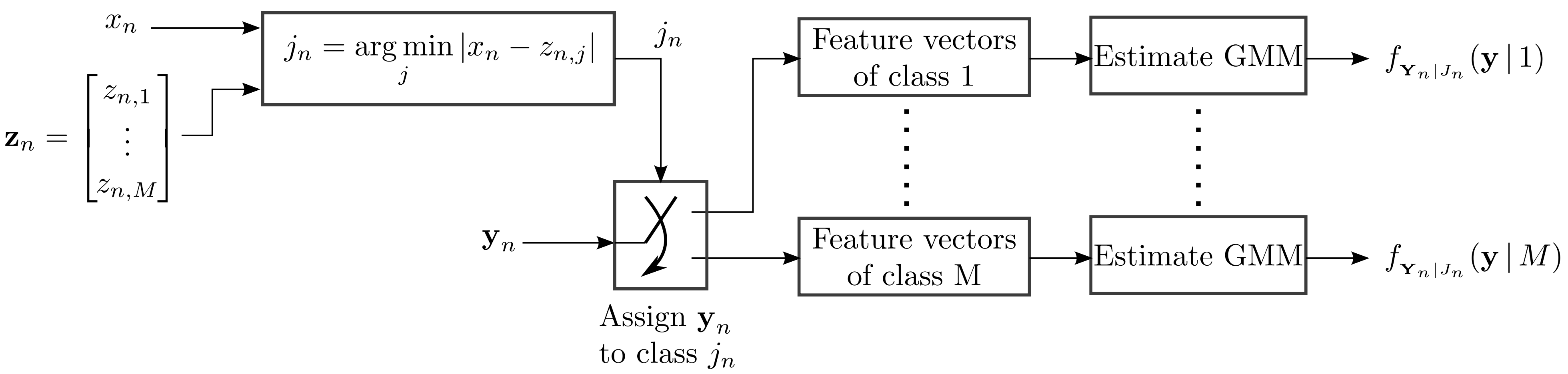
[1] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," *Proc. Int. Conf. Comput. Vis.*, 1998, pp. 839-846.

[2] T.-S. Wong, C. A. Bouman, I. Pollak, and Z. Fan, "A Document Image Model and Estimation Algorithm for Optimized JPEG Decompression," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2518-2535, Nov. 2009.

[3] C. B. Atkins, C. A. Bouman, J. P. Allebach, "Optimal image scaling using pixel classification" *IEEE Int'l Conf. on Image Proc.*, vol. 3, pp. 864-867, 2001.

[4] B. Zhang, J. S. Gondek, M. T. Schramm, and J. P. Allebach, "Improved Resolution Synthesis for Image Interpolation," *NIP22: International Conference on Digital Printing Technologies, Denver, CO* Sep. 2006. pp. 864-867, 2001.

HSF with Hard-Training (HSF-HT)



- During training
  - Classify each training sample using hard-classification

$$j_n = \arg\min_j |x_n - z_{n,j}|$$

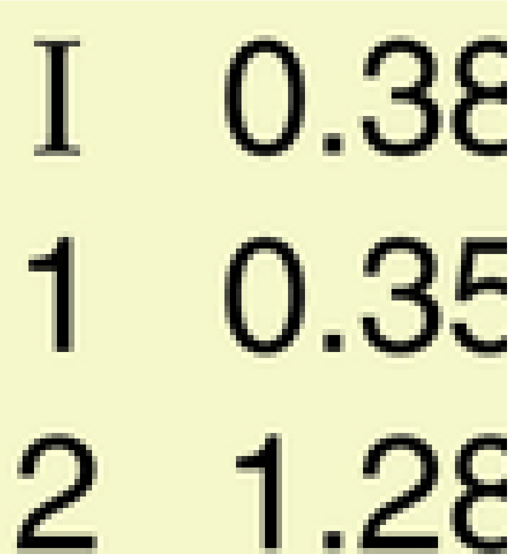
  - Partition the feature vectors into M sub-sets based on classification results.
  - Estimate a Gaussian mixture model from each sub-set.

Test Images

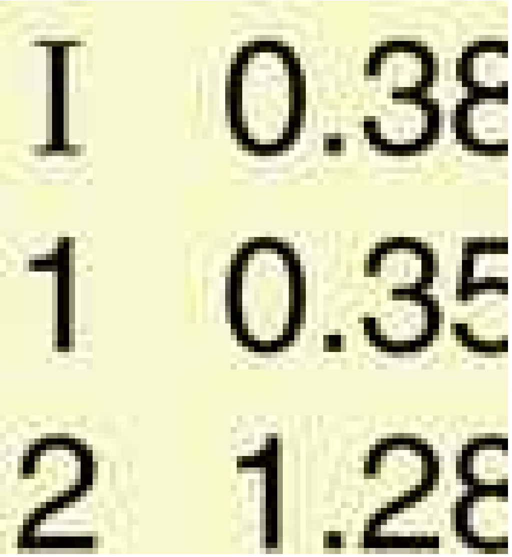
Thumbnail of Image 1, JPEG encoded at 0.69 bpp

Thumbnail of Image 2, JPEG encoded at 0.96 bpp

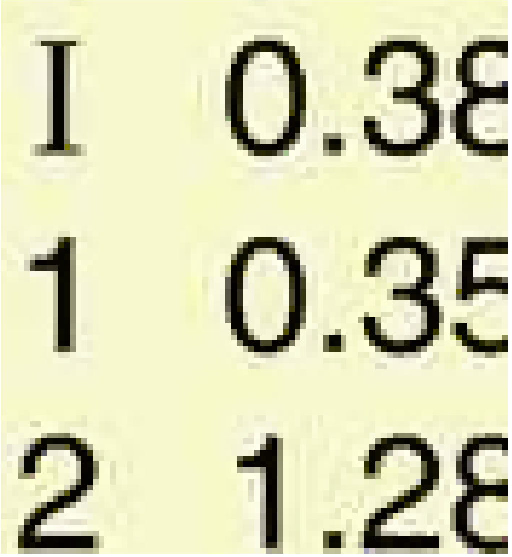
Results



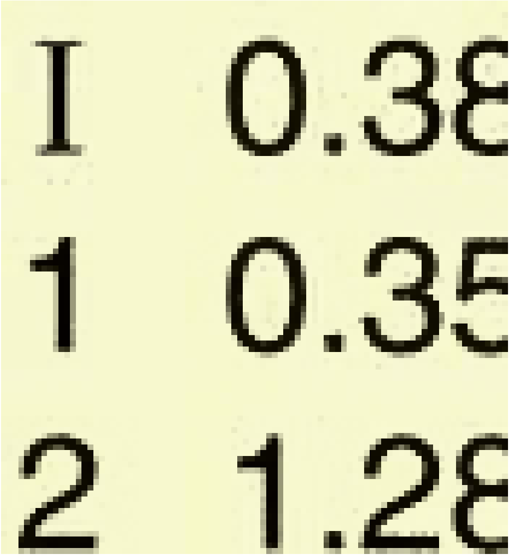
Original



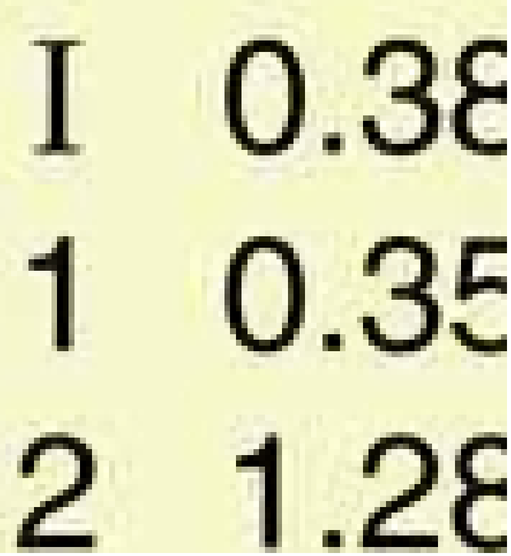
JPEG



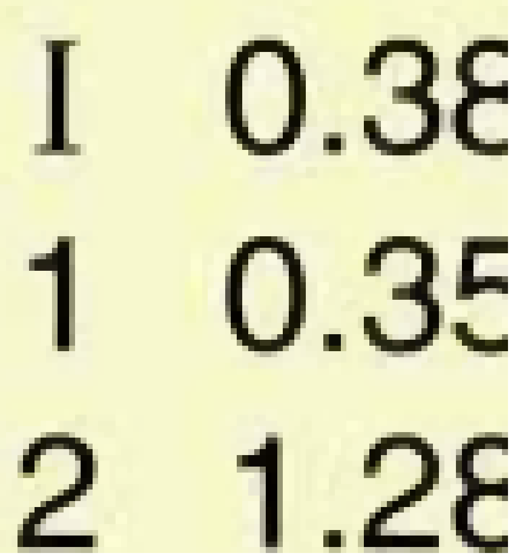
Bilateral Filter (BF)



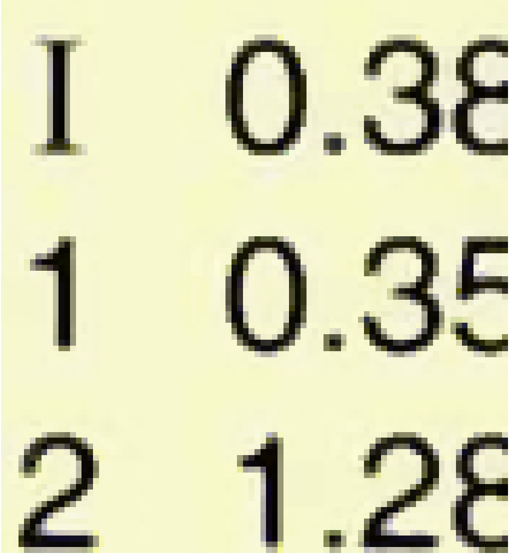
Bayesian Reconstruction (BR)



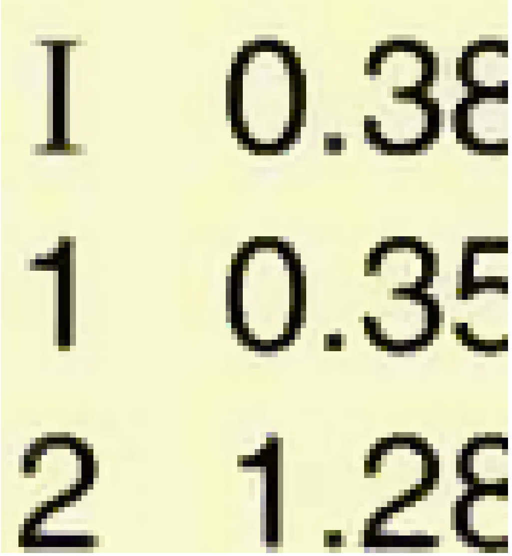
Resolution Synthesis (RS)



Improved Resolution Synthesis (IRS)



Hypothesis Selection Filter - Hard Training (HSF-HT)



Hypothesis Selection Filter (HSF)

Results

Original

JPEG

Bilateral Filter (BF)

Bayesian Reconstruction (BR)

Resolution Synthesis (RS)

Improved Resolution Synthesis (IRS)

Hypothesis Selection Filter - Hard Training (HSF-HT)

Hypothesis Selection Filter (HSF)

Results

Original

JPEG

Bilateral Filter (BF)

Bayesian Reconstruction (BR)

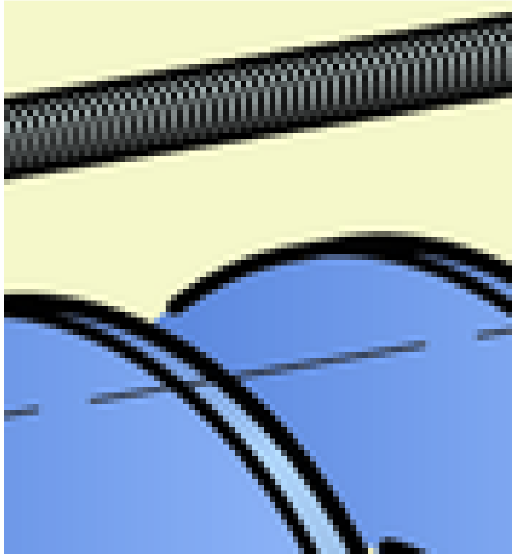
Resolution Synthesis (RS)

Improved Resolution Synthesis (IRS)

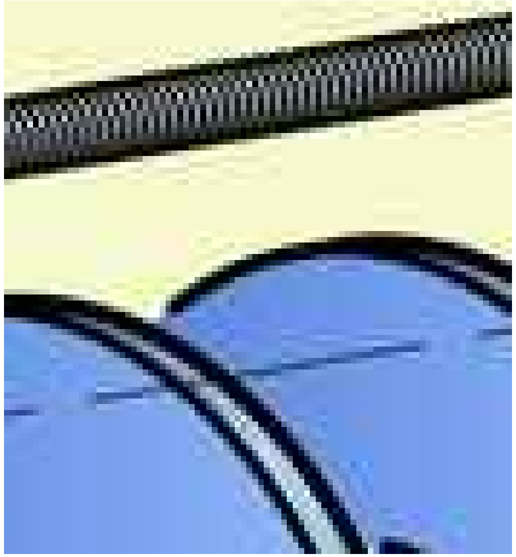
Hypothesis Selection Filter - Hard Training (HSF-HT)

Hypothesis Selection Filter (HSF)

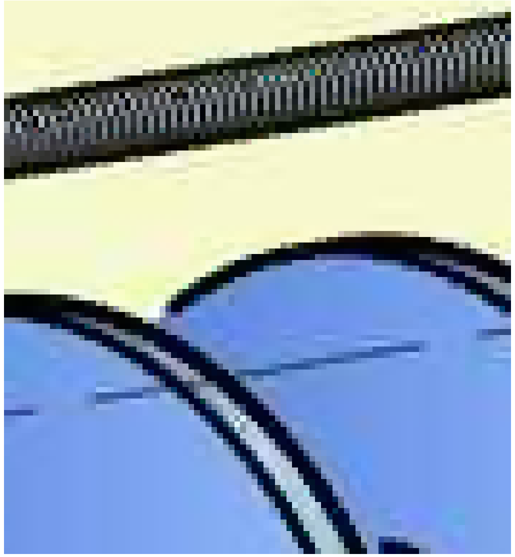
Results



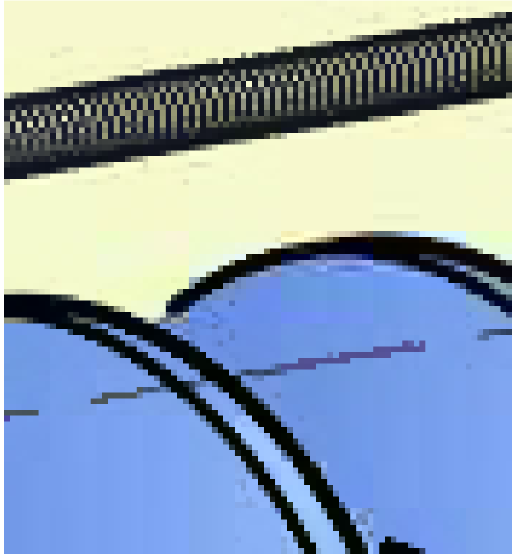
Original



JPEG

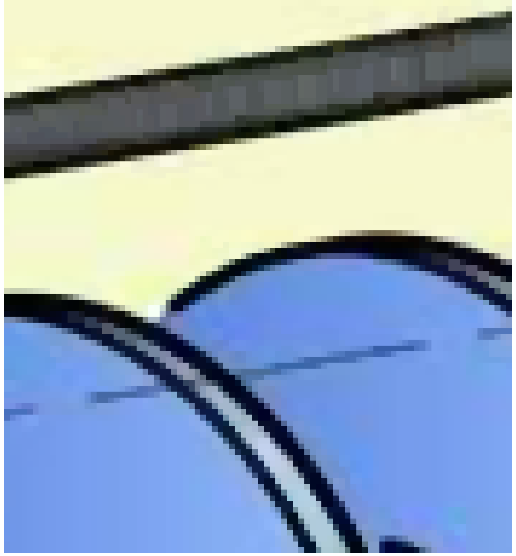


Bilateral Filter (BF)

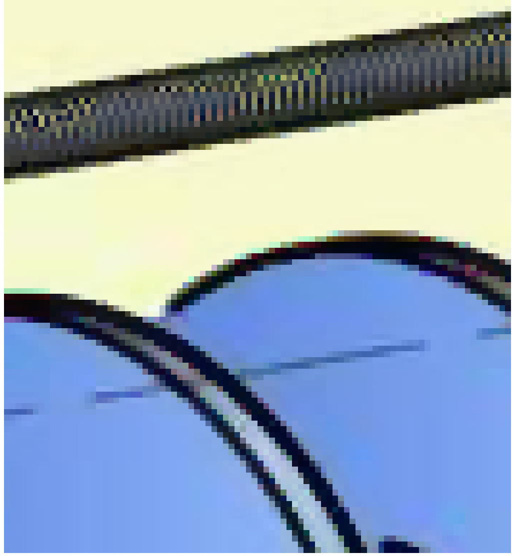


Bayesian Reconstruction (BR)

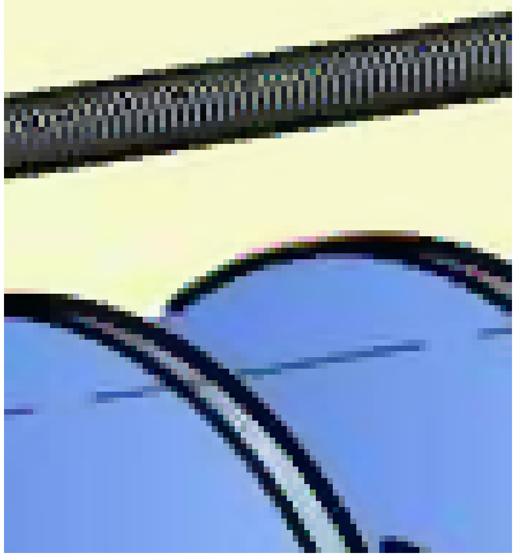
Resolution Synthesis (RS)



Improved Resolution Synthesis (IRS)



Hypothesis Selection Filter - Hard Training (HSF-HT)



Hypothesis Selection Filter (HSF)

Results


Original

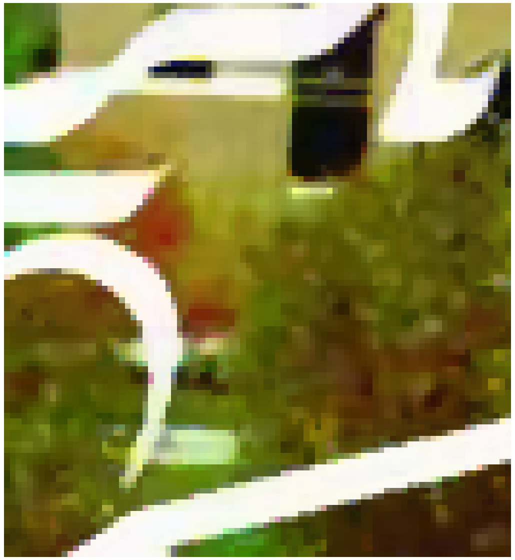

JPEG


Bilateral Filter (BF)


Bayesian Reconstruction (BR)

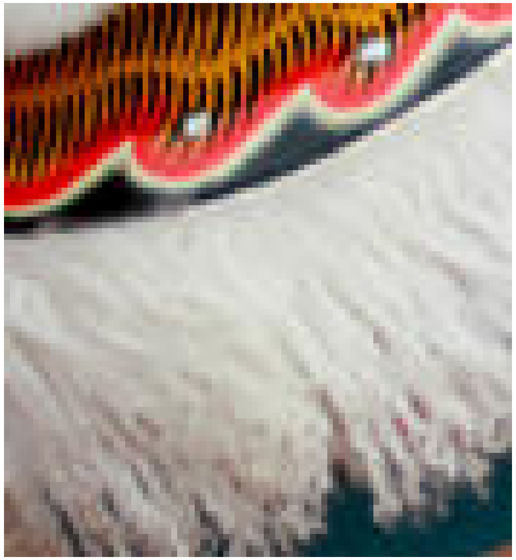

Resolution Synthesis (RS)


Improved Resolution Synthesis (IRS)

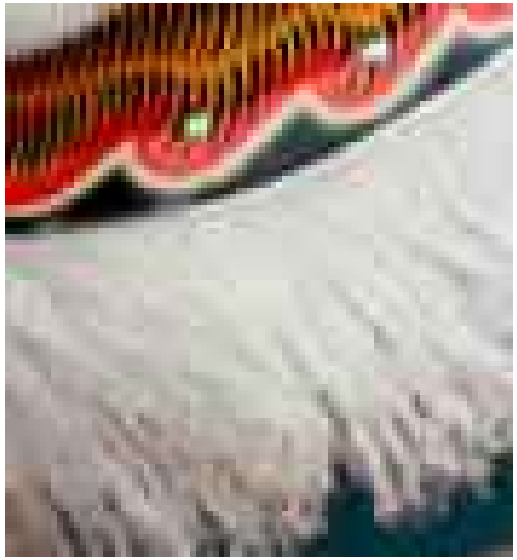

Hypothesis Selection Filter - Hard Training (HSF-HT)


Hypothesis Selection Filter (HSF)

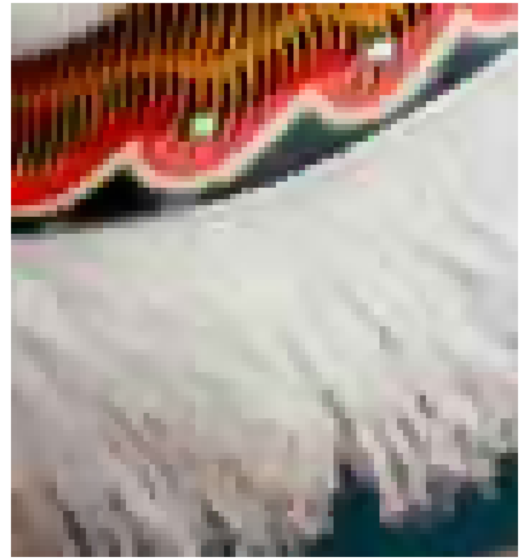
Results



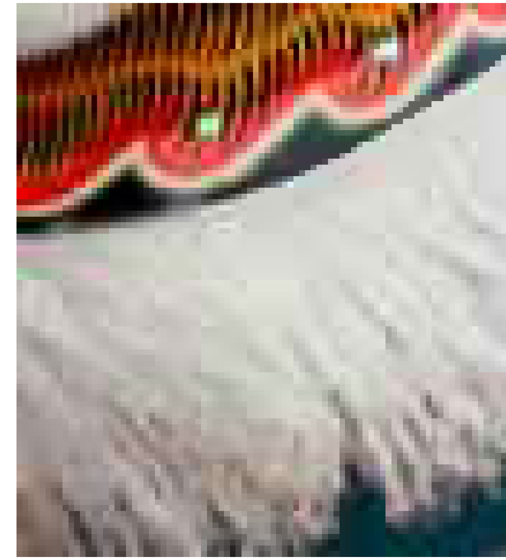
Original



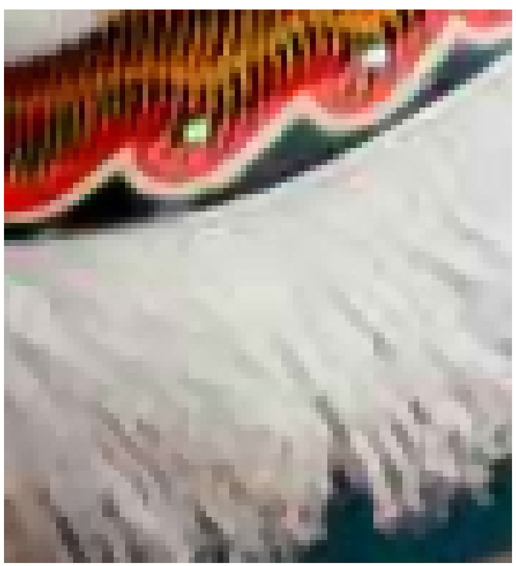
JPEG



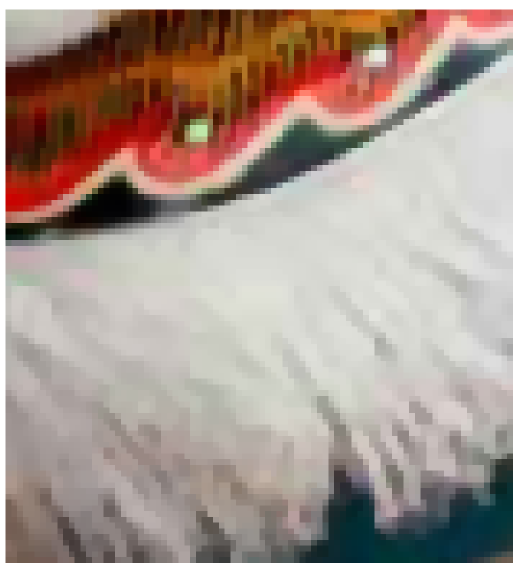
Bilateral Filter (BF)



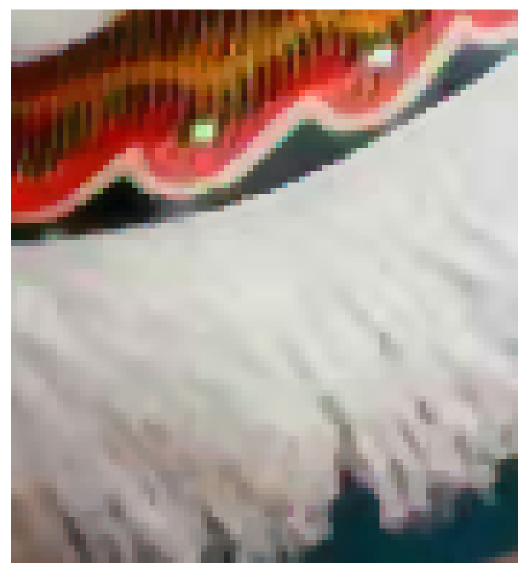
Bayesian Reconstruction (BR)



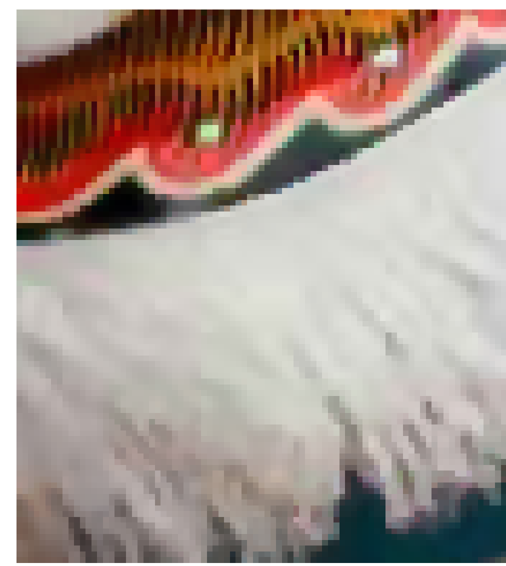
Resolution Synthesis (RS)



Improved Resolution Synthesis (IRS)

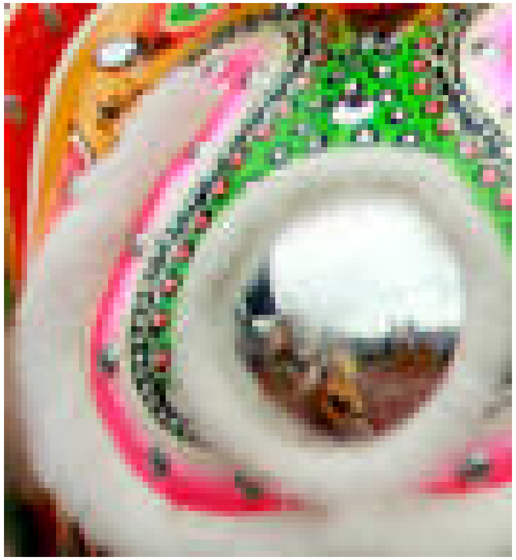


Hypothesis Selection Filter
- Hard Training (HSF-HT)



Hypothesis Selection Filter (HSF)

Results



Original

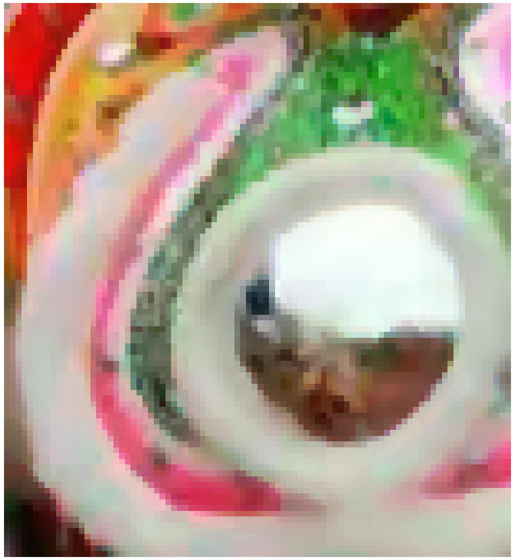
JPEG

Bilateral Filter (BF)

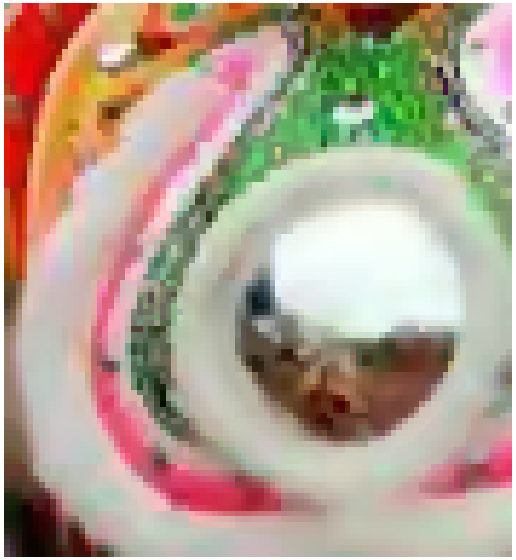
Bayesian Reconstruction (BR)

Resolution Synthesis (RS)

Improved Resolution Synthesis (IRS)



Hypothesis Selection Filter - Hard Training (HSF-HT)



Hypothesis Selection Filter (HSF)

Summary and Conclusion

- Document images contain different types of contents in complex layout.
- Using the specific properties of each content type, JPEG encoded document images can be decoded at high quality.
    - Reduced ringing artifacts.
    - Reduced blocking artifacts.
    - Improved sharpness.
- A Bayesian reconstruction approach:
    - Used segmentation to separate out different content types
    - Applied different prior models to make use of the unique characteristics of the different content types.
- A post-processing approach - the Hypothesis Selection Filter (HSF):
    - A systematic method for combining the advantages of multiple user selected flters.
    - Each filter is applied to the regions for which it is most appropriate.
    - Demonstrated the HSF by applying it as a post-processing step for JPEG artifact removal.
        - Used four filters: bilateral, Gaussian, foreground/background gray-level estimate.
        - Significantly reduced the ringing artifacts in text/graphics regions and blocking artifacts in natural images.
    - The HSF has a generic structure which is potentially applicable to other applications.