Document Image Segmentation and Compression *

Hui Cheng
Major Professor: Charles A. Bouman

School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana 47907-1285

*This research was supported by Xerox Corporation.

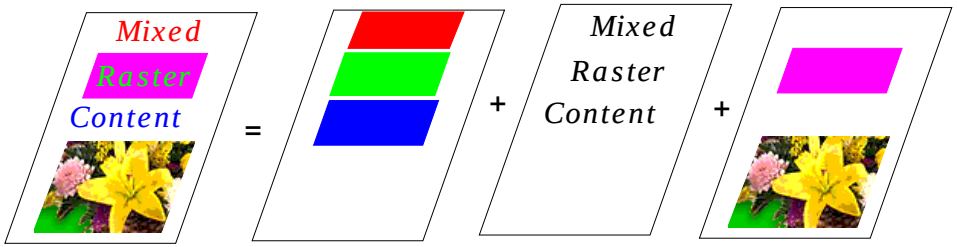
Outline

- Trainable Sequential MAP (TSMAP) segmentation algorithm
- Multilayer document image compression algorithm
- Rate-Distortion Optimized Segmentation (RDOS) algorithm

Document Image Compression

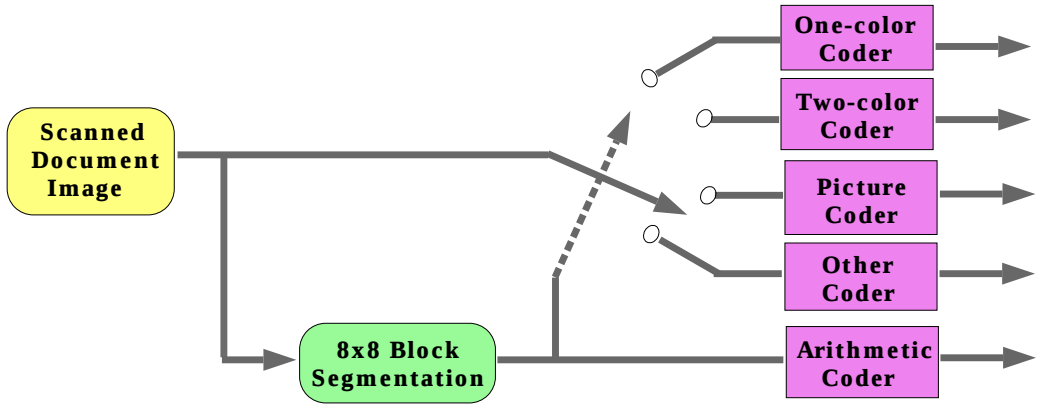
- Color documents scanned at 400 dpi are as big as 45 Megabytes.
- Effective compression of document images is needed for
  - Transmission
  - Storage
- Document images contain regions with distinct characteristics.
  - Text, line graphics: high spatial resolution, low color resolution.
  - Continuous-tone, halftone pictures: low spatial resolution, high color resolution.
- A good document compression should be spatially adaptive.

Previous Approaches



- Block-based approaches (Murata'96, Harrington & Klassen'97, etc.)
  - Segment non-overlapping blocks of pixels into different classes.
  - Compress each class differently according to its characteristics.
- Layer-based approaches (DjVu, de Queiroz, Buckley & Xu'98, etc.)
  - Partition a document into different layers.
  - Each layer is coded as an image independently from other layers.
  - A 3-layer, foreground/mask/background representation proposed in ITU recommendation T.44 for Mixed Raster Content (MRC).

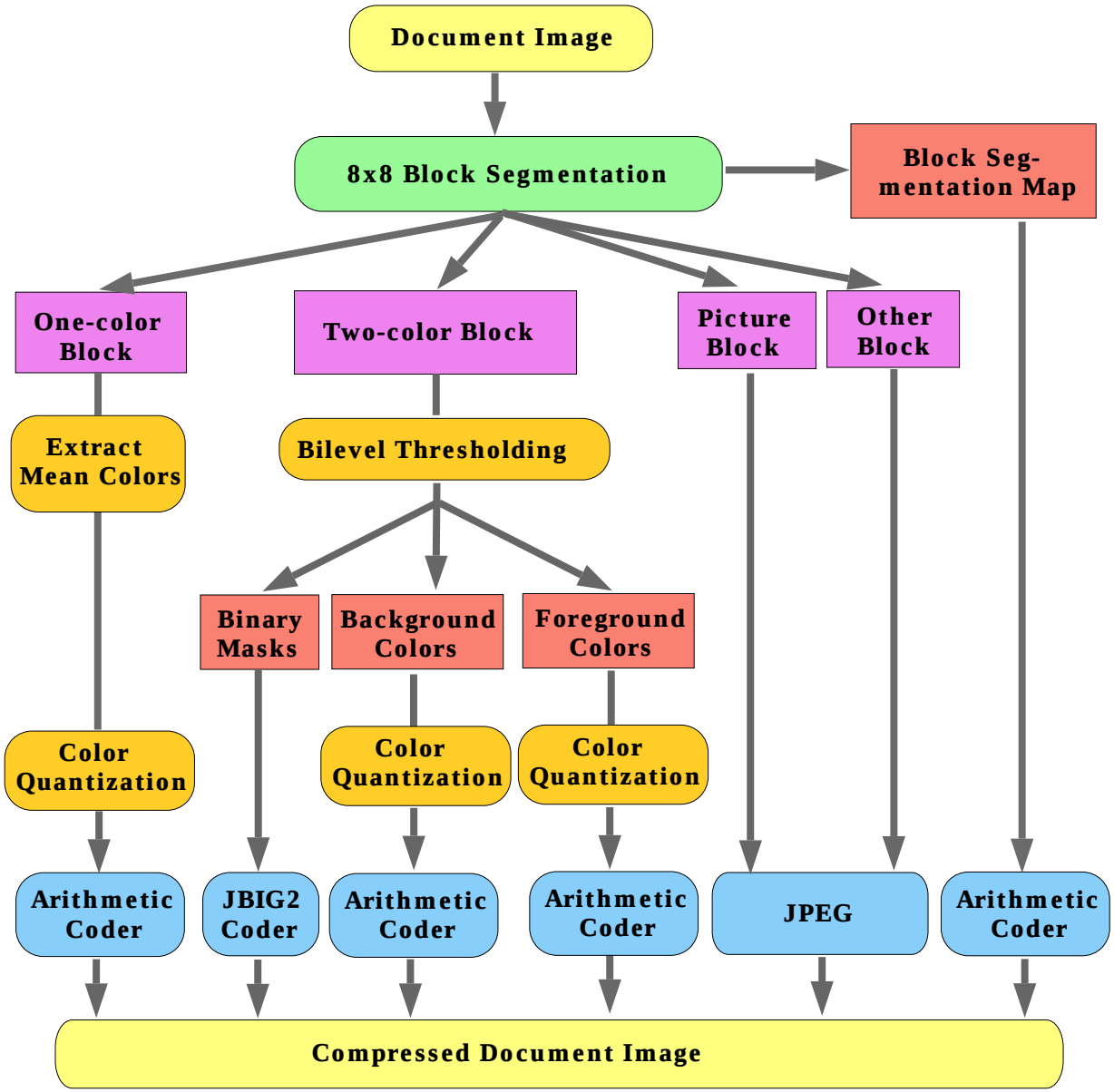
Multilayer Document Compression



- Segments 8×8 blocks into 4 classes.
  - One-color, Two-color, Picture, and Other blocks.
- Compresses each class using a different algorithm.
- Segmentation map is compressed and sent as side information.

Image Classes

- One-color block:
  - Mainly from background regions.
  - Coded as an indexed color.
- Two-color block:
  - Mainly from text, line graphics regions.
  - Coded as two indexed colors and a binary mask.
- Picture block:
  - Mainly from continuous-tone, halftone picture regions.
  - JPEG using customized quantization tables.
- Other block:
  - Blocks with sharp edges & need more than 2 colors to represent.
  - JPEG using standard quantization tables at quality 75.

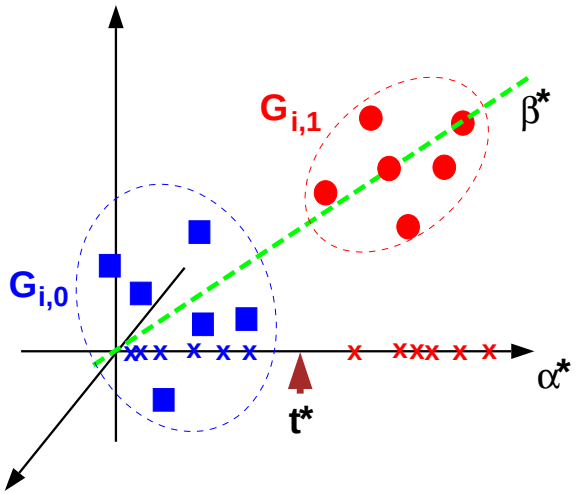

Document Image
8x8 Block Segmentation
Block Seg-mentation Map
One-color Block
Two-color Block
Picture Block
Other Block
Extract Mean Colors
Bilevel Thresholding
Binary Masks
Background Colors
Foreground Colors
Color Quantization
Color Quantization
Color Quantization
Arithmetic Coder
JBIG2 Coder
Arithmetic Coder
Arithmetic Coder
JPEG
Arithmetic Coder
Compressed Document Image

Bilevel Thresholding

- Apply bilevel thresholding to 8×8 Two-color blocks.
  - Extract 2 colors and a binary mask using minimal MSE thresholding.
  - Refine the 2 colors extracted by minimal MSE thesholding.
- For a block, if the number of pixels of one color region is too small, enlarge the 8×8 block to a 16×16 block.
- Apply bilevel thresholding to the 16×16 block.

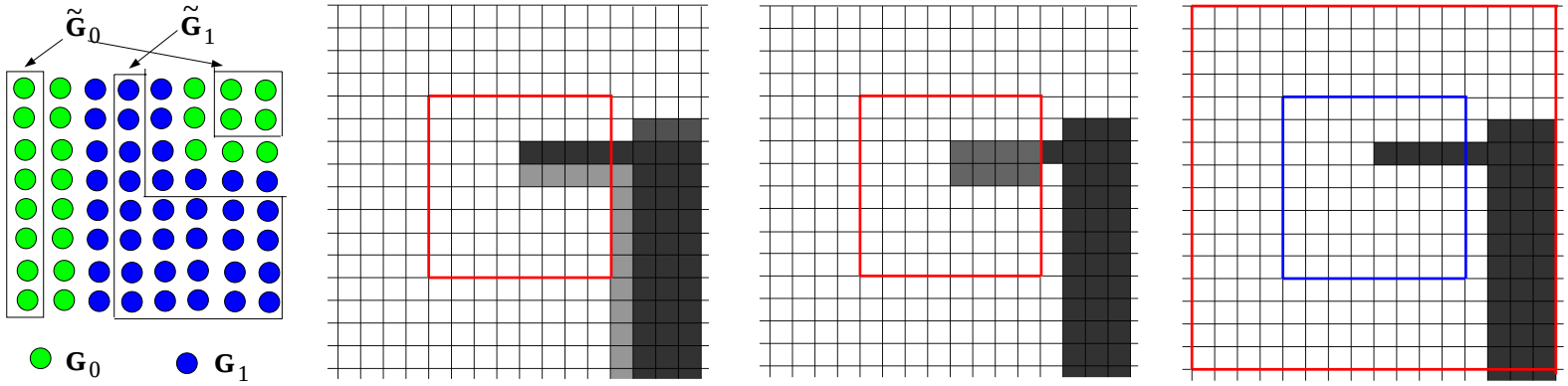
Minimal MSE Thresholding

- Goal: to partition a block into two groups, and minimize MSE.
- When calculating MSE, each pixel is represented by its group mean.
- Minimization is computationally expensive to perform in 3-D.



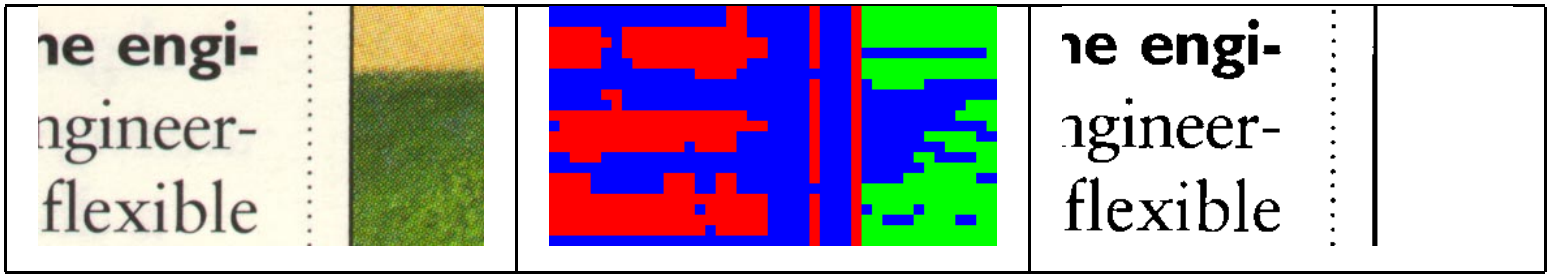
1. Project colors to the color axis with the largest variance, α^*.
2. Find t^*, such that $t^* = \arg\min_t E(t)$, where $E(t)$ is MSE.
3. Let $G_{i,j}$ be group j, and $c_{i,j}$ be the mean color of $G_{i,j}$, $j = 0, 1$.

Refinement



1. Find internal points of $G_{i,j}$ and denote them as $\tilde{G}_{i,j}$.
2. If $|\tilde{G}_{i,j}| > 0$, re-set $c_{i,j}$ to be the mean color of $\tilde{G}_{i,j}$.
3. If $|\tilde{G}_{i,j}| = 0$, enlarge the block to 16×16, then extract 2 colors and the binary mask from 16×16 block.

Compress Binary Masks

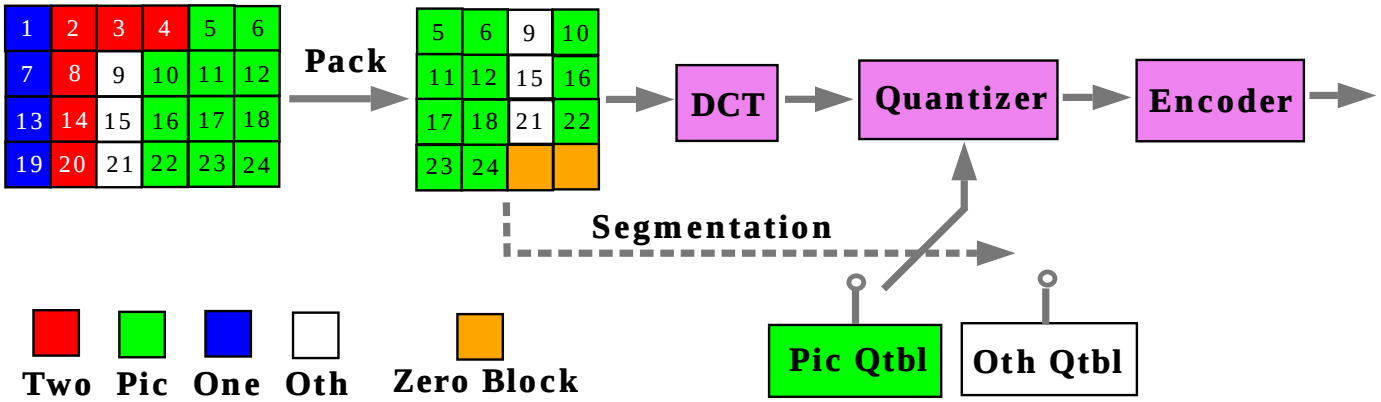


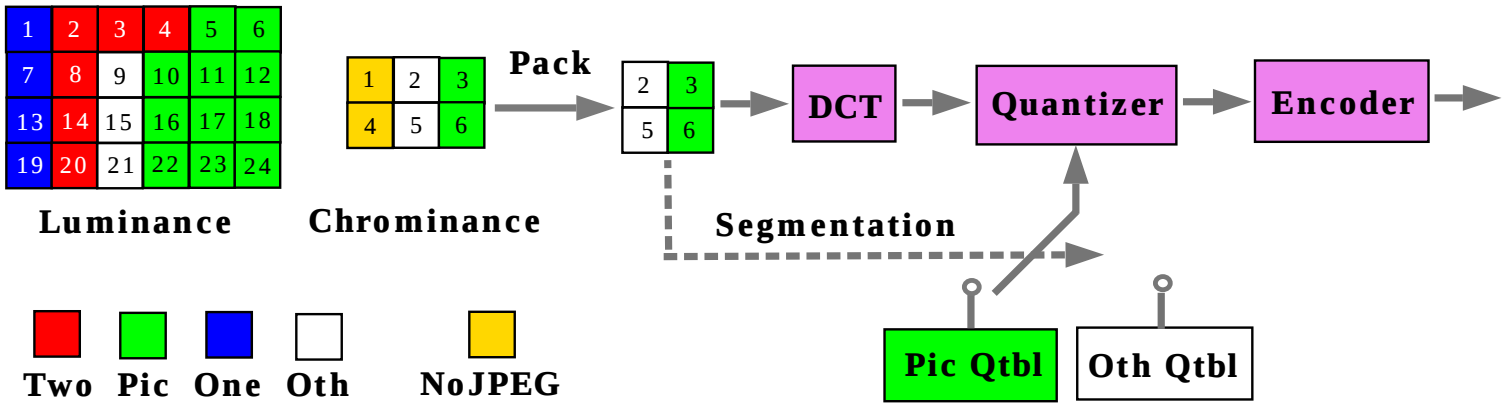
- Form a binary image B which has same size as y.
- Any block in B *not* corresponding to a Two-color block is set to 0's.
- Any block in B corresponding to a Two-color block is set to appropriate binary mask $b_{i,m,n}$.
- B is compressed by a JBIG2 coder using lossless soft pattern matching technique.

Code JPEG Blocks

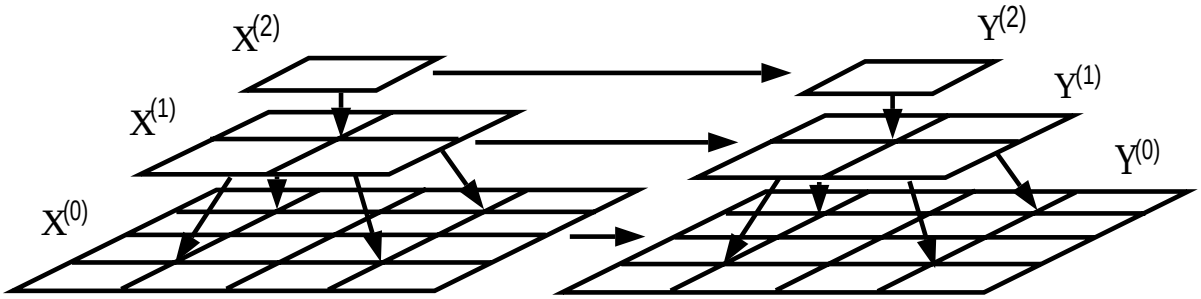
- JPEG blocks include Picture blocks and Other blocks.
- JPEG luminance blocks are packed in raster order, then JPEG'ed.
- JPEG subsamples chrominance 2×2, each 8×8 chrominance block corresponds to four 8×8 blocks in the input image.
- Chrominance segmentation is needed for coding chrominance.
- Chrominance classes: Picture, Other, NoJPEG blocks.
- JPEG chrominance blocks are packed in raster order, then JPEG'ed.

Code JPEG Blocks





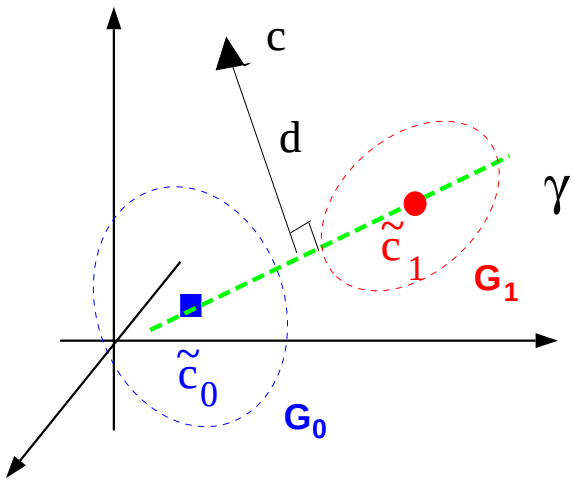
TSMAP Segmentation Algorithm



1. Based on a multiscale Bayesian approach (Bouman & Shapiro'94).
2. Has a novel multiscale context model and a multiscale image model.
3. Trained using typical scanned document images and their accurate segmentations.

Multilayer Compression Using TSMAP

- Segments each block into One-color, Two-color or Picture blocks.
- Other blocks are selected from Two-color blocks as follows:
  - Calculate average distance of boundary points to line determined by c_0 and c_1.
  - If average distance > 45, re-classify current block as Other block.



- For a Two-color block, if total number of internal points ≤ 8, re-classify the block as One-color block.

Chrominance Segmentation of TSMAP

- Chrominance segmentation is computed from 8×8 block segmentation as follows:
  - If any of 4 luminance blocks is *Other*, then set chrominance block to *Other*.
  - Else if any of 4 luminance blocks is *Picture*, then set chrominance block to *Picture*.
  - Else set chrominance block to $NoJPEG$.
- Chrominance segmentation does not need to be sent as side information.

Outline

- Trainable Sequential MAP (TSMAP) segmentation algorithm
- Multilayer document image compression algorithm
- Rate-Distortion Optimized Segmentation (RDOS) algorithm

Segmentation for Compression

- Performance of a document compression system depends on its segmentation algorithm.
  - A good segmentation can lower the bit rate, and the distortion.
  - Most damaging artifacts are often caused by misclassifications.
- Previous segmentation algorithms for document compression.
  - Murata'96 – absolute values of DCT coefficients
  - Konstantinides & Tretter'98 – a DCT activity measure
  - DjVu'98 – multiscale bicolor clustering algorithm
  - Huang etc.'98 – morphological filters followed by thresholding
  - Ramos and de Queiroz'99 – block activity measure

Direct Segmentation for Compression

- Direct approaches – use only the document image data
- Advantages – simple, computationally efficient.
- Disadvantages
  - Do not consider the properties of the coders.
  - Result in infrequent, but serious misclassifications.
  - Segmentation is computed independent of the desired rate-distortion trade-off by the user.

Rate-Distortion Optimized Segmentation

- RDOS method works in a close-loop fashion by
  - Applying each coder to each region
  - Selecting coder for each region to optimize rate-distortion trade-off of entire image
- Let y be original image, x be 8×8 block segmentation. Then,

$$x^* = \arg \min_{x \in \mathcal{N}^L} R(y|x) + R(x) + \lambda D(y|x). \quad (1)$$

- Constant λ controls the trade-off between bit rate and distortion.
- $\mathcal{N} = \{One, Two, Pic, Oth\}$.

Properties of RDOS

- RDOS produces more robust segmentations.
- RDOS allows user to control trade-off between rate and distortion.
- RDOS is different from previous approaches (Ramchandran & Vetterli'94, Effros & Chou'95) in that
  - We switch among different types of coders, instead of parameters of the same coder.
  - We use class-dependent distortion measure to approximate the perceived distortion in text, and picture regions.

Computing RDOS

- For simplicity and computational efficiency, we assume
  - Number of bites for coding a block only depends on image data and class labels of that block and previous block in raster order.
  - Distortion of a block is independent from other blocks.
- Let y_i denote i-th 8×8 block in raster order, x_i denote its class label, and L be the number of 8×8 blocks. Then,

$$x^* = \arg \min_{x \in \mathcal{N}^L} \sum_{i=1}^{L-1} R_i(x_i|x_{i-1}) + R_x(x_i|x_{i-1}) + \lambda D_i(x_i) \qquad (2)$$

- (2) can be solved using dynamic programming techniques.
- Since bit rate for coding segmentation is usually less than 0.01 bpp, we assume that $R(x_i|x_{i-1}) = 0$.

Rate & Distortion of One-color Coder

- If $x_i = One$, y_i is represented by an indexed color denoted as μ_i.
- With 1st order approximation, we have

$$R_i(x_i|x_{i-1}) = \begin{cases} -\log_2 p_\mu(\mu_i|\mu_{i-1}), & if\ x_{i-1} = One \\ -\log_2 p_\mu(\mu_i), & if\ x_{i-1} \neq One \end{cases}$$

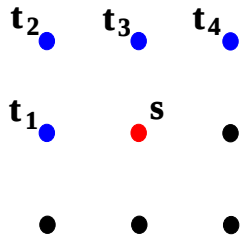
- To estimate $p_\mu(\mu_i|\mu_{i-1})$ and $p_\mu(\mu_i)$, we assume that all blocks are One-color blocks.
- Total squared error in YCrCb is used for One-color blocks.

$$D_i(x_i) = \sum_{m=0}^{7}\sum_{n=0}^{7} \|y_{i,m,n} - \mu_i\|^2.$$

where $y_{i,m,n}$ is the color of pixel (m, n) in y_i, and $\|a\| = \sqrt{a^t a}$.

Rate of Two-color Coder

- A Two-color block i is represented by 2 indexed colors $\tilde{c}_{i,0}$, $\tilde{c}_{i,1}$, and a binary mask $b_{i,m,n}$.



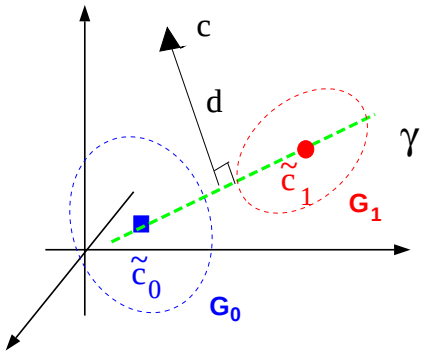
- $R_i(x_i|x_{i-1}) = R_i^I(x_i|x_{i-1}) + R_i^b(x_i|x_{i-1})$
- $R_i^I(x_i|x_{i-1}) = \begin{cases} -\sum_{j=0}^{1} \log_2 p_j(\tilde{c}_{i,j}|\tilde{c}_{i-1,j}) & if\ x_{i-1} = Two \\ -\sum_{j=0}^{1} \log_2 p_j(\tilde{c}_{i,j}) & if\ x_{i-1} \neq Two \end{cases}$
- Assume bits for coding $b_{i,m,n}$ only depend on 4 neighbors, $V_{i,m,n}$.

$$R_i^b(x_i|x_{i-1}) = -\sum_{n=0}^{7}\sum_{m=0}^{7} \log_2 p_b(b_{i,m,n}|V_{i,m,n})$$

- Estimate probabilities from blocks whose maximal dynamic range among 3 color channels ≥ 8.

Distortion of Two-color Coder

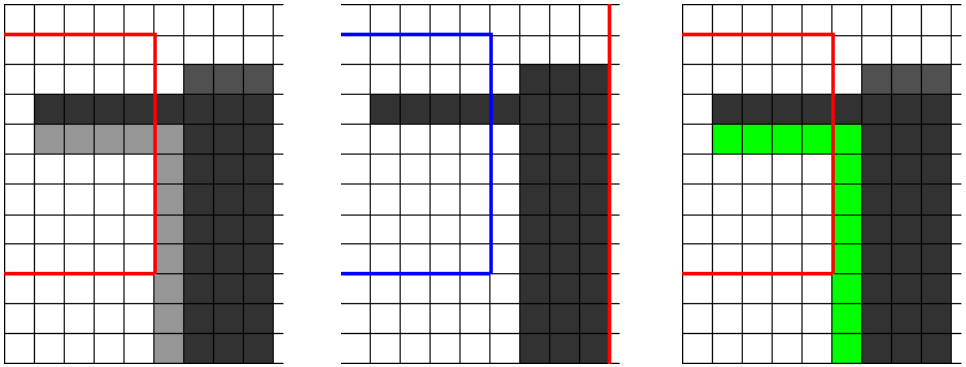
- Sharpening may cause large error in pixel values along boundaries.
- A third color often occurs along boundaries.
- Let $I_{i,m,n} = 1$, if (m, n) is an internal point. $I_{i,m,n} = 0$, otherwise.

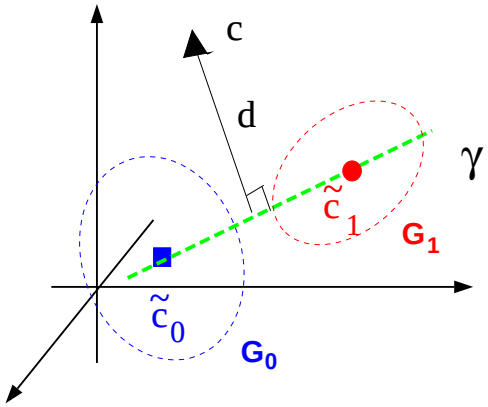


$$D_i(x_i) = \begin{cases} \sum_{m=0}^{7}\sum_{n=0}^{7} \left[I_{i,m,n} \| y_{i,m,n} - \tilde{c}_{i,b_{i,m,n}} \|^2 \right. \\ \qquad \left. +(1 - I_{i,m,n}) d^2(y_{i,m,n}; \tilde{c}_{i,0}, \tilde{c}_{i,1}) \right], & if \ \sum_{j=0}^{1} |\tilde{G}_{i,j}| > 8 \\ 255^2 \times 64 \times 3, & if \ \sum_{j=0}^{1} |\tilde{G}_{i,j}| \leq 8 \end{cases}$$

where $d(c; \tilde{c}_0, \tilde{c}_1)$ is distance between c and line determined by $\tilde{c}_0$ & $\tilde{c}_1$.

Distortion of Two-color Coder





- Let $I_{i,m,n} = 1$, if (m, n) is an internal point. $I_{i,m,n} = 0$, otherwise.

$$D_i(x_i) = \begin{cases} \sum_{m=0}^{7}\sum_{n=0}^{7} \Big[I_{i,m,n} \| y_{i,m,n} - \tilde{c}_{i,b_{i,m,n}} \|^2 \\ \qquad +(1 - I_{i,m,n}) d^2(y_{i,m,n}; \tilde{c}_{i,0}, \tilde{c}_{i,1}) \Big], & if \ \sum_{j=0}^{1} |\tilde{G}_{i,j}| > 8 \\ 255^2 \times 64 \times 3, & if \ \sum_{j=0}^{1} |\tilde{G}_{i,j}| \leq 8 \end{cases}$$

where $d(c; \tilde{c}_0, \tilde{c}_1)$ is distance between c and line determined by $\tilde{c}_0$ & $\tilde{c}_1$.

Rate of JPEG Coder

- $R_i(x_i|x_{i-1}) = R_i^l(x_i|x_{i-1}) + R_i^c(x_i|x_{i-1})$
- $\alpha_i^d(x_i)$ is quantized DC, $\alpha_i^a(x_i)$ is quantized AC of luminance.

$$R_i^l(x_i|x_{i-1}) = r_d \left[\alpha_i^d(x_i) - \alpha_{i-1}^d(x_{i-1})\right] + r_a \left[\alpha_i^a(x_i)\right].$$

- $\beta_{j,k}^d(z_j)$ is quantized DC, $\beta_{j,k}^a(z_j)$ is quantized AC of k-th chrominance component.

$$R_i^c(x_i|x_{i-1}) = \frac{1}{4}\sum_{k=0}^{1} \left\{ r'_d \left[\beta_{j,k}^d(x_i) - \beta_{j-1,k}^d(x_{i-1})\right] + r'_a \left[\beta_{j,k}^a(x_i)\right] \right\}.$$

- Note: we split number of bits for coding chrominance equally among 4 corresponding 8×8 blocks.
- We assume $\alpha_{i-1}^d(x_{i-1}) = \beta_{j-1,k}^d(x_{i-1}) = 0$, if $x_{i-1} \notin \{Pict, Oth\}$.

Distortion of JPEG Coder

- Total squared error in YCrCb is used as JPEG distortion.
- Let $e_i^l(x_i)$, $e_{j,k}^c(z_j)$ be quantization error of DCT coefficients of luminance and chrominance, respectively.

$$D_i(x_i) = \left\| e_i^l(x_i) \right\|^2 + \sum_{k=0}^{1} \left\| e_{j,k}^c(x_i) \right\|^2$$

- $D_i(x_i)$ is calculated in DCT domain. No IDCT is needed.
- We approximate distortion due to chrominance channels by dividing chrominance error among 4 corresponding 8×8 blocks.

JPEG Chrominance Segmentation

- Let chrominance segmentation be $z = \{z_0, z_1, \ldots, z_{L/4-1}\}$.
- Compute RDOS for chrominance with constrain, $z_j \in \{Pic, Oth\}$.

$$z = \arg \min_{z' \in \{Pic,Oth\}^{L/4}} \sum_{j=0}^{L/4-1} \left\{ \tilde{R}_j(z'_j | z'_{j-1}) + \lambda \tilde{D}_j(z'_j) \right\}$$

$$\tilde{R}_j(z_j | z_{j-1}) = \sum_{k=0}^{1} \left\{ r'_d \left[\beta^d_{j,k}(z_j) - \beta^d_{j-1,k}(z_{j-1}) \right] + r'_a \left[\beta^a_{j,k}(z_j) \right] \right\}.$$

$$\tilde{D}_j(z_j) = \sum_{k=0}^{1} \left\| e^c_{j,k}(z_j) \right\|^2$$

- Then, z_j is set to NoJ, if none of 4 corresponding 8×8 blocks is JPEG block (Picture or Other).

Conclusion

- A spatially adaptive compression algorithm is developed for document images.
- We also proposed a way to compute a rate-distortion optimized segmentation for our compression algorithm.
- At similar bit rates, our algorithm can achieve a higher subjective quality than DjVu, SPIHT and JPEG.