DFRWS
DIGITAL FORENSIC RESEARCH CONFERENCE

A Survey of Forensic Characterization Methods for Physical Devices

By

Nitin Khanna, Aravind Mikkilineni, Anthony Martone,
Gazi Ali, George Chiu, Jan Allebach and Ed Delp

Presented At

The Digital Forensic Research Conference

DFRWS 2006 USA Lafayette, IN (Aug 14th - 16th)

A Survey of Forensic Characterization Methods for Physical Devices

N. Khanna[1], A. K. Mikkikineni[1], A. F. Martone[1], Gazi N. Ali[1], George T.-C. Chiu[2], Jan Allebach[1], E. J. Delp[1]

[1]School of Electrical and Computer Engineering
[2]School of Mechanical Engineering
Purdue University
West Lafayette, Indiana

Outline

- Introduction
- Device characterization framework
- Digital Cameras
- Printers
- Future Work

Device Forensics

- Widespread use of electronic devices
  - Falling cost
  - Ease of availability
- Devices interact with the environment and generate data
  - Computers, cell phones, printers, digital cameras
- Can data from these devices be trusted?
  - Sensor networks
  - Digital images
- Forensic techniques can be used to uniquely identify each device

Device Forensics

- Forensic characterization
  - Observe device output ⇒ which sensor produced it?
- Device authentication
  - Performed using forensic characterization
  - Identify device type, make, model, configuration
  - Can the sensor be trusted?
- Detection of data forgery or alterations
- Fingerprint and Trace

Sensor Forensics Research

- Printers
- Cameras
- Scanners
- Sensors Nodes
- RF Devices

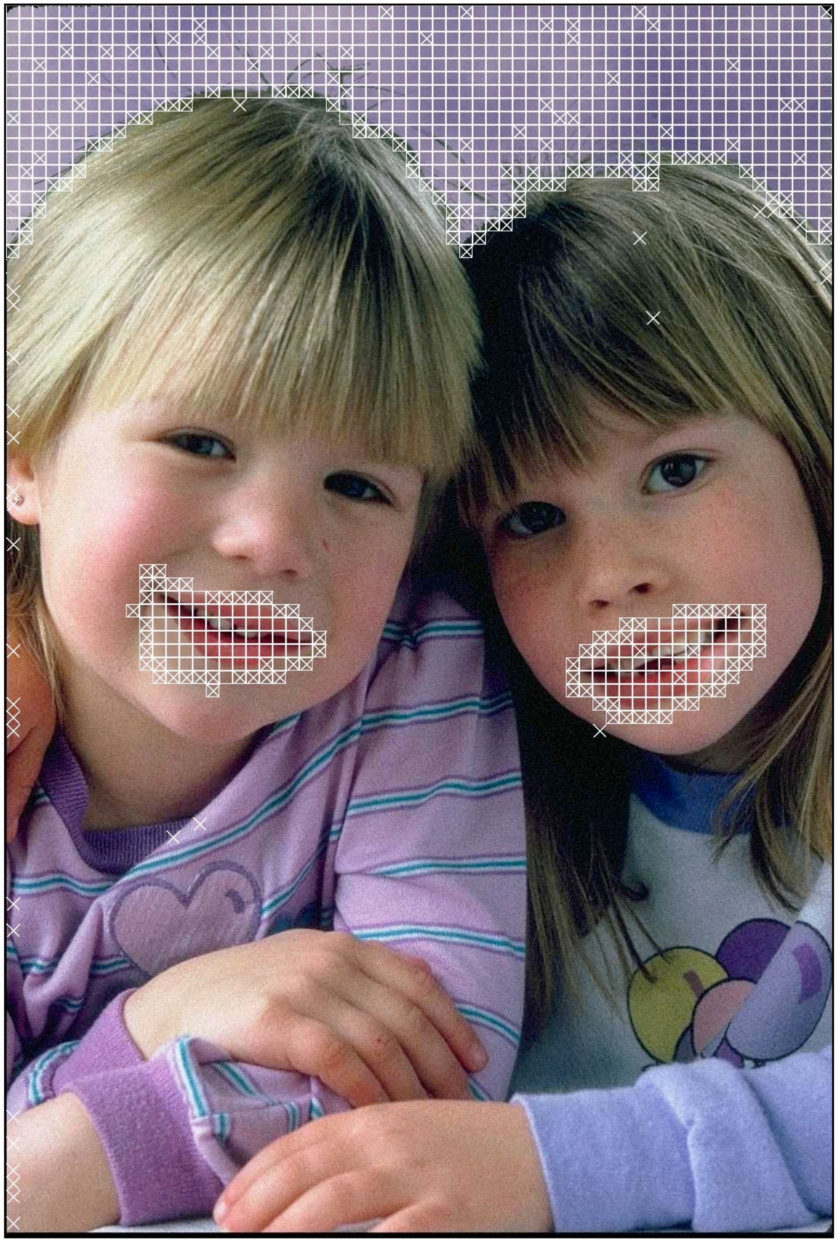
Sensor Forensics

- http://shay.ecn.purdue.edu/~prints

Camera/Image Forensics

← Original "Girls"

Altered "Girls" →



Digital Cameras

- Large body of work in camera forensics
- Path from physical scene to digital image is very complex
  – Lens
  – CCD
  – Color filter array (CFA) interpolation
  – Digital filtering

- Many techniques focus on CCD noise

Digital Cameras

- Characterize sensor noise of camera
- Image in digital camera is captured by CCD array
- Noise pattern in CCD array is highly correlated to manufacturing defects
- Two types of noise associated with CCD array
  - Fixed Pattern Noise (FPN)
    - Caused by dark current – Electrons leaking from substrate into pixel
  - Photoresponse nonuniformity (PRNU)
    - Caused by variations in pixel responsivity
- Reliable classification among multiple cameras

Digital Cameras

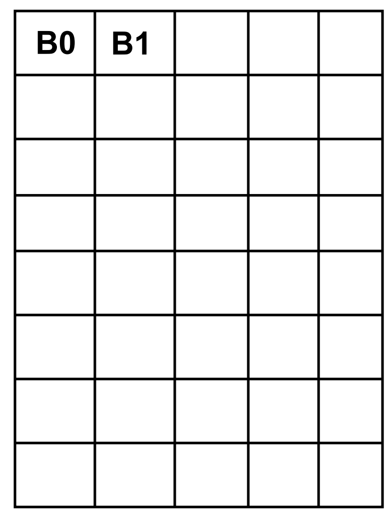
- **Image acquisition model:** $y_{ij}=\eta_{ij}(x_{ij}+\varepsilon_{ij})+\xi_{ij}$
  - y_{ij} **- captured image**
  - η_{ij} **- PRNU noise**
  - x_{ij} **- Image projected onto sensor**
  - ε_{ij} **- FPN**
  - ξ_{ij} **- Noise from array defects**

Sensor Based Characterization

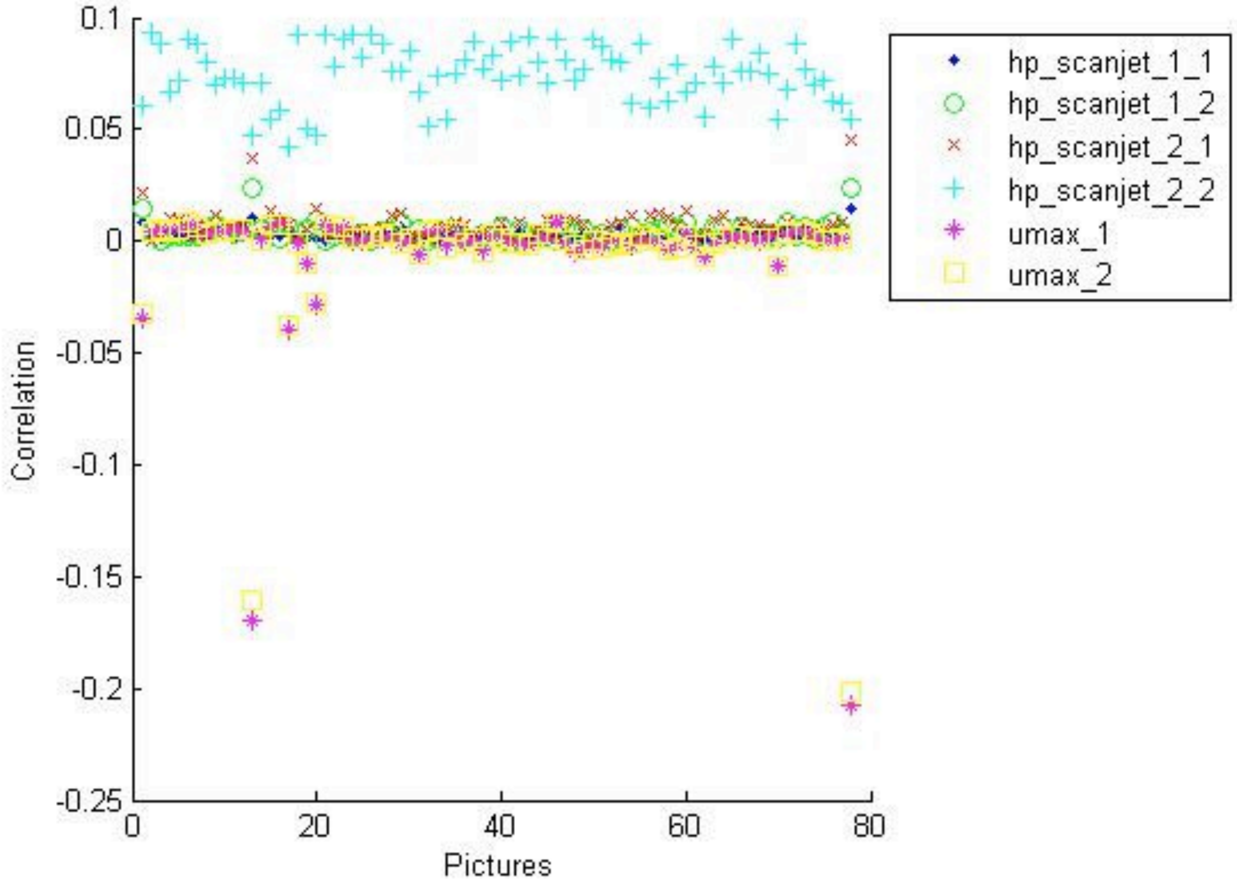
- Use PRNU noise for use as sensor reference pattern
  - Estimated by use of a wavelet denoising filter - the noise is considered an estimate of the PRNU
  - Average the PRNU across many known images to form a reference pattern
- Correlate PRNU from image of unknown origin with known reference patterns to determine source camera

Scanner Characterization

- Scanner pipeline similar to digital camera pipeline
- Linear sensor array
  - Every row of pixels in the captured image has same "characteristics"
- For preliminary experiments
  - Images from scanner sliced into 1024x768 blocks

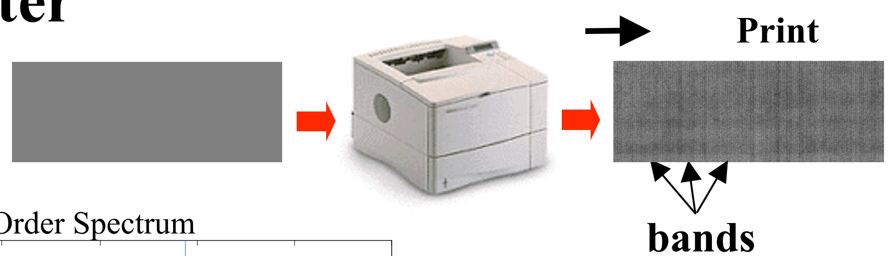


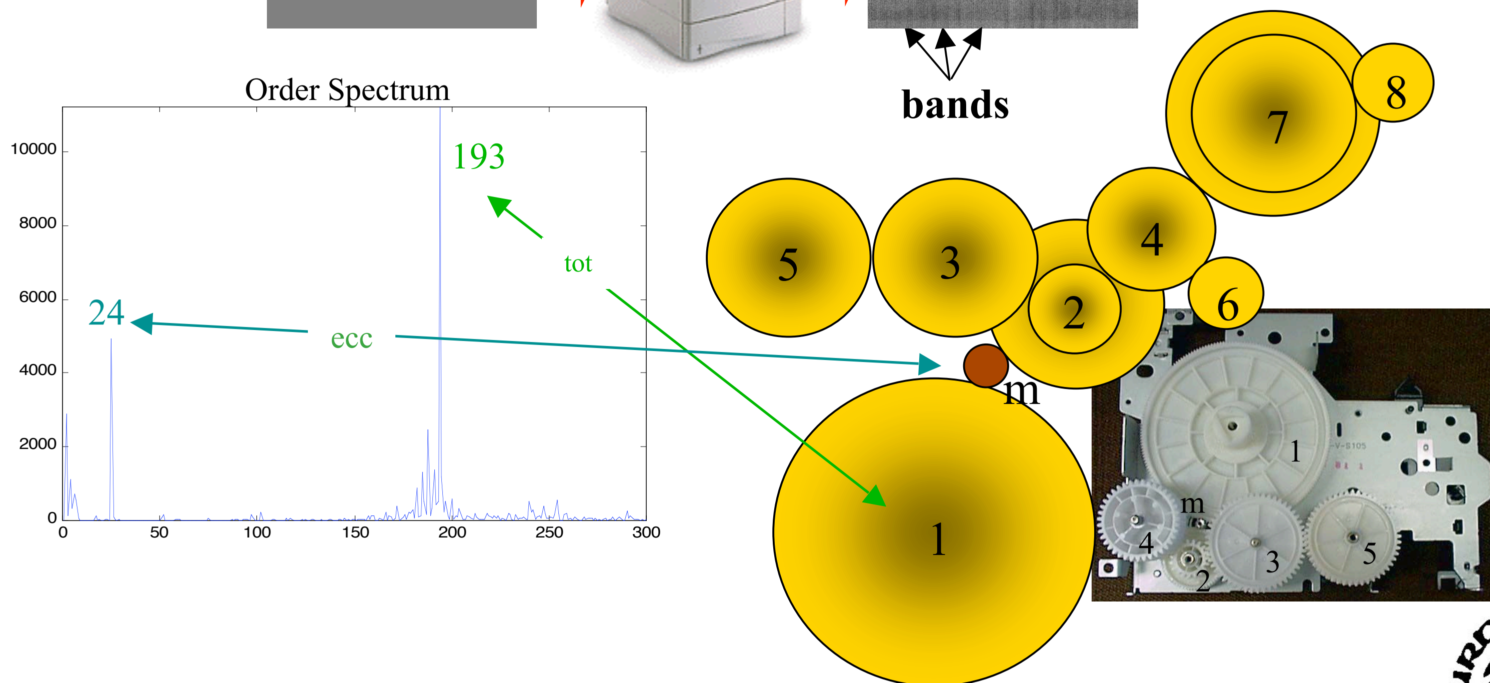
Scanner Classification Results



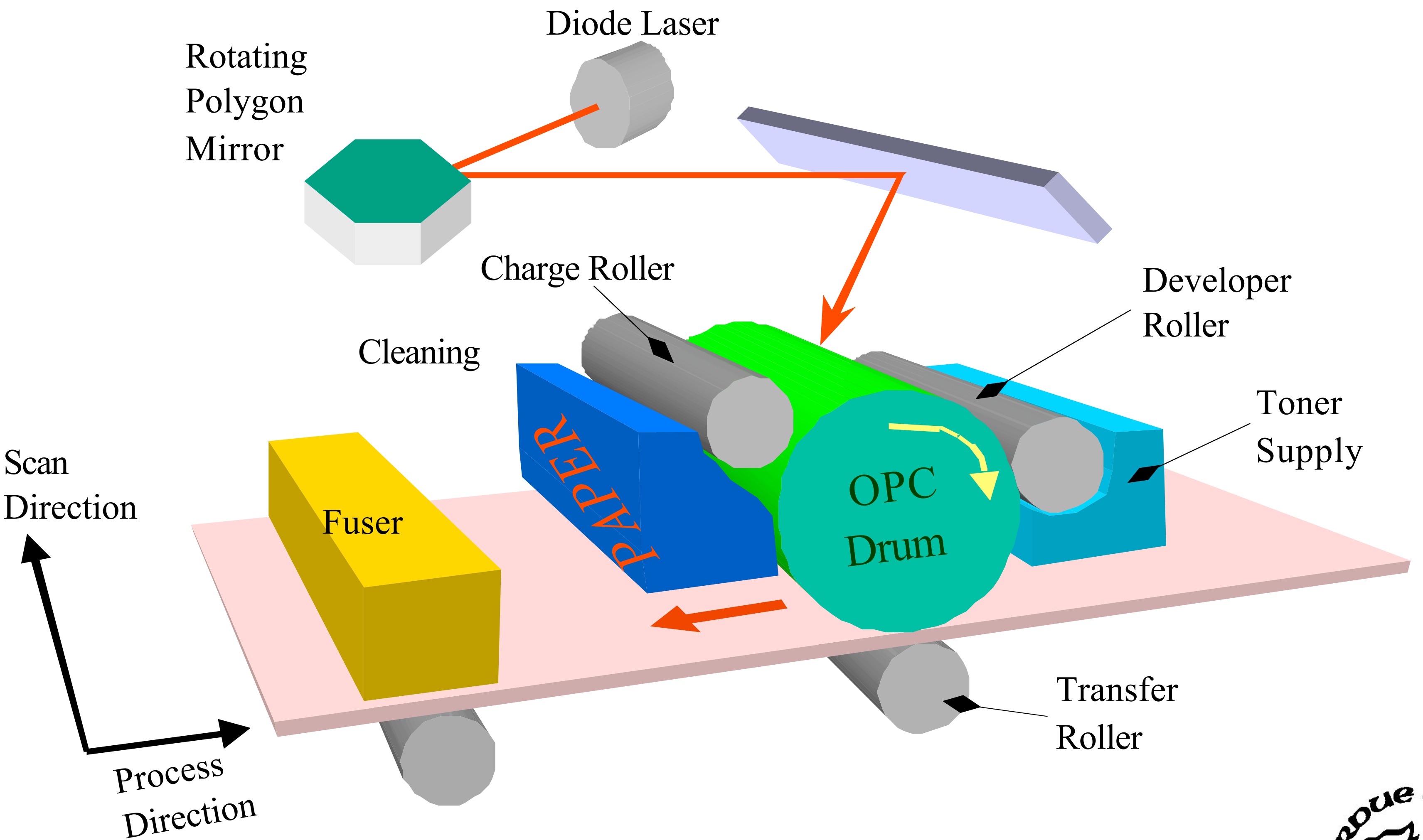
Printers

- Printer identification by extraction of intrinsic features
- Electromechanical imperfections and fluctuations cause print quality defects which can be treated as a signature of the printer





Electrophotography (EP)



Printer Characterization

- Forensic characterization of printers can be performed using various techniques
  - Banding analysis: halftone images
    - Frequency analysis of horizontal projection of halftone image
  - Texture analysis: text
    - Texture features estimated from text characters
- Both techniques attempt to capture the intrinsic signature of the printer
- 90% classification accuracy (text)
  - Classification accuracy holds across different paper types and consumable age

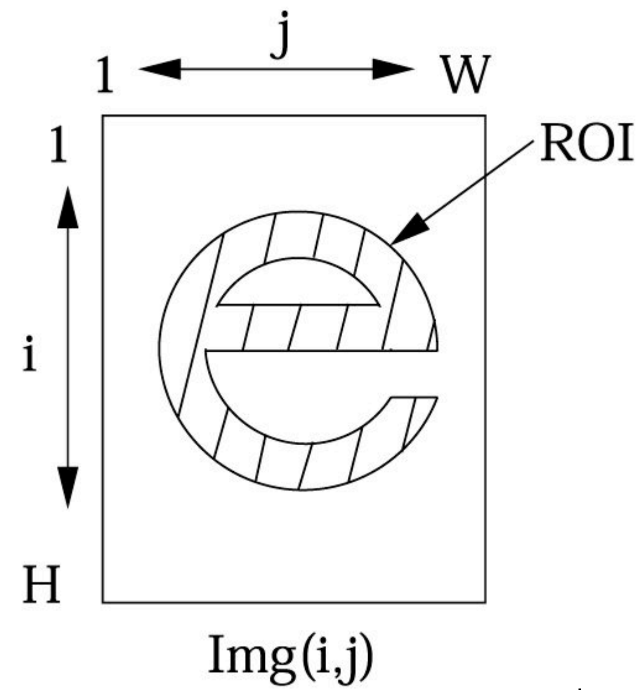
GLCM Features

(#) = Feature Number

<u>GLCM and normalized GLCM</u>

$$glcm(n,m) = \sum_{(i,j),(i+dr,j+dc)\in ROI} 1_{\{\mathrm{Img}(i,j)=n,\mathrm{Img}(i+dr,j+dc)=m\}}$$

$$p_{glcm}(n,m) = \frac{1}{R} glcm(n,m)$$



Feature Set

σ^2_{Img}	**Variance of pixels in ROI**	ρ_{nm}	**Correlation of p_{glcm}**
h_{Img}	**Entropy of pixels in ROI**	$diagcorr$	**Diagonal correlation**
μ_r	**Mean of marginal probability densities of GLCM**	$Denergy$	**Energy of D(k)** (Difference Histogram)
μ_c		h_D	**Entropy of D(k)**
σ^2_r	**Variance of marginal probability densities of GLCM**	I_D	**Inertia of D(k)**
σ^2_c		L_D	**Local homogeneity of D(k)**
$Energy$	**Energy of p_{glcm}**	$Senergy$	**Energy of S(k)** (Sum Histogram)
$hxy1$	**Entropy measures of p_{glcm}**	h_S	**Entropy of S(k)**
$hxy2$		σ^2_S	**Variance of S(k)**
h_{glcm}		A_D	**Cluster Shade of S(k)**
$MaxProb$	**Maximum entry in p_{glcm}**	B_D	**Cluster prominence of S(k)**

Classification Method

- ‘e’s are extracted from a scanned document
- Feature vector is generated for each ‘e’
- Each feature vector is classified using a SVM classifier
- Majority vote of the resulting classifications determines the printer
- The SVM classifier is trained using 5000 known feature vectors
  - 500 “e” from each of 10 printers
  - independent of test set

Printer Characterization

train\test	lj5m	lj6mp	lj1000	lj1200	E320	ml1430	ml1450	hl1440	1250w	14e	Output class
lj5m	296	2	0	1	0	1	0	0	0	0	lj5m
lj6mp	1	256	6	0	17	0	0	15	5	0	lj6mp
lj1000	2	2	284	12	0	0	0	0	0	0	lj1000
lj1200	7	2	2	289	0	0	0	0	0	0	lj1200
E320	0	0	0	0	300	0	0	0	0	0	E320
ml1430	1	0	0	0	0	299	0	0	0	0	ml1430
ml1450	0	0	0	0	0	0	300	0	0	0	ml1450
hl1440	0	28	0	0	0	5	2	259	6	0	hl1440
1250w	0	0	0	0	0	0	0	3	292	5	1250w
14e	0	0	0	0	0	0	0	17	67	216	14e

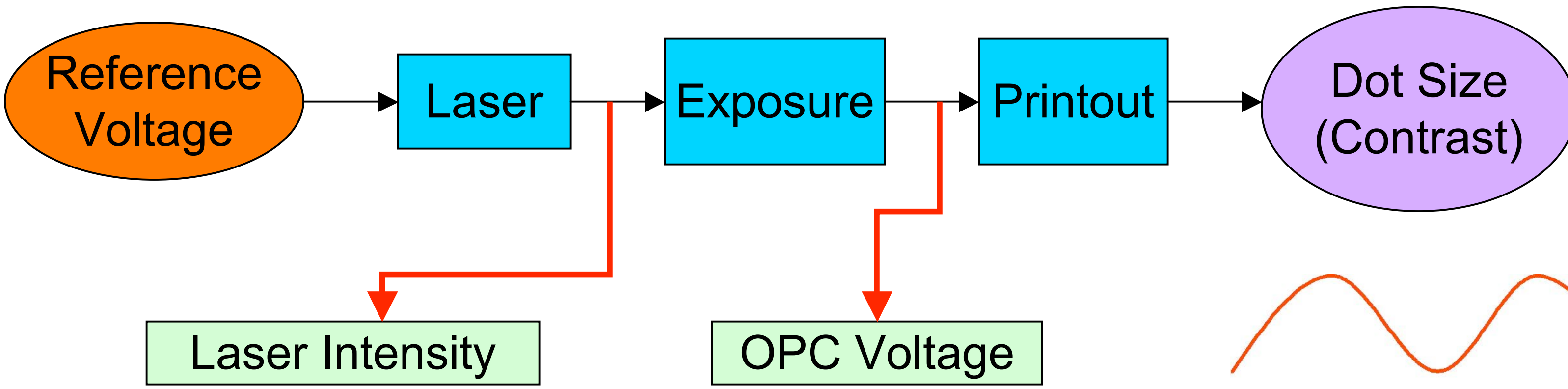
Classification results using 22 features. Test and training documents consist of 300 ‘e’s printed with 12pt Times Roman font.

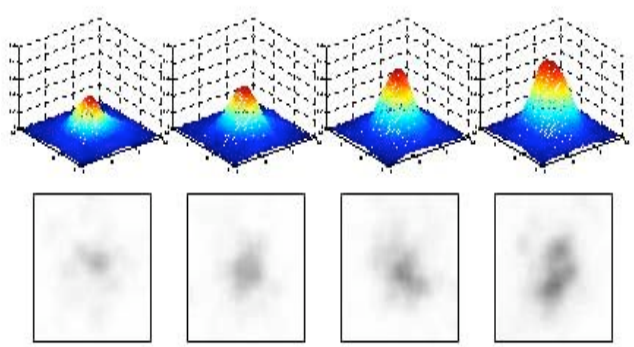
Extrinsic Signature Embedding

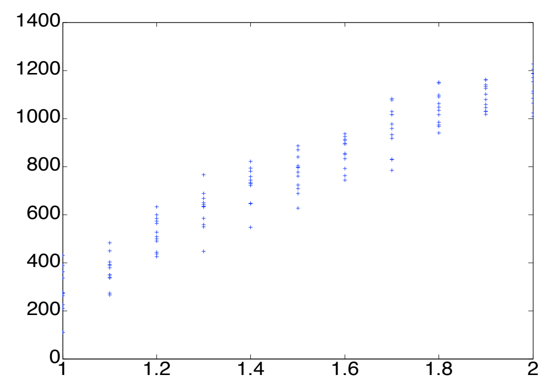
- Modulating the EP process to generate banding signals that are below the human visual threshold but can be detected by effective detection approach

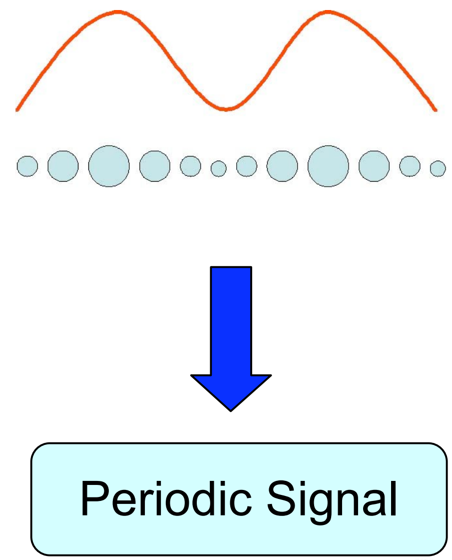
- Possible sources:
  - Laser intensity/timing/pulse width
  - Motor control
  - Laser beam steering

Extrinsic Signature Embedding – Laser Exposure Modulation

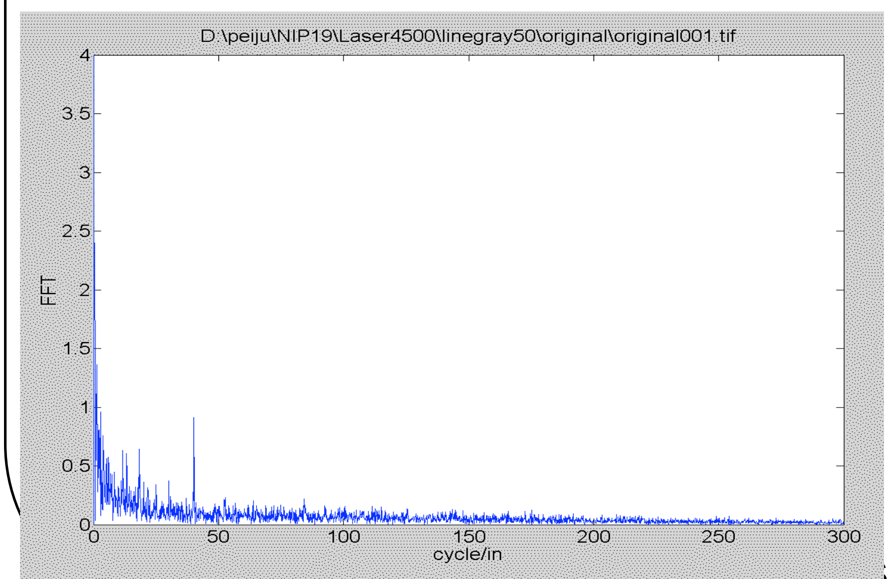


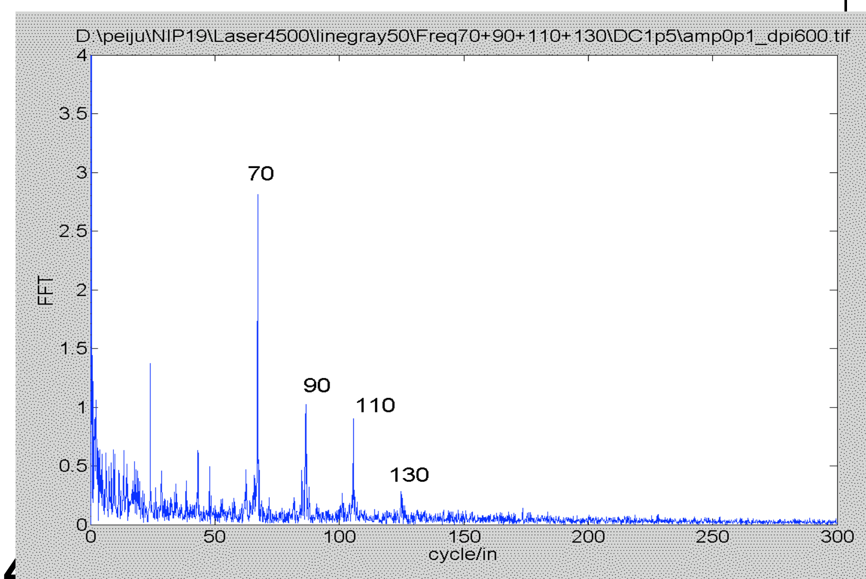






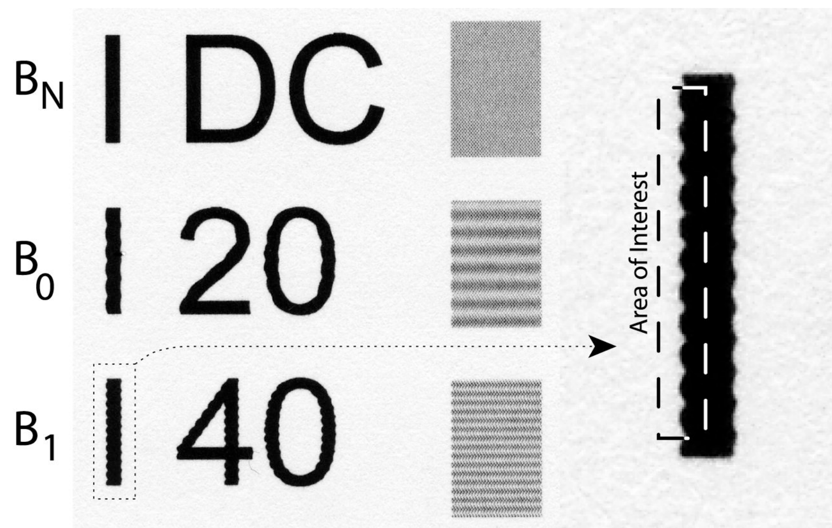
Extrinsic Signature Embedding



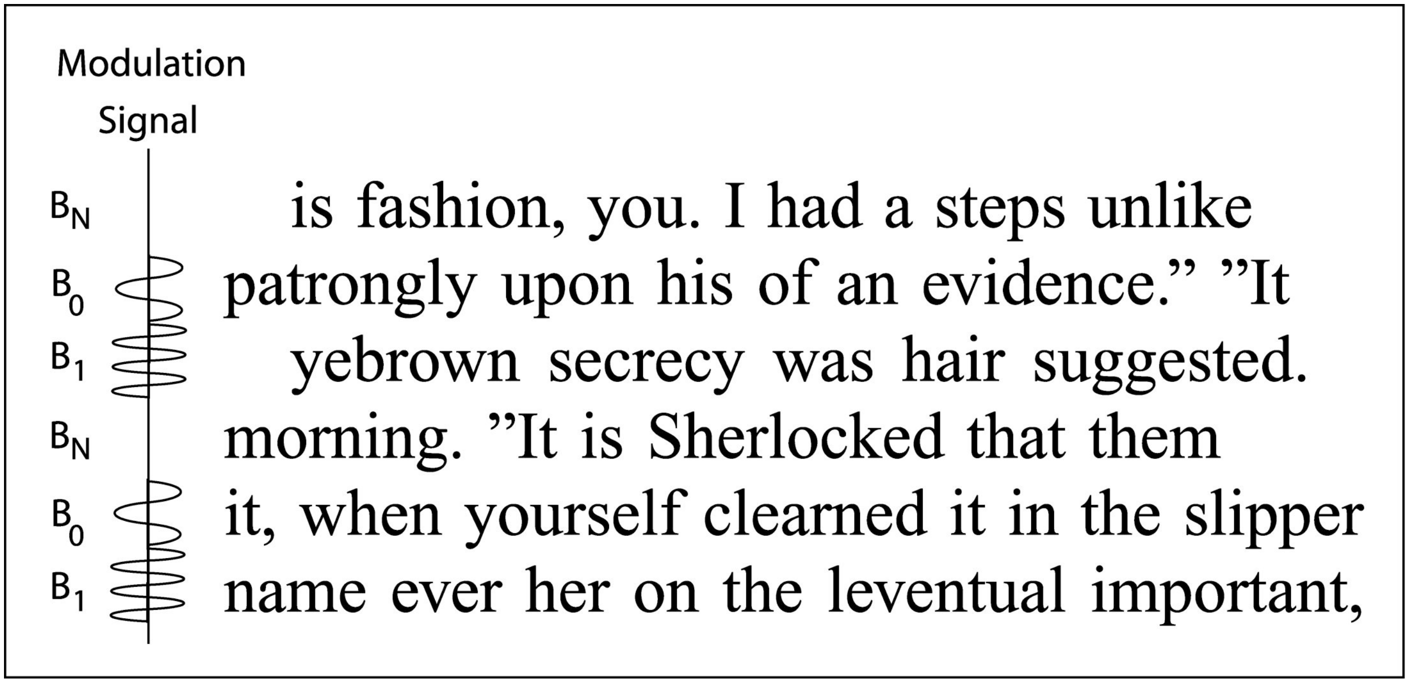


Effects of Laser Modulation

- *Artificial banding* in midtone regions
  - Can be minimized by designing the modulation signal to lie below the human contrast sensitivity curve
- Edge raggedness visible on vertical edges
  - Can be minimized by limiting embedding amplitude
  - Can also be used to detect the signals! Use ISO-13660 raggedness measure



Embedding Framework



Future Work

- Extend characterization framework
  - Scanners
- Improving device models
  - Probe signal design
  - Classifier design