Multiscale Document Compression *

Hui Cheng

Digital Imaging Technology Ctr.
Xerox Research and Technology
Webster, NY 14580
Hui.Cheng@crt.xerox.com

Charles A. Bouman

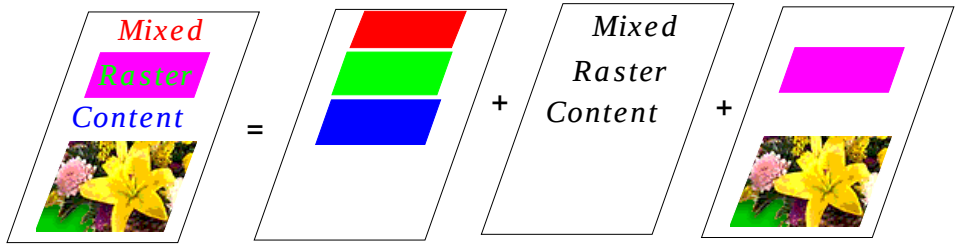
School of Electrical & Comp. Eng.
Purdue University
West Lafayette, Indiana 47907
bouman@enc.purdue.edu

*This research was supported by Xerox IMPACT Imaging and Xerox Foundation.

Document Image Compression

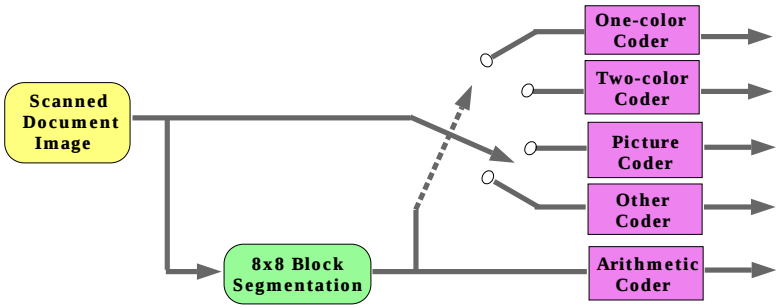
- High quality documents are scanned at 400-600 dpi and 24 bpp.
- Uncompressed scanned images can be as big as 45-90 Megabytes.
- Effective compression of document images is needed for
  - Transmission and storage.
- Document compression is targeted at
  - Visually lossless when printed at original resolution.
  - Very low bit rate ($\leq$ 0.30 bpp, 80:1 compression).
- 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.

Approaches to Document Compression



- 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).

Our View of Document Compression



- 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.

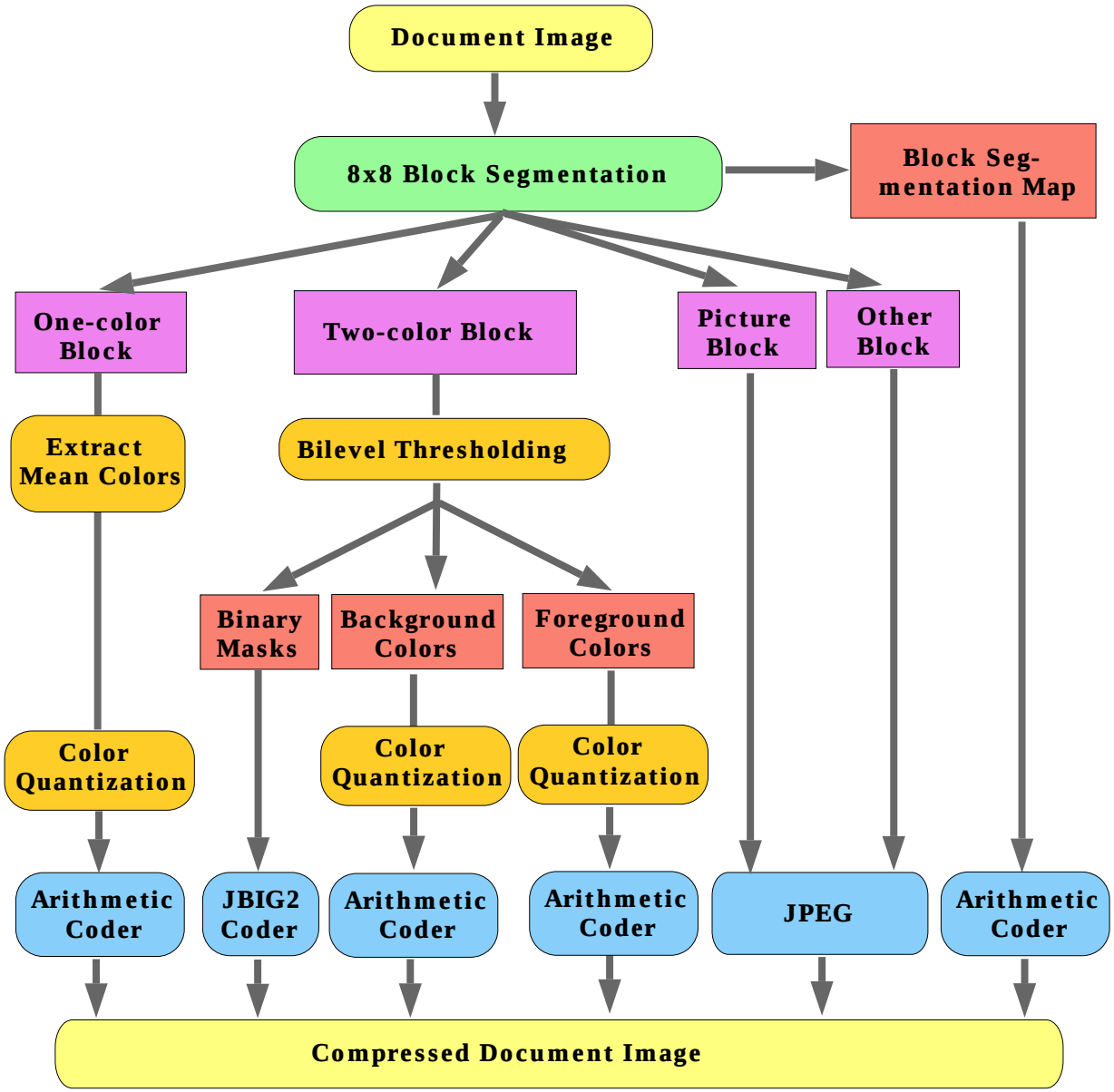
How to compute segmentation?

- Direct segmentation based on semantic categories
  - Murata'96 – absolute values of DCT coefficients
  - Konstantinides & Tretter'98 – a DCT activity measure
  - DjVu'98 – multiscale bicolor clustering algorithm
  - Ramos and de Queiroz'99 – block activity measure
- Segment to maximize coder performance
  - Adapts to desired bit rate
  - Selects segmentation based on result rather that semantic class
  - Minimizes serious defects
  - Minimum rate-distortion segmentation
  - Need to make some simplifying assumptions

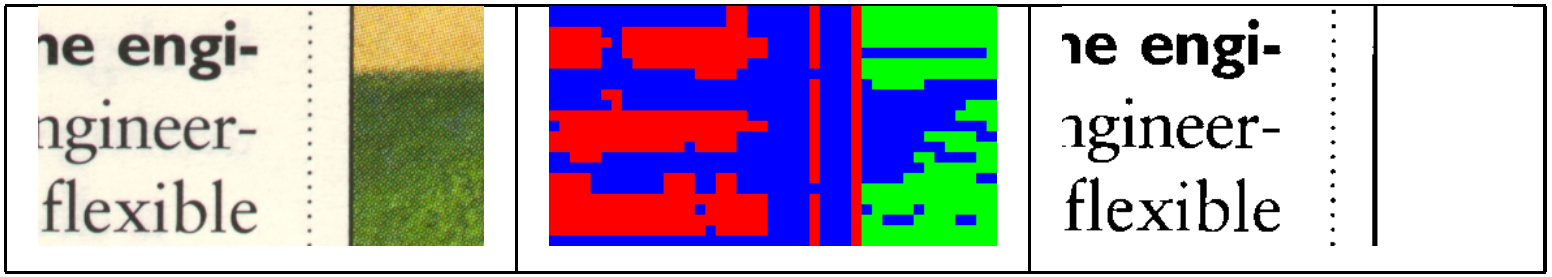
- Rate distortion **optimized** segmentation (RDOS)
  - Approximately optimizes rate-distortion performance
  - Draws on related work of:
    * Ramchandran and Vetterli - JPEG/MPEG bit allocation
    * Effros and Chou - DCT/VQ/KLT selection and design
    * Schuster and Katsaggelos - Dynamic programming for MPEG

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

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.

Computing RDOS

- For tractability, we assume
  - Rate depends on image data of block and class labels of current and previous block.
  - Distortion of a block is independent from other blocks.

  y_i - i-th 8×8 block in raster order

  x_i - class of y_i

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

- Solve using dynamic programming
- Assume $R(x_i|x_{i-1}) = 0$.
- Formulate espressions for rate and distortion of each class

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

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

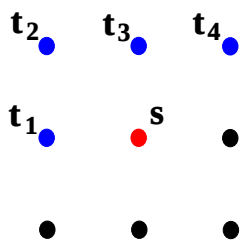
- To estimate $p_\mu(\mu_i|\mu_{i-1})$ and $p_\mu(\mu_i)$, we assume that all blocks are One-color blocks.
- Total squared error in YCrCb is used for One-color blocks.

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

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 $b_{i,m,n}$.



- $R_i(x_i|x_{i-1}) = R_i^I(x_i|x_{i-1}) + R_i^b(x_i|x_{i-1})$

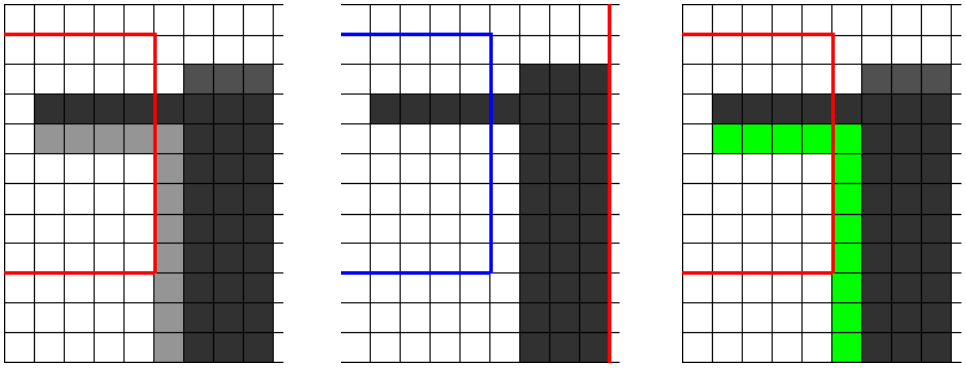
- $$R_i^I(x_i|x_{i-1}) = \begin{cases} -\sum_{j=0}^{1} \log_2 p_j(\tilde{c}_{i,j}|\tilde{c}_{i-1,j}) & if\ x_{i-1} = Two \\ -\sum_{j=0}^{1} \log_2 p_j(\tilde{c}_{i,j}) & if\ x_{i-1} \neq Two \end{cases}$$

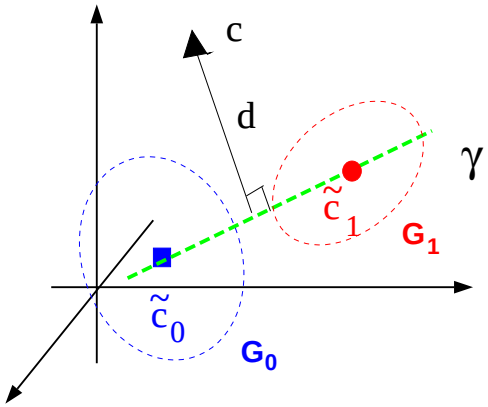
- Assume bits for coding $b_{i,m,n}$ only depend on 4 neighbors, $V_{i,m,n}$.

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

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

Distortion of Two-color Coder





- 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.

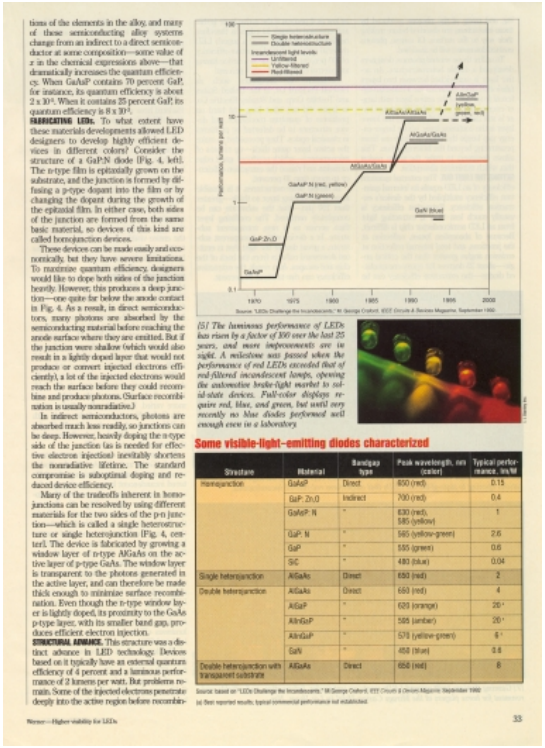
Experiments

- Image data base
  - 30 document images scanned at 400 dpi and 24 bpp.
  - 1 synthetic document image.
- Experimental results are compared with
  - DjVu: designed for web browsing, can achived very high compression ratio.
  - SPIHT: wavelet coder, good for natural images.
  - JPEG.
- Segmentation results are compared with those computed by TSMAP segmentation algorithm (Cheng & Bouman'98).

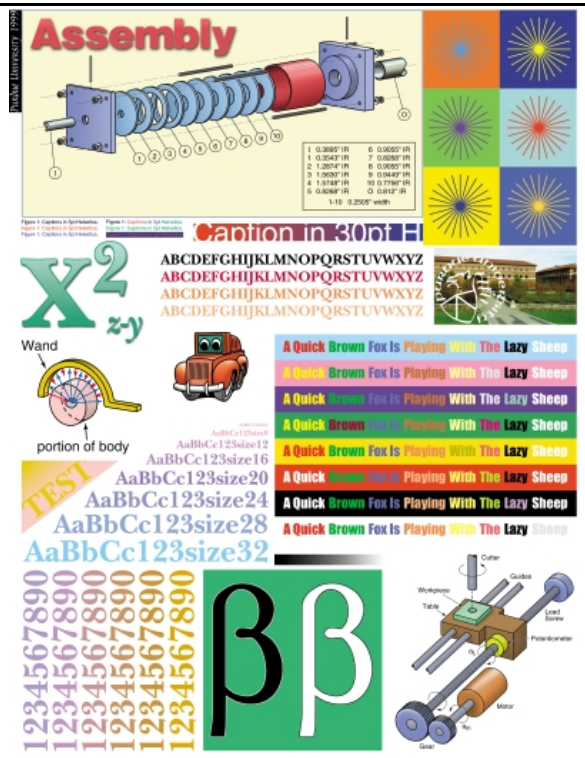
Test Images



Test Image I



Test Image II



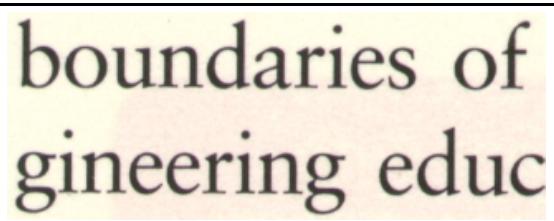
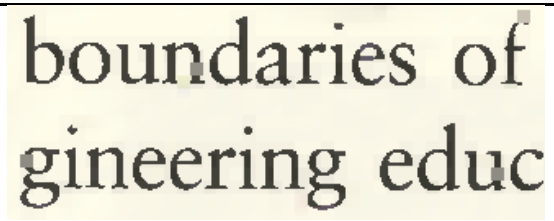
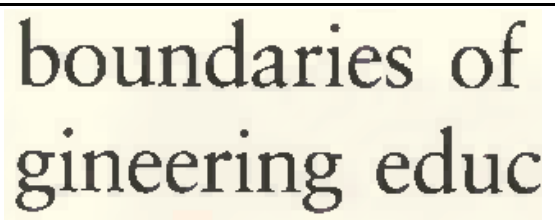
Test Image III

Bit Rates, CR and RDOS Distortion

image	coder	bit rate (bpp)	CR	distortion*	λ
Test image I	TSMAP	0.138	173:1	27.58	n/a
	RDOS	0.132	182:1	23.47	0.0021
	RDOS	0.125	192:1	24.99	0.0018
	RDOS	0.095	253:1	31.00	0.0013
Test image II	TSMAP	0.120	200:1	40.33	n/a
	RDOS	0.114	210:1	32.14	0.0018
Test image III (Synthetic)	TSMAP	0.089	245:1	32.12	n/a
	RDOS	0.101	237:1	3.40	0.0042

* RDOS distortion per pixel per color channel

Images Compressed Using TSMAP and RDOS

		
Test Image I	TSMAP 0.138 bpp (173:1)	RDOS with $\lambda = 0.0021$ 0.132 bpp (182:1)

Average Rates of Coding Each Class

classes	average bit rate (bpp)	standard deviation
One-color	0.0240	0.0092
Two-color	0.3442	0.1471
JPEG	0.8517	0.3260
Segmentations	0.0097	0.0002

- Mean and standard deviation are calculated from 30 scanned document images compressed using RDOS with $\lambda = 0.0018$.
- Segmentations include both 8×8 block segmentation and 16×16 chrominance segmentation.

RDOS segmentations with different λ's

Two-color Block | Picture Block | One-color Block | Other Block



Test Image I



RDOS with $\lambda = 0.0013$

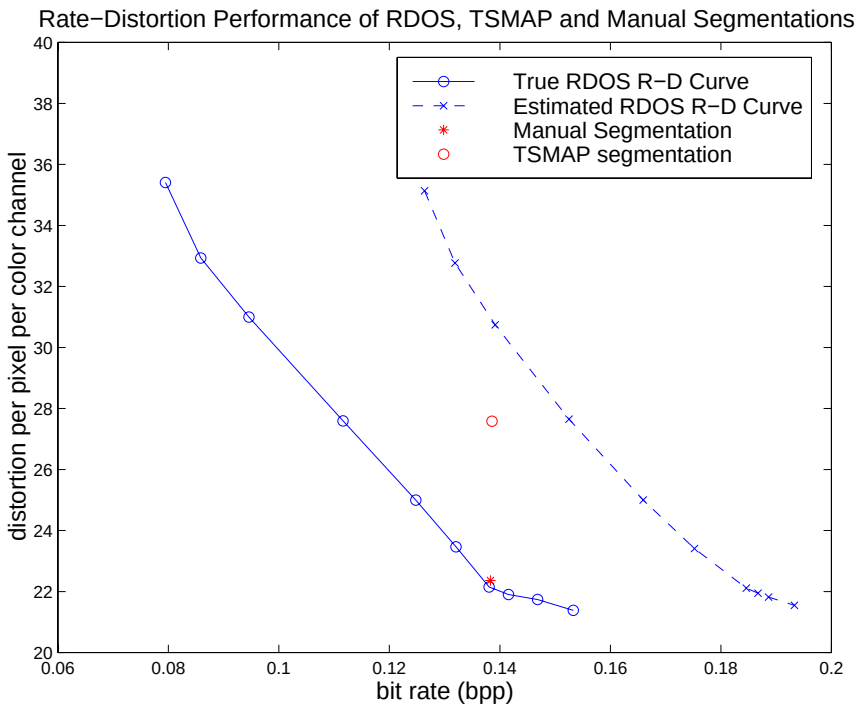
0.095 bpp (253:1)



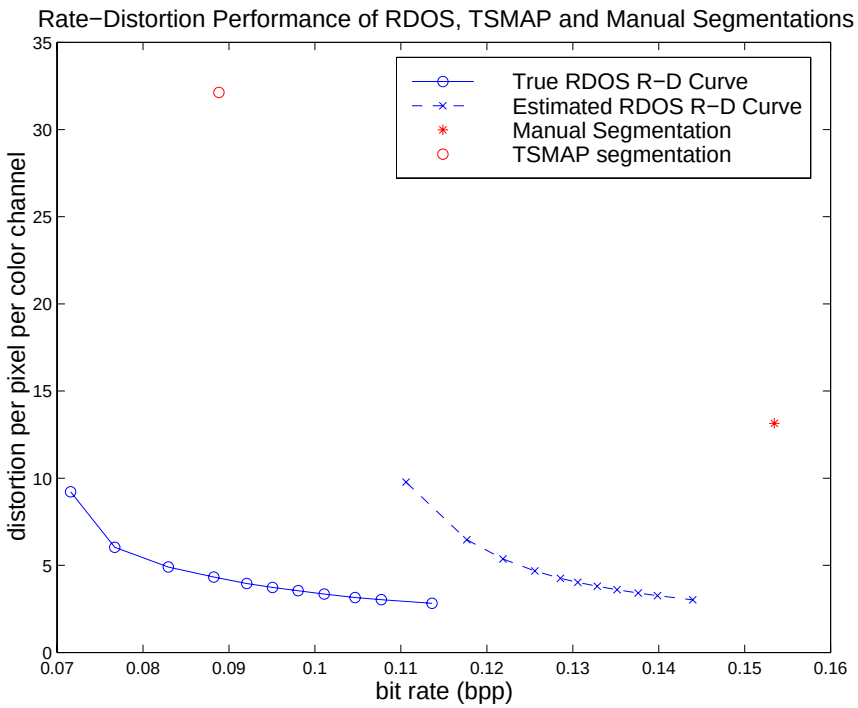
RDOS with $\lambda = 0.0018$

0.125 bpp (192:1)

Rate-distortion Performance

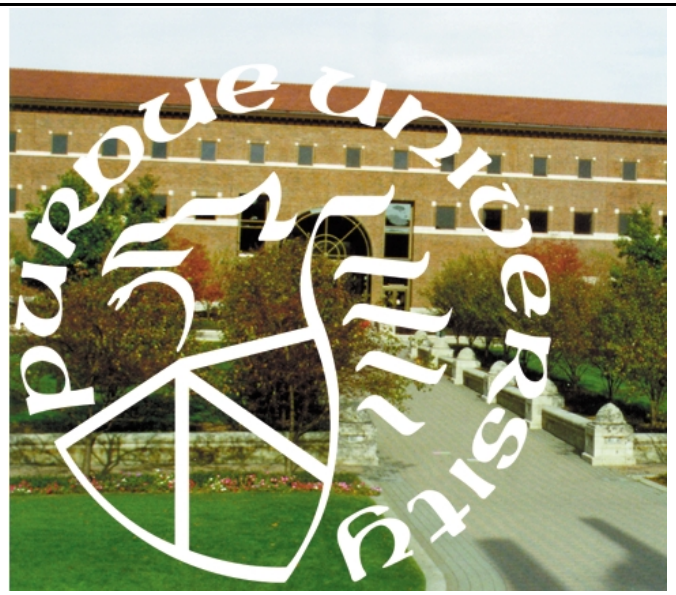
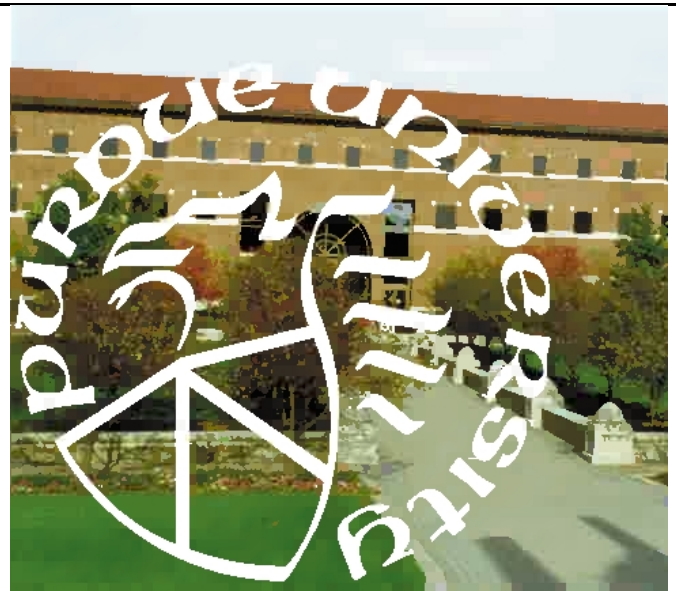
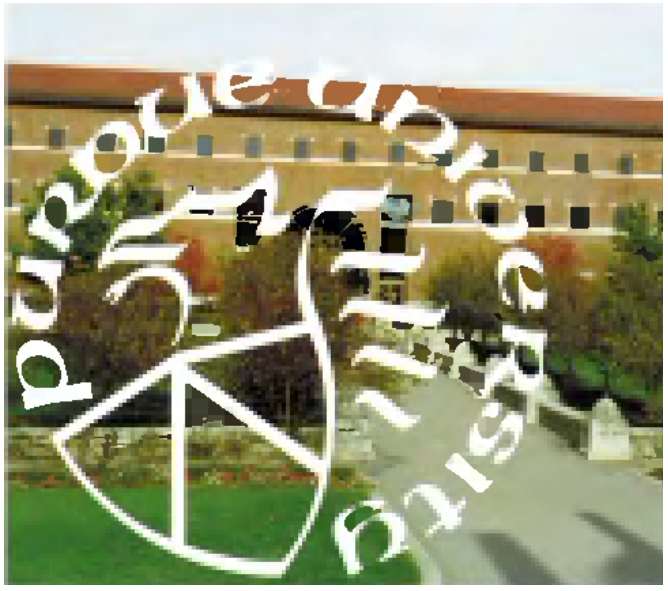


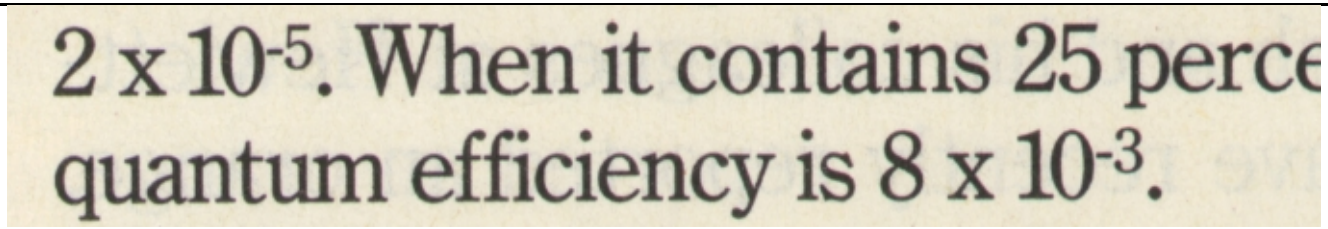
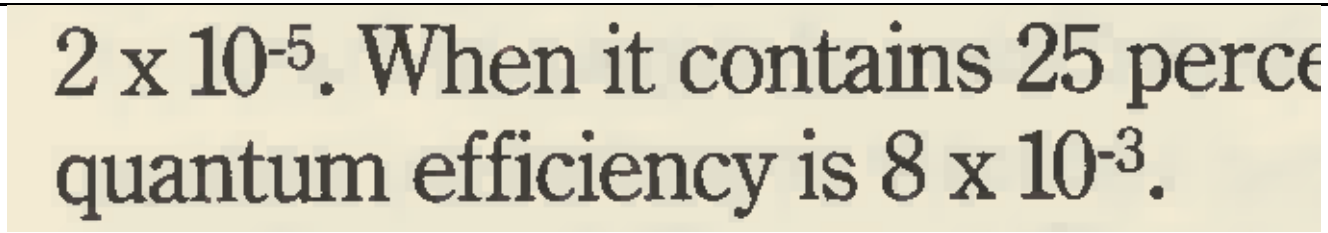
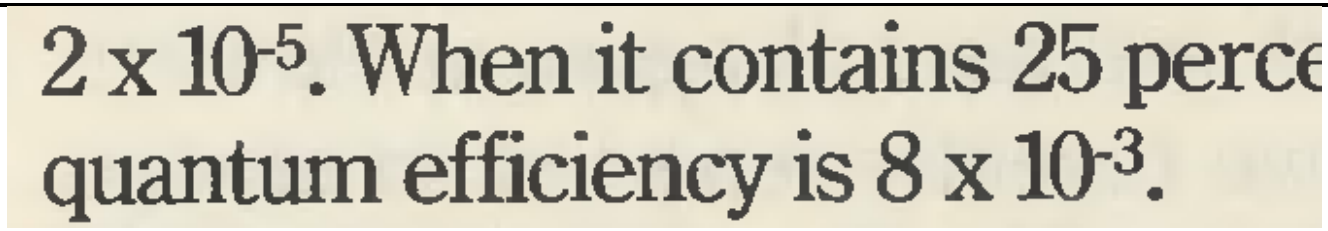
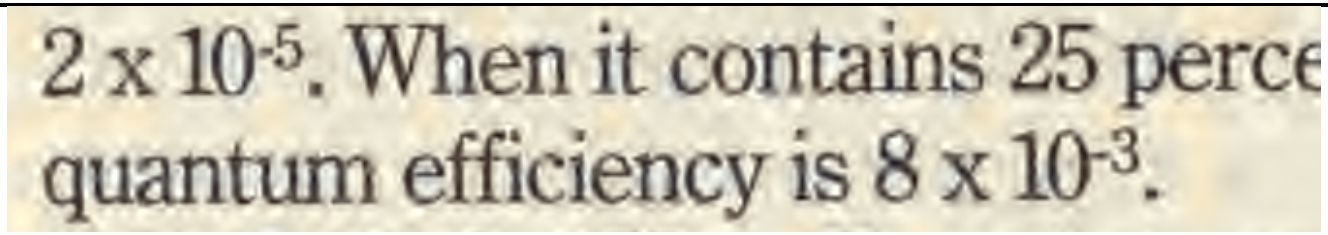
(a) Test Image I

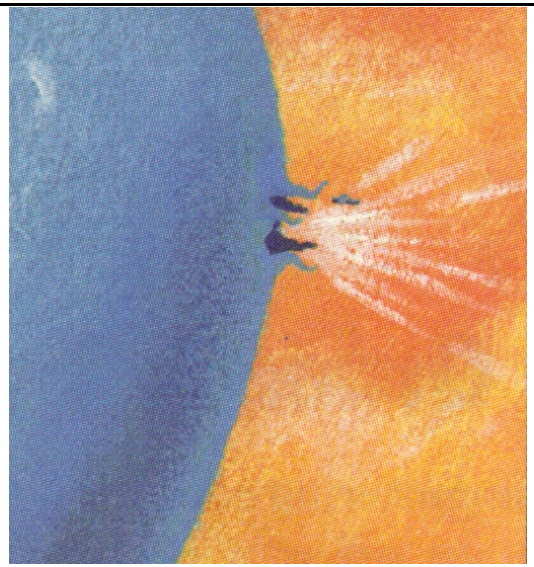
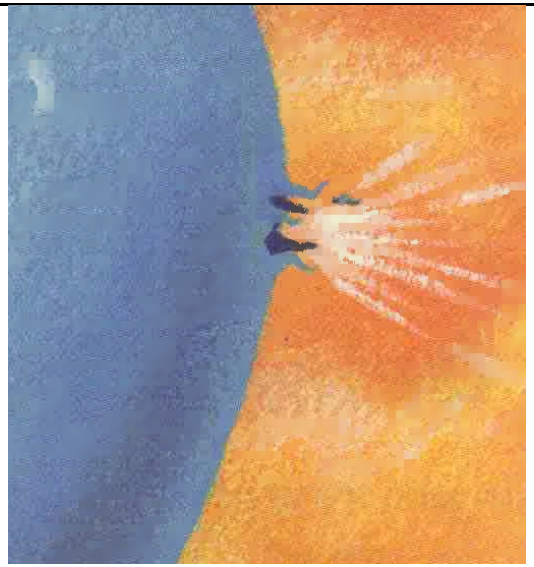
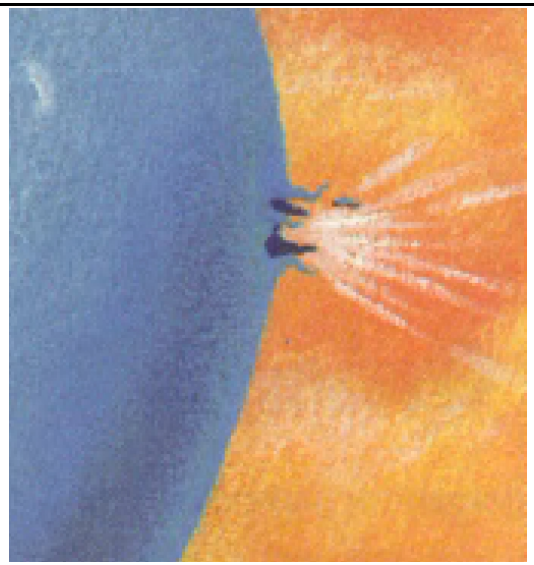


(b) Test Image III

		
Original	RDOS, 0.101 bpp	DjVu, 0.103 bpp
		
SPIHT, 0.103 bpp	JPEG, 0.184 bpp	

Org			RDOS
DjVu			SPIHT

Original	
RDOS 0.114 bpp	
DjVu 0.114 bpp	
SPIHT 0.114 bpp	

Original			RDOS 0.125 bpp
DjVu 0.132 bpp			SPIHT 0.125 bpp

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.

Acknowledgement

- Xerox IMPACT Imaging and Xerox Foundation.
- Dr. Faouzi Kossentini and Mr. Dave Tompkins, University of British Columbia.
- ASEE, ASEE Prism
- IEEE, IEEE Spectrum
- Stanley Electric Sales of America