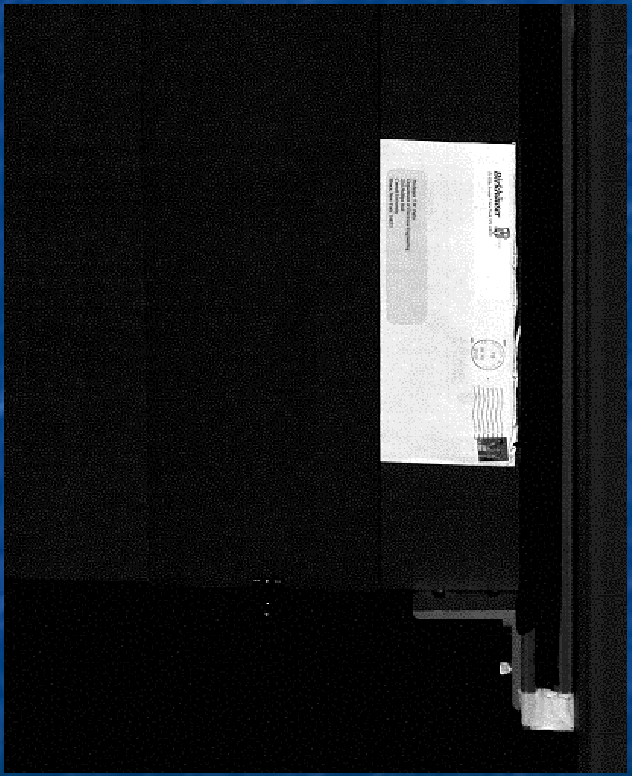
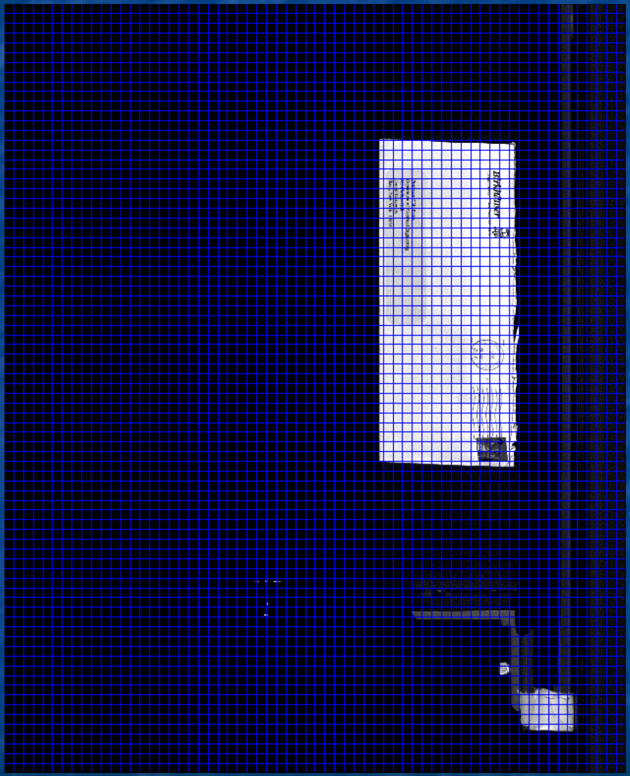
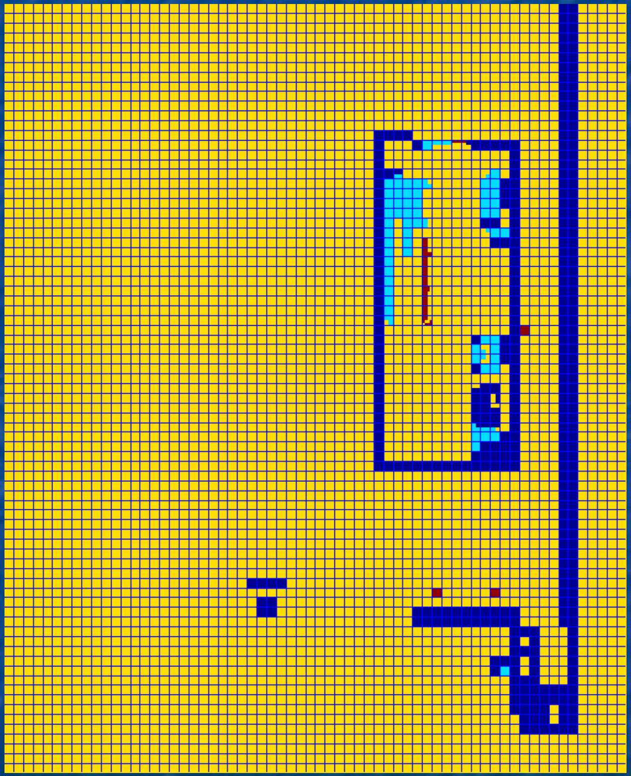
Block-segmentation and Classification of Grayscale Postal Images

Kush R. Varshney
School of Electrical and Computer Engineering
Cornell University, Ithaca, NY

Problem

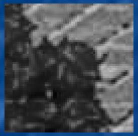




Wavelet Approach

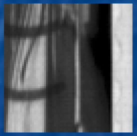
- J. Li and R. M. Gray, “Context-Based Multiscale Classification of Document Images Using Wavelet Coefficient Distributions.”
- First pass classification based on Wavelet Coefficient Distributions
- Multiscale approach
  - Context-based microclassification
- Implemented by Lau, Chen, Ong, Koo

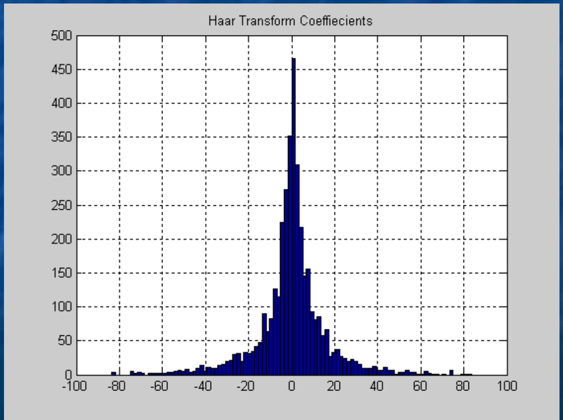
Distribution in Pictures

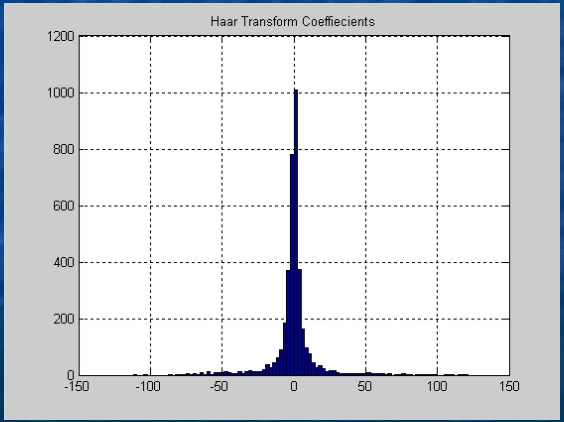
Computer-Generated



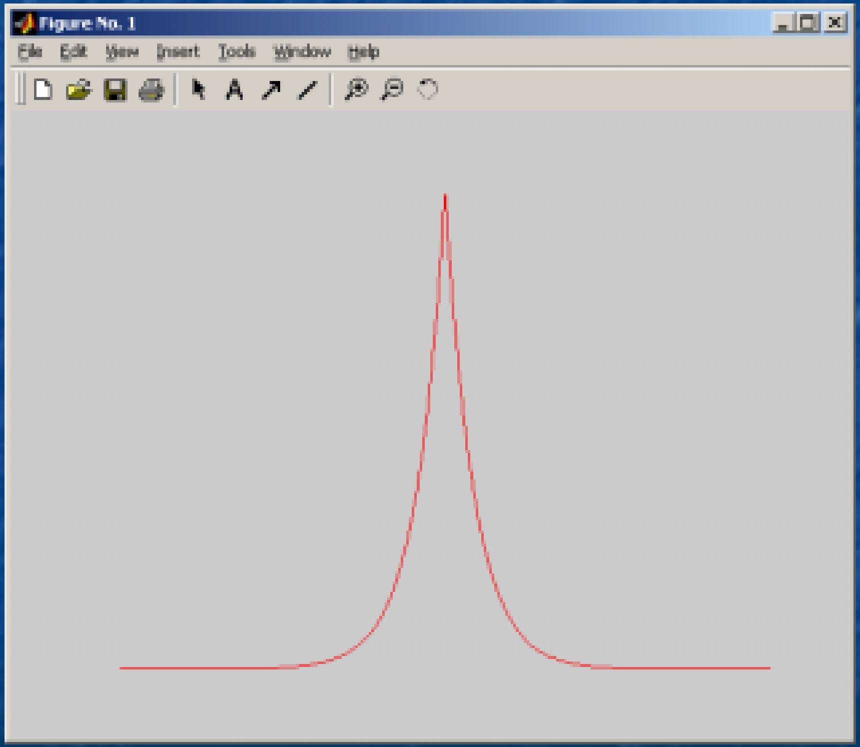
Postal

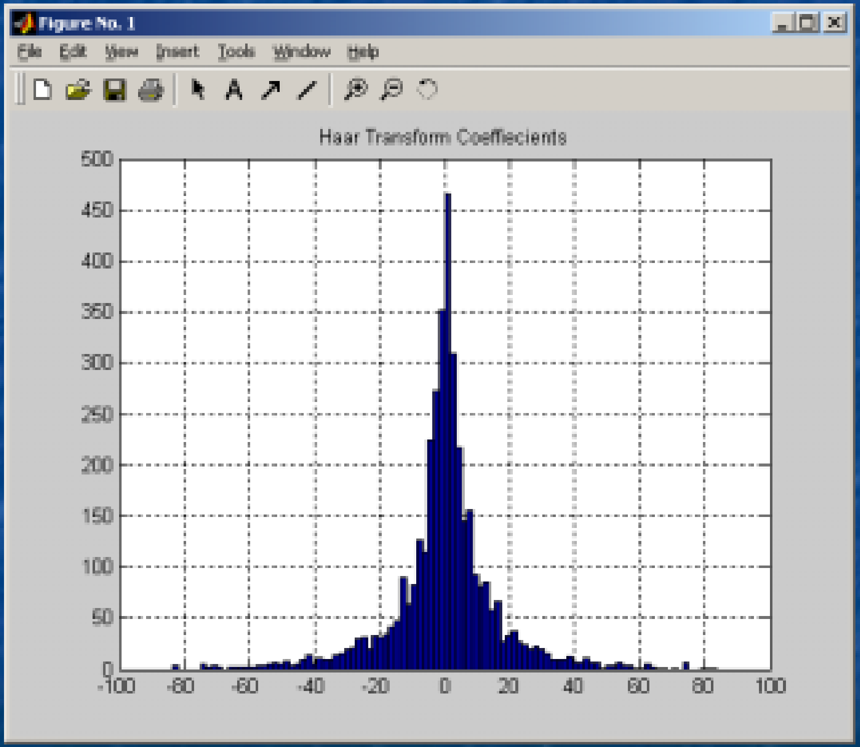






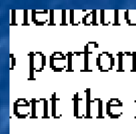
Laplacian Distribution

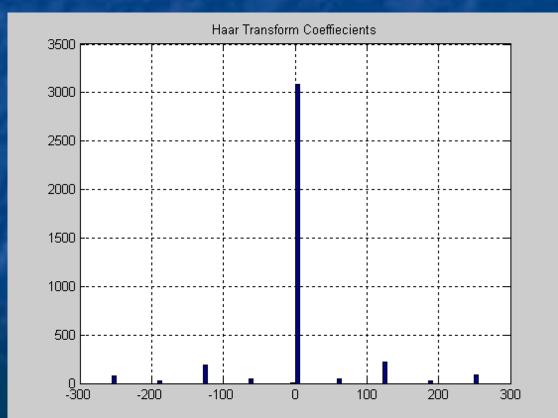




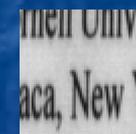
Distribution in Text

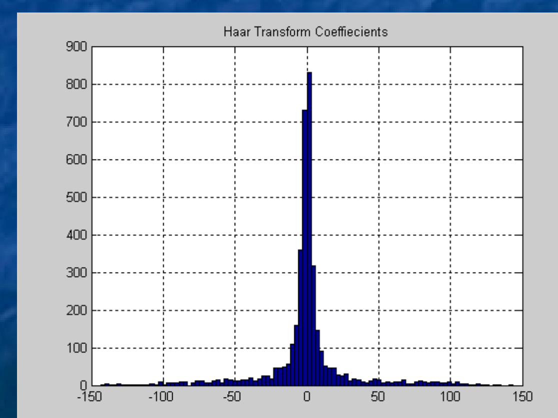
Computer-Generated





Postal





Haar Transform Coefficient Distributions Summary

- Pictures (computer-generated and postal), have Laplacian Distribution of Haar Transform Coefficients
- Computer-generated text has extremely discrete distribution clustered around certain values
- Postal text has almost Laplacian distribution
  - Caused by 'Dirty Text' problem
    - Transitions from characters to background are not abrupt
    - Many different transitions produce smooth coefficient distribution, not clustered distribution

Context-based Classification

- Classifier uses information gathered at large block sizes to aid classification of ‘undetermined’ smaller blocks
- Procedure delineated by Li and Gray; implemented by Lau, et al.
  - Interpolates from ‘determined’ adjacent blocks

Evaluation of Context-based Classification

- When algorithm is applied to computer-generated data, it works extremely well
  - Shows effectiveness of context-based classification
- When applied to postal data, first pass is not effective enough to enable context-based classification to work

Discussion

- Classification approach was designed for computer-generated document images
  - Relies on sharp edges among text characters
  - Fails due to ‘dirty text’
- Multiscale context-based classification may be useful if first pass classification is suited to postal data
- Therefore, I propose a new first pass classification scheme

New First Pass Classification: 6-D Approach

- Use a 6-dimensional feature space for first pass classificaton
- 5 features proposed by A. Suvichakorn, S. Watcharabusaracum, and W. Sinthupinyo, in “Simple Layout Segmentation of Gray-Scale Document Images.”
- 1 feature not in the literature

6 Classification Features

- μ – Mean Intensity
- σ – Standard Deviation Intensity
- α – Active Pixels
- D_x – Sum of Second Derivatives in x
  - via Savitzky-Golay Filter
- D_y – Sum of Second Derivatives in y
  - via Savitzky-Golay Filter
- g_c - Sum of Cardinal Gradient Vector magnitudes

μ – Mean Intensity

- Mean value of all pixel intensities in block
- For envelopes, three ranges apparent for three classes:
  - μ ~ 240 background
  - μ ~ 90 picture
  - μ ~ 180 text
- Does not work well for magazines

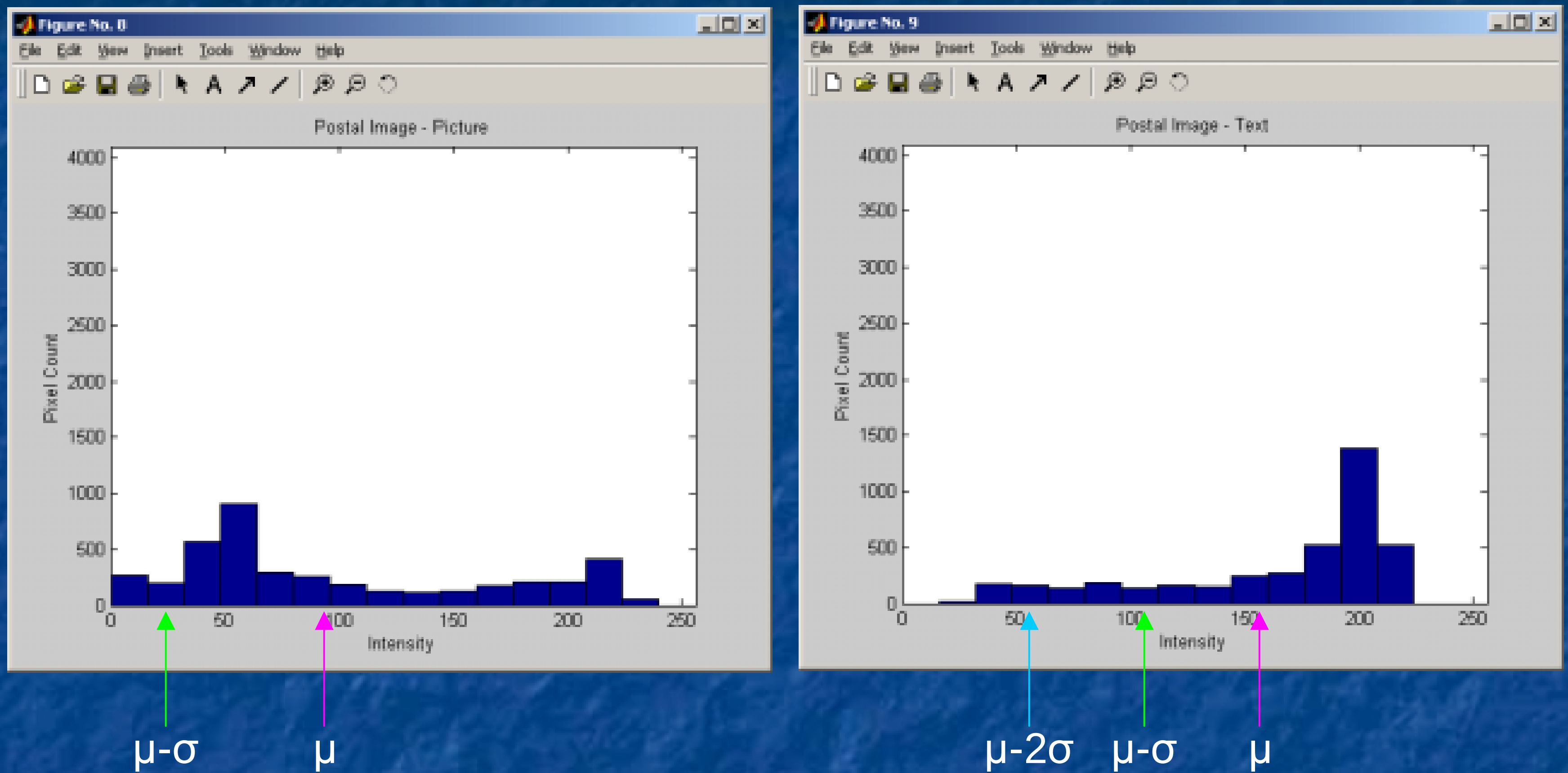
σ – Standard Deviation Intensity

- Standard Deviation of all pixel intensities in block
- Used to distinguish background
  - $\sigma \sim 0$ background

α – Active Pixels

- $\alpha = \Sigma (I < \mu - k \cdot \sigma)$
- A count is taken of pixels with intensity less than the threshold $\mu - k \cdot \sigma$
  - k is a chosen constant
- The method of Adaptive Thresholding is employed
  - Each block has its own threshold
  - Avoids problems associated with predetermined global threshold
    - Necessity to characterize data set *a priori*
    - Irregular lighting of document

Why it Works



- Text blocks have preponderance of light pixels (background), which brings up the mean. Dark pixels (characters) are also present. Therefore, pixels more than 2 standard deviations below the mean are abundant.

Evaluation of α

- Pictures and text can be distinguished using α
- Results are not heavily influenced by choice of k

D_x and D_y – Sum of Second Derivatives of Average Intensity

- $I_{av,x}$ denotes the vector containing the means of the image block columns
- $I_{av,y}$ denotes the vector containing the means of the image block rows
- $D_x = \Sigma(|(d^2 I_{av,x}/dx^2)|$, evaluated at all points)
- $D_y = \Sigma(|(d^2 I_{av,y}/dy^2)|$, evaluated at all points)
- Second Derivatives are obtained with Savitzky-Golay Filter

Why it Works

- ↑ Edges, ↑ |Second Derivatives|, ↑ D
- Other edge-based features have been proposed
  - Sobel Gradient Sum
  - Prewitt Operator
  - Canny Edge detector
- Savitzky-Golay second derivative is computationally fast

Review of Savitzky-Golay Filter

- Also known as Least-Squares Smoothing Filter
- A frame of points surrounding the current point is fit to a polynomial of specified order by the technique of least-squares
- From the fitted polynomial, the function value of the current point is retained, while the function values of other points in the frame are discarded

Review of Savitzky-Golay Filter
(ex: Polynomial Order=2; Frame Length=5)

- $\mathbf{w}$ is noisy data
- Basis Vectors
  - $\mathbf{s}_0 = [1;1;1;1;1]$
  - $\mathbf{s}_1 = [-2;-1;0;1;2]$
  - $\mathbf{s}_2 = [4;1;0;1;4]$
  - $\mathbf{S} = [\mathbf{s}_0\ \mathbf{s}_1\ \mathbf{s}_2]$
- $\hat{\mathbf{w}} = c_0\mathbf{s}_0 + c_1\mathbf{s}_1 + c_2\mathbf{s}_2$
- Find optimal $\mathbf{c}$
- $\mathbf{c} = (S^TS)^{-1}S^T\mathbf{w}$

Review of Savitzky-Golay Filter
(ex: Polynomial Order=2; Frame Length=5)

n	w
1	4
2	0
3	0
4	1
5	6
6	10
7	17
8	23

current point

Frame

- $(S^T S)^{-1} S^T [0;1;6;10;17] = [5.229;4.300;0.786]$

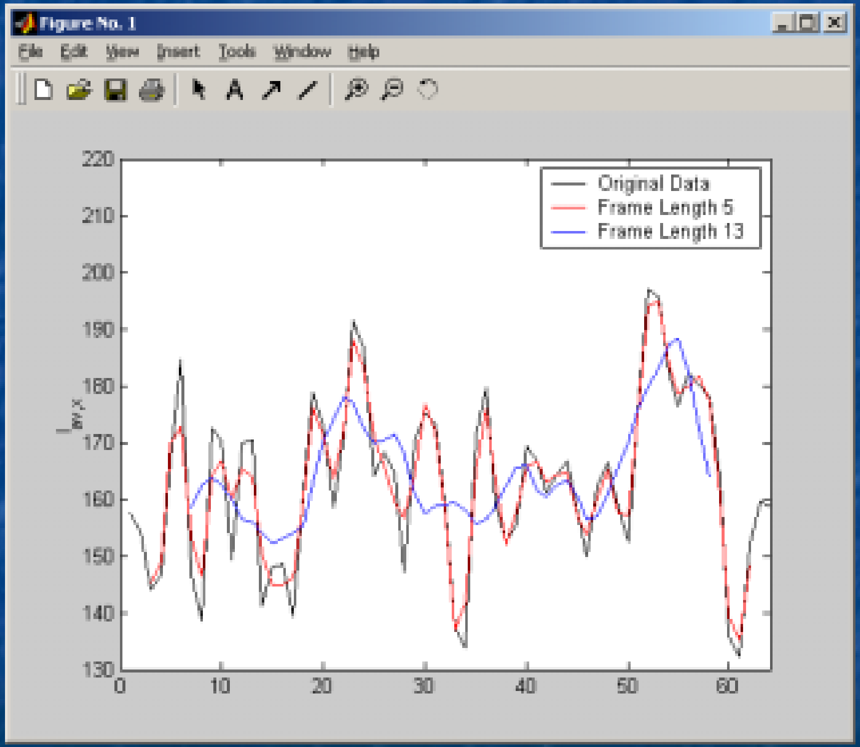
- $\hat{\mathbf{w}}[5] = 5.229$
- $\hat{\mathbf{w}}'[5] = 4.300$
- $\hat{\mathbf{w}}''[5] = 0.786$

Review of Savitzky-Golay Filter

- In practice, filtering is performed by a single matrix multiplication
  - No loop necessary to move the frame
- The resulting matrix contains the filtered data, the first derivative of the filtered data, and all other order derivatives up to the order of the polynomial used in the least-squares fit

Effect of Frame Length

- Generally, D_x and D_y remain banded for different frame lengths
- However, smaller frame length works better
  - Fine structure of characters is lost with large frame length



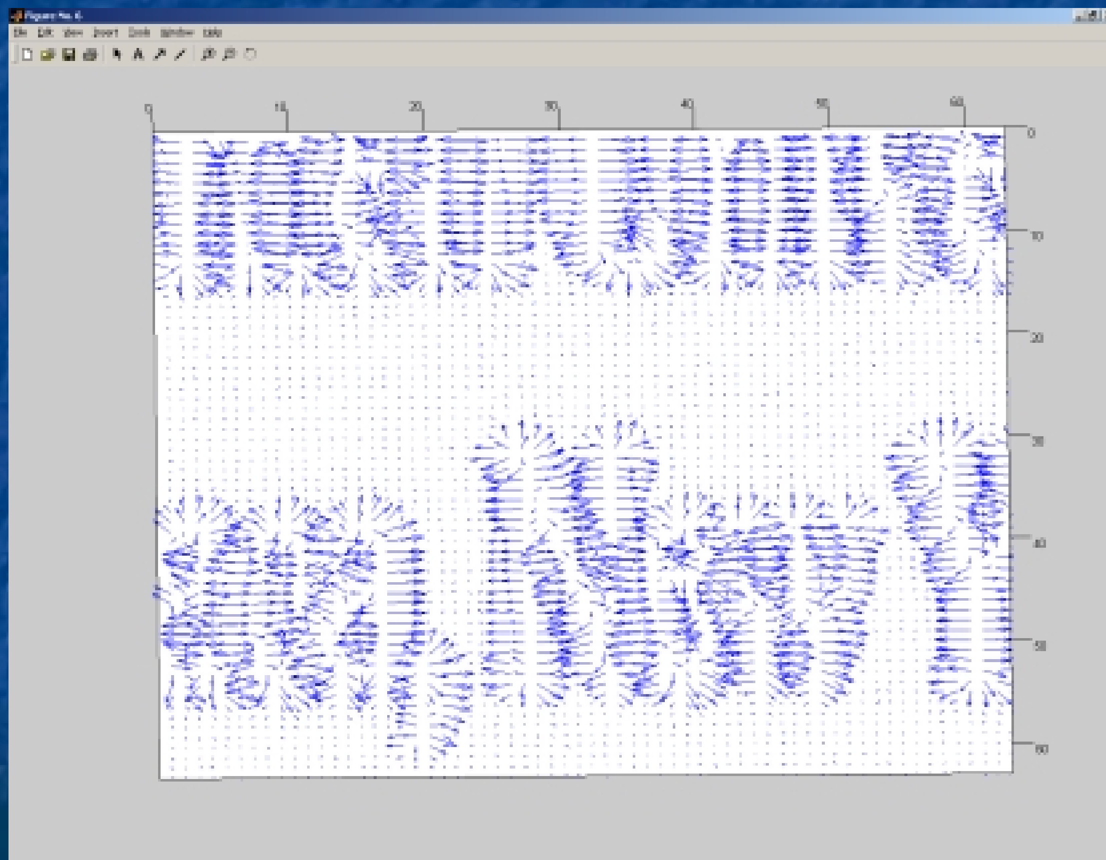
Evaluation of D_x and D_y

- D_x and D_y are useful metrics to separate text from pictures and background
- Edgy pictures will cause failure

Gradient Vector Direction

- Text has maximum gradient in the four cardinal directions $-\pi/2$, 0, $\pi/2$, and π
- Pictures have no direction in which gradient is expected to be maximum

Illustration of Gradient Vector Field



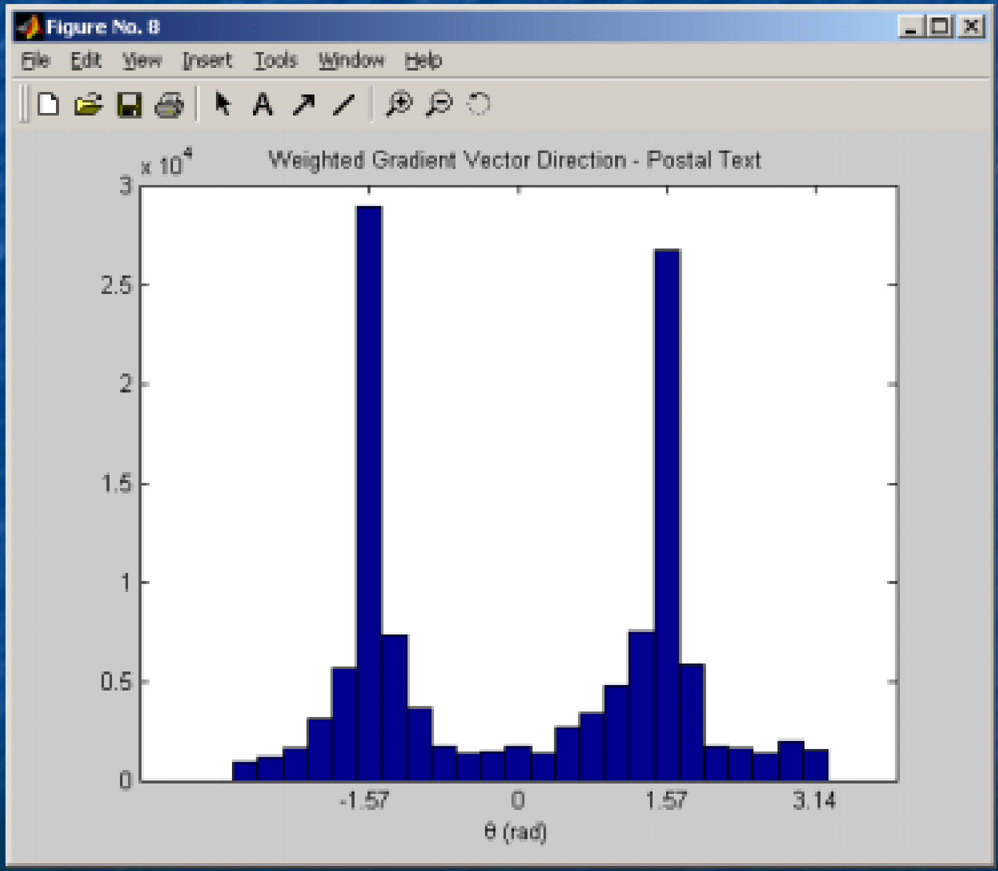
Gradient Vector Direction – $\theta_{i,j}$

- Definition:
  - $\Delta h_{i,j} = I_{(i+1),j} - I_{i,j}$ (Horizontal Gradient)
  - $\Delta v_{i,j} = I_{i,(j+1)} - I_{i,j}$ (Vertical Gradient)
  - $\theta_{i,j} = \tan^{-1}(\Delta v_{i,j} / \Delta h_{i,j})$
    - $-\pi < \theta_{i,j} \leq \pi$
    - i, j: 1, 2, ..., N-1

g_c – Cardinal Gradient Vector Magnitude

- Gradient vectors whose θ falls within a 15° range around the four cardinal directions is classified as a cardinal gradient vector.
- I propose a classification feature that is the summation of cardinal gradient vector magnitudes (g_c).

Histogram of Gradient Vector Direction Weighted by Magnitude



Evaluation of g_c

- Text and pictures can be classified using this feature
- Prone to error in rotated images
- Relatively fast to compute

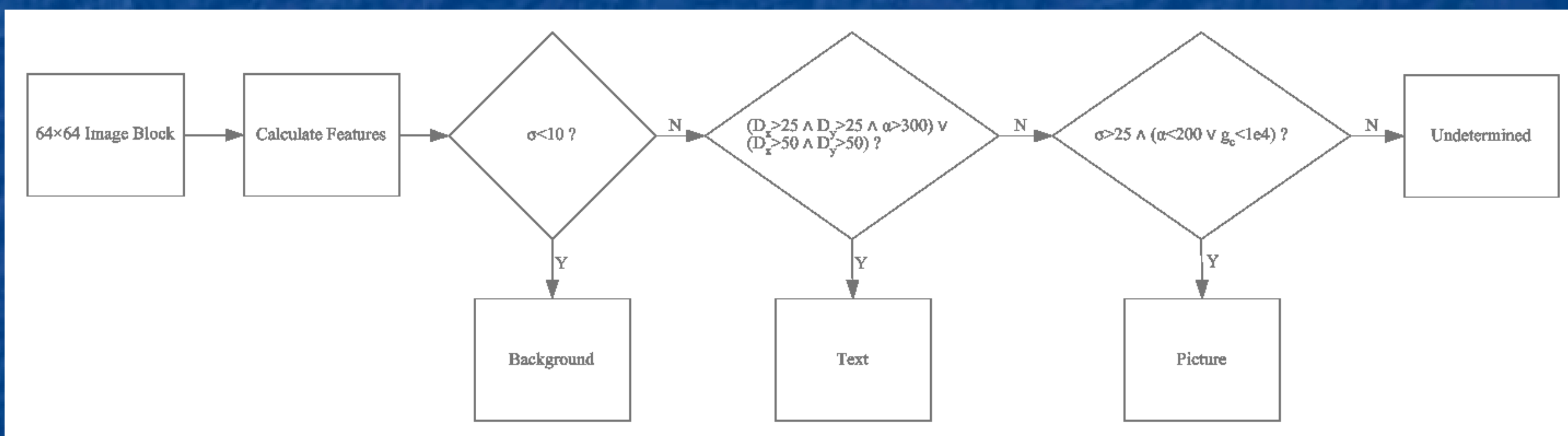
Summary of 6 Features

- μ: three bands, but prone to error depending on lighting, background color, etc.
- σ: good for separating out background
- α: good for separating out text
- D_x, D_y: three bands, but prone to error on edgy pictures
  - Not prone to error based on lighting and background color
- g_c: good for separating pictures and text, but prone to error on misaligned pictures

6 Dimensional Feature Space

- There are numerous ways to classify image blocks
- Too many ways to look at
- I developed a decision tree which is hardly optimal, but still gives good results
  - With more experimentation and better techniques, the decision rule can be significantly improved

First Pass Classification: Decision Tree

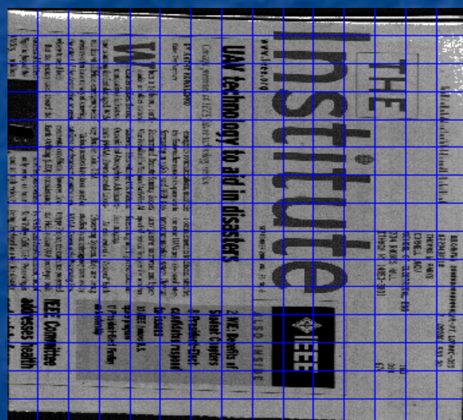


Comparison of Classification Techniques

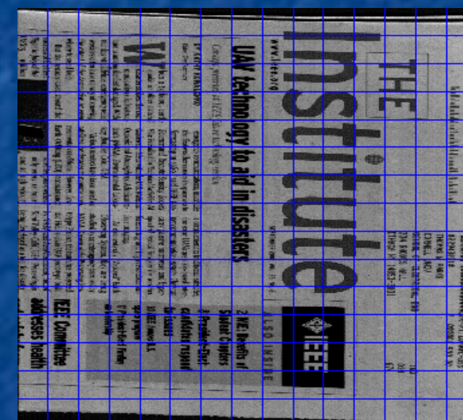
- Wavelet approach is computationally slow, whereas 6-D approach is fast
- 6-D approach can be improved by better decision rule, whereas wavelet approach cannot be improved much
- Comparative results will be shown in the following slides

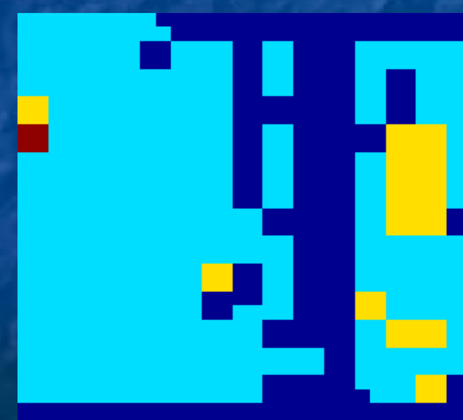
Comparative Results

Wavelet



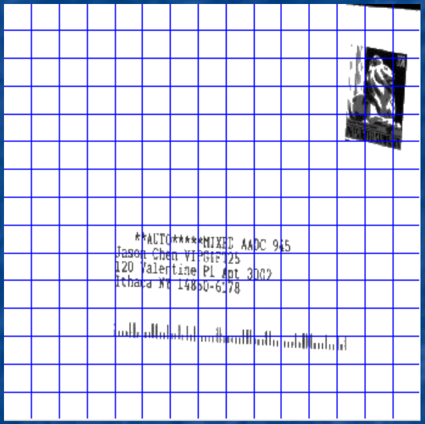
6-D

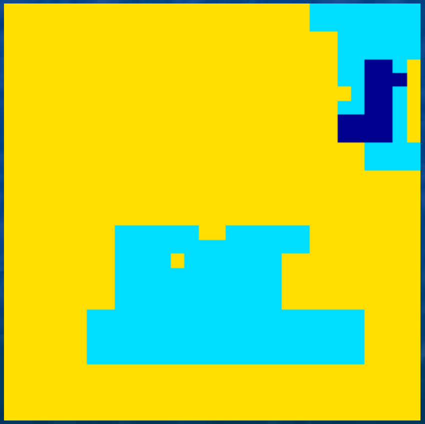




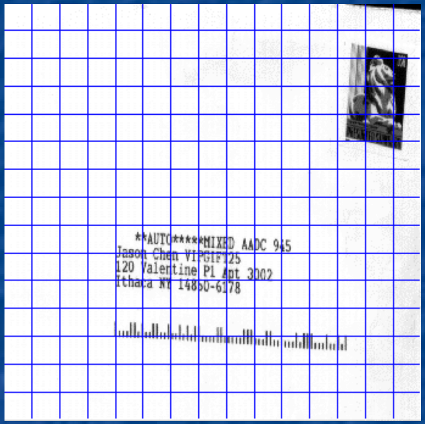
Comparative Results

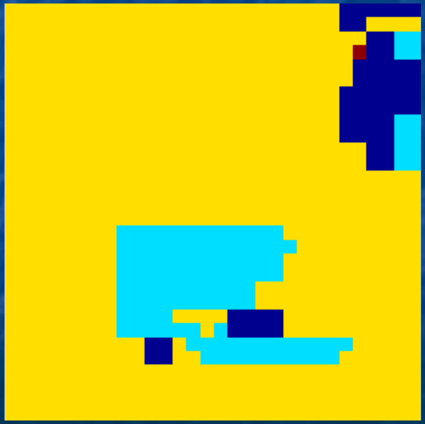
Wavelet





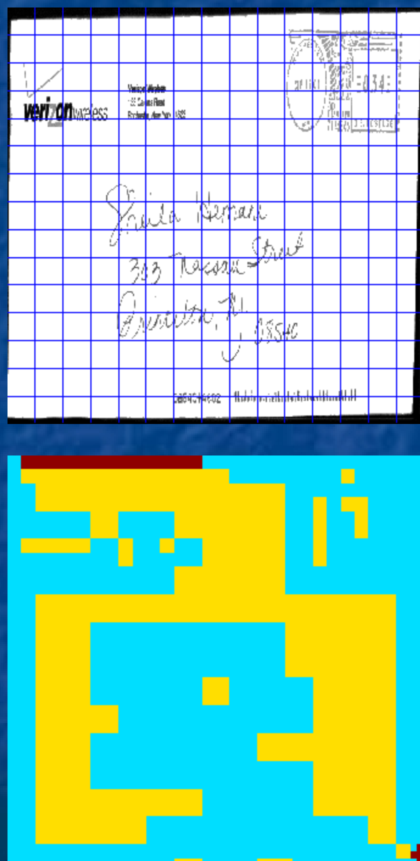
6-D





Comparative Results

Wavelet



6-D

Table of Results

	Wavelet	6-D	Wavelet	6-D
ID	% Error		Execution Time	
1	34.67	23.06	157.76	5.81
2	39.17	28.50	148.37	5.64
3	29.78	21.94	183.01	6.65
4	40.11	31.53	176.66	5.26
5	43.86	36.69	237.90	5.94
6	35.08	33.33	211.75	4.62
7	26.72	25.67	225.68	4.42
8	14.44	9.00	174.21	3.54

Final Discussion

- Many false classifications occur at the envelope boundary
  - Wavelet approach classifies as text
  - 6-Dimensional approach classifies as picture
- Other deviations from hand segmentation are a matter of precision rather than accuracy
- Remaining misclassifications are true errors

- Classification error is worse with wavelet-based approach for all test images
- Execution time is 25 times longer for wavelet-based approach
- Classification error is still not acceptable
  - Can be improved by better decision rule