Forensic Markings for Progressive Barcodes

Steven Simske, Marie Vans, Stephen Pollard & Guy Adams
HP Labs

Outline

- Barcodes
  - Standard & Non-Standard Barcodes
  - Color Tiles, IIOs, & Progressive Barcodes
- Tri-Purpose IIOs
  - Forensics with SDED
  - Applications of SDED
- Results & Conclusions

Barcodes

- 1 Dimensional
- 2 Dimensional
- 3 Dimensional
- 4 Dimensional

Barcode Standards

- Data Matrix: high density marks
  - Electrical components, food industry
    - ISO/IEC 16022:2006 - Data Matrix bar code symbology specification
    - Data capacity: 2335 bytes
- Aztec: marks with no quiet zone
  - Transportation/Travel industry, airline & train tickets
    - ISO/IEC 24778:2008 - Aztec Code bar code symbology specification
    - Data capacity: 1914 bytes

Barcode Standards - Continued

- QR-Code: probably the most ubiquitous
  - Marketing, URLs to Websites
    - Data capacity: 3706 bytes

Non-Standard Barcodes

- ColorTile
- Progressive Barcode

The example shown here comprises:
1. 84 data tiles, max 165 bits total
2. 8 non-payload tiles:
   a. Two black for orientation & corner detection
   b. 6 color (RGBCMY) for color calibration
   c. Colors are 180 rotated from their color opponency pair, providing the greatest contrast in hue space and thus the most reliable opposite-corner orientation detection possible
IIOs – Statistical Probability

- Progressive barcodes instantiation of IIO: Incremental Identifying Object
- Statistical probability assigned at each step of the progression
- Associated with any transition between two steps in a workflow
- Based on how many bits are written and how many remain
  \[ \frac{N_{IU}!}{N_{RB}!} \geq \prod_{i} P_i \]
  
  \[ P_i = \text{Step } i \]
  
  \[ N_{IU} = \text{the number of residual (0 bits) at end of the workflow} \]
  
  \[ N_{RB} = \text{number of initial unwritten bits} \]

- If barcode is unique at step \( i-1 \), then total number of barcodes of the current state is 1.
- If the progressive barcode is binary, then the number of bits in the workflow is \( N_{RB} - N_{IU} \).
- If there are \( N_c \) colors, number of bits increases to:
  \[ \left[ \frac{\ln(N_c)}{\ln(2)} \right] \times (N_{RB} - N_{IU}) \]

- The size (height and width in tiles) of progressive barcode used in the workflow can be determined from these equations along with the number of bits to write at each state.

IIOs – Data Capacity

- Each color tile independent
- Define a tile to be \( N \)-ary, where \( N = \) number of colors allowed at each tile:
  \[ \log_2(n) / \log_2(2) = \log_2(n) \text{ bits at any stage} \]
- \( n = 2 \), e.g. DataMatrix, QR, Aztec: 1 bit per tile
- Color tile with six colors (RGBCMY): 2.585 bits/tile
- 8 colors allowed (RGBCMYWK): 3.0 bits/tile
- Color tile barcode \( X \) data tiles wide \& \( Y \) data tiles high contains exactly:
  \[ XY \log_2(n) \text{ bits} \]

Standard Barcodes with Progressive Barcodes

- Allows use of “static” data encoded in black & white modules for standard purposes:
  - point of sale
  - serial numbers (serialization)
  - product information
- Allows “separate channel” for encoding changing workflow-related information
- Static data: off-the-shelf reader reads the “black-as-black” modules and the “rest-as-white” (white is a “dual channel”)

Example Progression

- Reader-Friendly Progressive Barcodes
- Columns indicate progression along workflow
- Colors saturated enough so that 2D barcode reader still interprets them as “white”
- 2D barcode reads the same throughout the progression.

Example GS1 Workflow

- GS1 world-wide track and trace standards
- Progressive barcodes for GS1 product workflows using multiple barcodes
- Original GS1-compliant barcode contains a product ID and remains static as its second channel of content progresses
- Concurrently, information encoded in the colors changes at each step.
- The data within the white-as-N-ary modules used for data normally encoded in separate barcodes
Tri-Purpose Barcodes – Adding Forensics

- Static black & white + color progression + forensics
- Progressive barcodes provide a secure means to transition from node to node in a workflow
- Does not incorporate the physical attributes of the current printed barcode.
- The shape distortion encoding difference (SDED) approach solves this issue.

SDED - Continued

- Sum in $SSE_{Det}$ is divided by $40n$ to determine the atomic unit of encoding
  - $n$ is the number of samples of $p$ per element
- A 40-position string, $P$, is created for the deterrent (progressive barcode)
  - dividing SSE of each element by atomic unit & rounding
- SDED for comparing any two deterrents defined as:
  \[
  SDE_{\text{D}} = \sum_{j=0}^{39} |P(j) - P_{\text{det}}(j)|
  \]
- A form of modified Hamming Distance: expected value of $P(*)$ is 1 at each element.

SDED Workflow

1. Capture previous IIO while capturing the document, compute $P$ string, e.g. [0300100100401010002012301240502040120].
2. Expand the IIO into a binary form (simplified one shown), (0111001001001111010100011001111110111111100111110011011110111011).  
3. Use this string (with padding, deleting, scrambling, encryption, or other encoding techniques needed for specific workflow & IIO type) to determine new information for the IIO.
4. Archive previous image.
5. Print new document with altered IIO (and other altered workflow data) included.

Note that this new document has cleaned-up IIO and new forensic boundary, which is not reflected in the IIO (yet) in the scenario depicted.

This approach resets the forensic information for IIO boundary each time there is a new document.

SDED Applications

- Combine barcodes directly with the current forensics.
  - Reading device collects forensics at same time as current image of the IIO
  - Incremental elements written to IIO reflect the forensics of original IIO, which “never change”.
  - There is only ever one physical item printed to—such as a unique label, packaging, special document, etc.
    - Encryption/signing of the incremental data generally required
    - Note that image registration issue here is not solved

SDED Example

- Original image $P$: [0300100100401010002012301240502040120]
  - implies the original image had significant variability at positions 1, 11, 24, 28, 31 and 36 since the encoding is for 3, 4, 3, 4, 5 and 4, respectively, at those positions.
- Same barcode imaged later, some variability in the encoding likely
  - E.g. the following may be recorded:
    - Second image $P$: [0210100000500110003001130115006002030120]
  - The modified Hamming Distance between the two is 12.
  - Threshold between matching images and non-matching images difficult to replicate during printing, difficult to tamper with

Shape Distortion Encoding Difference-SDED

- Divide idealized perimeter of the barcode
  - E.g. 10 line segment elements along each of the 4 sides
- One element for each of 40 exposed “sides” of one of 36 edge modules of 100 module example.
- Compute a sum squared error (SSE) of the residual, of some image related metric $p$, for each of the 40 elements
- Overall SSE of progressive barcode designated $SSE_{Det}$ defined as:
  \[
  SSE_{Det} = \sum_{i=0}^{39} \left( (p(i) - \mu_{\text{element}})^2 \right)
  \]
- $p(i) = \text{orthogonal displacement with respect to the deterrent model of each point on the perimeter for a particular element}$
- $\mu_{\text{element}} = \text{mean of such over the whole of that element}$
SDED Applications - Workflow kiosk

- All-in-one or copier
  - workflow-enhancing appliance or dual path print required

- Steps:
  1. Print document with IIO updated to previous stage
  2. Capture IIO and compute its forensics, e.g. sequence P
  3. Convert P into additional IIO modules
     Note: all conversions are possible here:
     E.g. we can convert magenta to red by overprinting with yellow, etc.
  4. Run the page through workflow kiosk (e.g. laserjet printer) again,
     register & overwrite IIO to make it the current IIOfm
     Note: Registration aided by page find/placement information coming
     from scanning platen

Conclusions

- Progressive Barcodes/IIOs can significantly increase data density
- SDED approach can be used to generate secure IIOs for document workflows
  - A series of experiments have been performed between false and valid images using a very high resolution scanning device
  - Number of segments N and atomic unit of coding was varied
  - Results showed that by using the SDED measure on between 50 and 400 samples of the string P best forensic security (at 200 samples) had a probability of false detection less than $10^{-9}$.
  - Encoding the position string P into the payload modules of the barcode can provide authentication and forensic capabilities on the fly.

Questions?

Thank you for your kind attention