Digital Multimedia Forensics and Anti-Forensics

Matthew C. Stamm

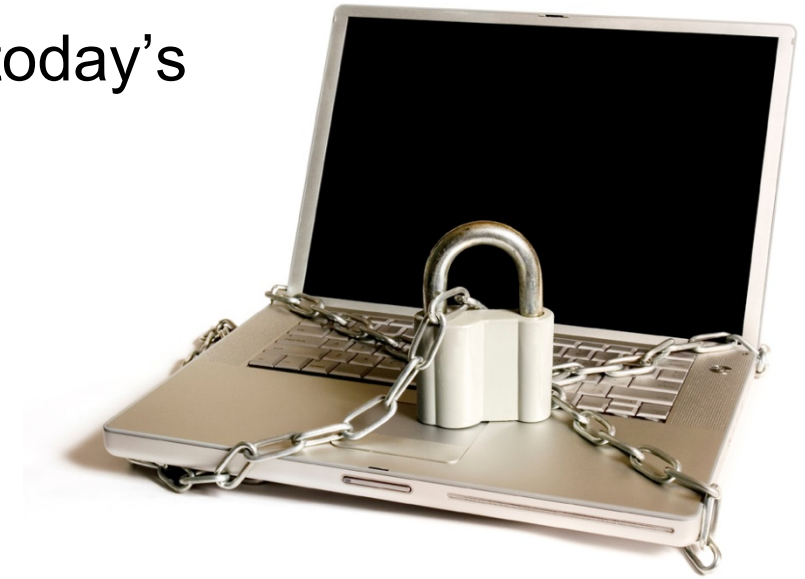
Signals and Information Group

Department of Electrical and Computer Engineering

University of Maryland, College Park

Information Security

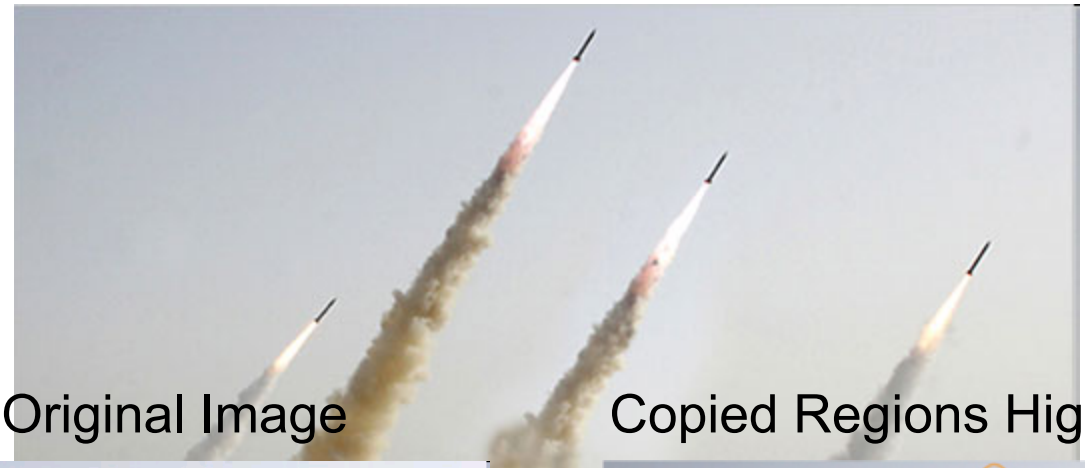
- Digital multimedia information is everywhere
- Information security is one of today's greatest challenges
- Cryptographic approach
  - Prevent unauthorized access
  - Secure transmission



- *What if information is manipulated first?*

Digital Forgeries

- Editing software can create *perceptually realistic* digital multimedia forgeries



Original Image

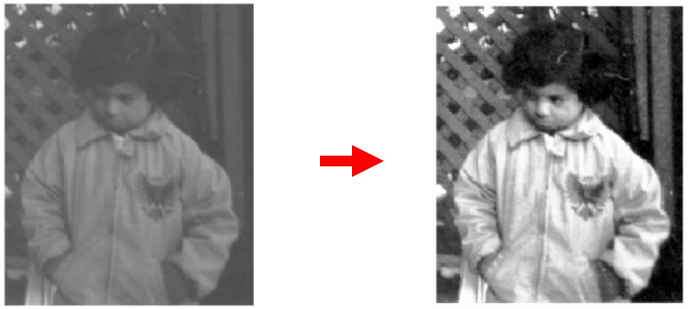
Copied Regions Highlighted

Information Forensics

- Solution: *Information forensics*
  - Identify editing
  - Verify authenticity
  - Determine origin
- Provide information security when the *information source is not trusted*
- Detect traces known as *fingerprints* left by information manipulation

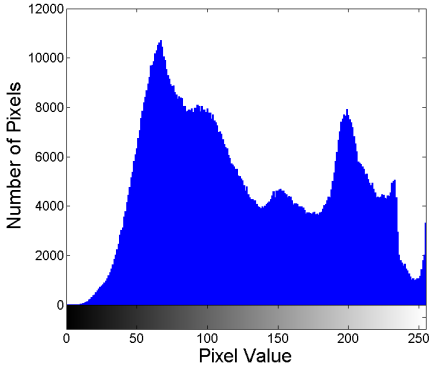
Contrast Enhancement Forensics

- Contrast enhancement
  - Used to alter lighting conditions
  - Nonlinear pixel value mapping

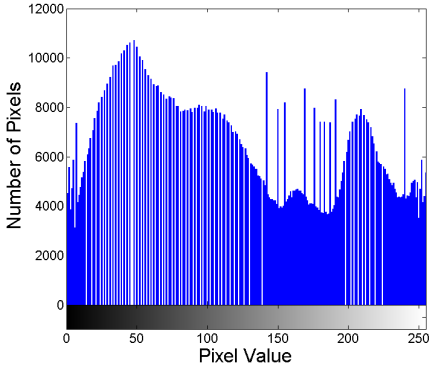


- Contrast enhancement mappings *leave fingerprints in histograms*

Unaltered



Contrast Enhanced



M. C. Stamm and K. J. R. Liu, "Forensic Detection of Image Manipulation Using Statistical Intrinsic Fingerprints", *IEEE Trans. on Information Forensics and Security*, vol. 5, no. 3, pp. 492 - 506, Sep. 2010.

Information Anti-Forensics

- Very little consideration has been given to *anti-forensic operations*
  - Designed to remove fingerprints
  - Create undetectable forgeries
- *The study of anti-forensics is critical*
  - Identifies weaknesses in existing forensic techniques
  - Develop tools to detect the use of anti-forensics

Today's Talk

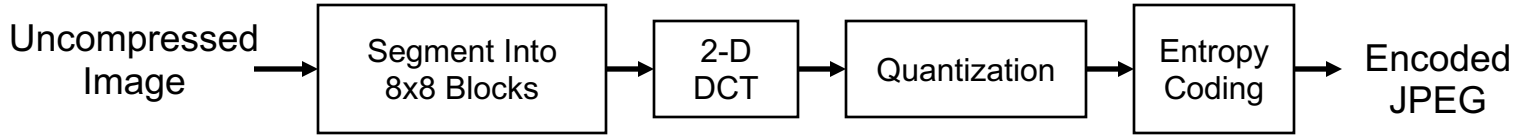
Goal: *Examine forensics from attacker's point of view*

- Image Compression Anti-Forensics
- Video Frame Deletion Forensics and Anti-Forensics
- Game Theoretic Evaluation of Interplay Between Forger and Investigator
- Reverse Engineering Prevention Using Anti-Forensics
- Summary
- Future Work

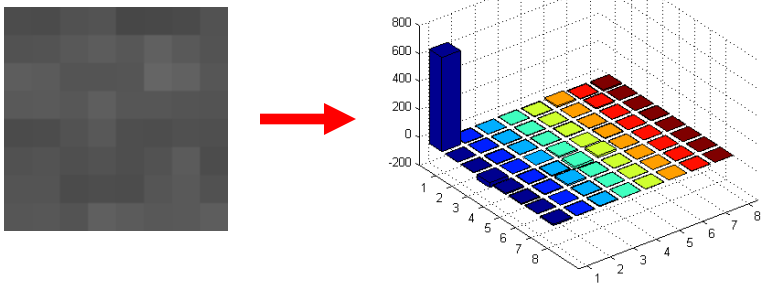
Image Compression Anti-Forensics

- Most digital images are compressed
- Image compression *leaves behind fingerprints*
- Widely used to perform many forensic tasks
  - Trace processing history
  - Detect cut-and-paste forgeries
  - Identify an image's origin
- Intelligent forger wishes to *remove these fingerprints*

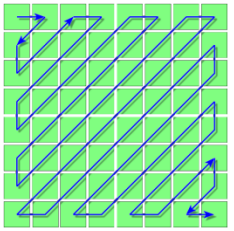
JPEG Compression Overview



- Image is segmented into blocks
- Discrete cosine transform (DCT) performed on each block
- DCT coefficients are quantized
- Quantized coefficients are reordered into one dimensional sequence
- Bitstream is losslessly encoded



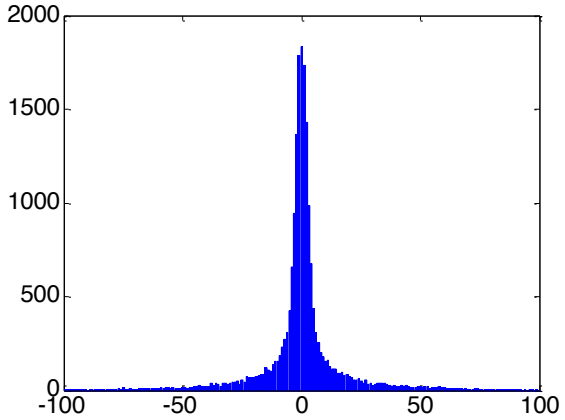
16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99



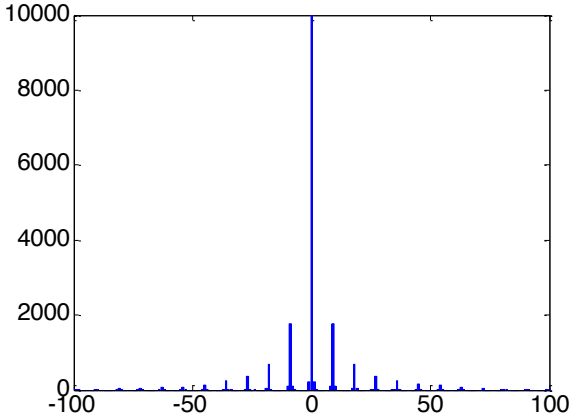
JPEG Compression Fingerprints

- Decompression performed by inverting each step
  - Quantization is not invertible
  - Dequantization $Y = Q_{i,j}\text{round}(\frac{X}{Q_{i,j}})$
- Dequantization results in compression fingerprints
  - DCT coefficient quantization fingerprints
  - Blocking artifacts

Uncompressed



JPEG Compressed



Anti-Forensic Attack

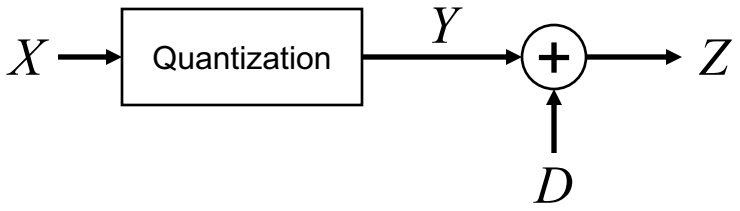
- Developed anti-forensic techniques to *erase image compression fingerprints*

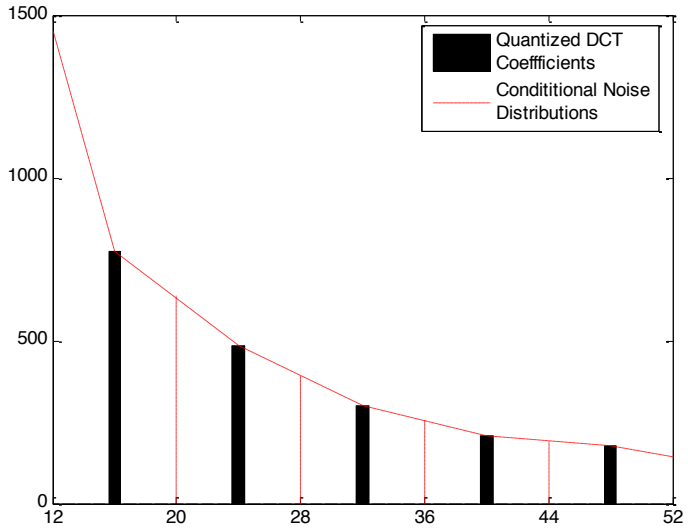
- Can *fool existing forensic techniques* that use compression fingerprints

M. C. Stamm and K. J. R. Liu, "Anti-Forensics of Digital Image Compression", *IEEE Trans. on Information Forensics and Security*, vol. 6, no. 3, pp. 1050 - 1065, Sep. 2011.
[#1 Most accessed article of IEEE TIFS from August 2011 through January 2012]

Anti-Forensic Attack

- Parametrically model DCT coefficient distribution
- Estimate unquantized coefficient distribution from quantized coefficients
- Remove DCT quantization artifacts by adding *anti-forensic dither*
- Dither distribution dependant on
  - Estimated unquantized DCT coefficient distribution
  - *DCT coefficient value to which dither is added*



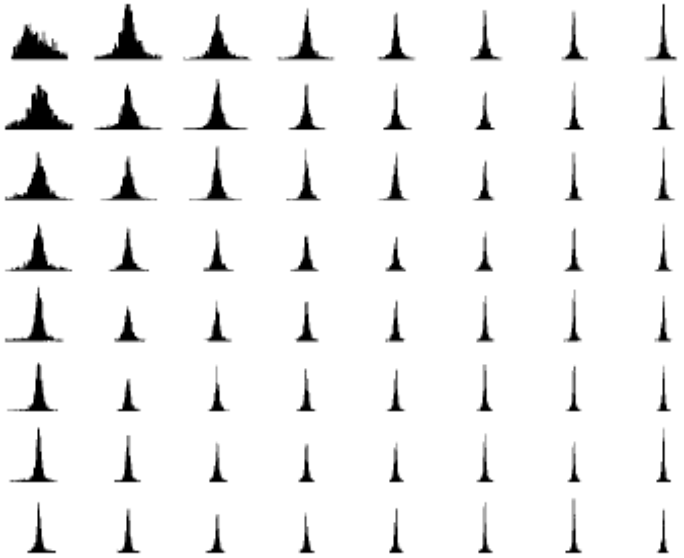


DCT Coefficient Distribution

- Unquantized DCT coefficient distribution model

$$P(X = x) = \frac{\lambda}{2} e^{-|x|}$$

Laplace Distribution



- Quantized DCT coefficient distribution model

$$P(Y = y) = \begin{cases} 1 - e^{-\lambda Q_{i,j}/2} & \text{if } y = 0 \\ e^{-\lambda |y|} \sinh\left(\frac{\lambda Q_{i,j}}{2}\right) & \text{if } y = kQ_{i,j} \\ 0 & \text{otherwise} \end{cases}$$

Discrete Laplace Distribution

- Maximum likelihood estimate of λ on the basis of y

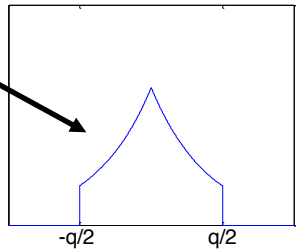
Anti-Forensic Dither

- Anti-forensically modified coefficient Z given by $Z=Y+D$
- Use conditional anti-forensic dither distributions
  - For zero valued DCT coefficients

$$P(D=d \mid Y=0)=\begin{cases}\frac{1}{c_0}e^{-\lambda|d|} & \text{if } \frac{-Q_{i,j}}{2}\geq d\geq\frac{Q_{i,j}}{2}\\ 0 & \text{otherwise}\end{cases}$$

where $c_0=1-e^{-\lambda_{ML}Q_{i,j}/2}$

Truncated Laplace Distribution

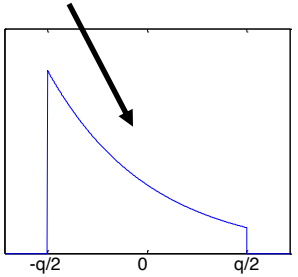


  - For nonzero valued DCT coefficients

$$P(D=d \mid Y=y)=\begin{cases}\frac{1}{c_1}e^{-\operatorname{sgn}(y)\lambda(d+Q_{i,j}/2)} & \text{if } \frac{-Q_{i,j}}{2}\geq d\geq\frac{Q_{i,j}}{2}\\ 0 & \text{otherwise}\end{cases}$$

where $c_1=\frac{1}{\lambda_{ML}}\left(1-e^{-\lambda_{ML}Q_{i,j}}\right)$

Truncated Exponential Distribution



Anti-Forensic Dither

- Can prove adding anti-forensic dither ensures that DCT distribution *matches uncompressed distribution!*
  - Assume estimated λ is true value
  - Use Law of Total Probability

Expression for $P(Z=z|Y\neq 0)P(Y\neq 0)$

$$P(Z=z)=\sum_{y} P(Z=z\,|\,Y=y)P(Y=y)$$

$$=\sum_{y\neq 0}\frac{1}{c_1}e^{-\mathrm{sgn}(y)\lambda(z-y+Q_{i,j}/2)}e^{-\lambda|y|}\sinh\left(\frac{\lambda Q_{i,j}}{2}\right)\mathbb{1}\left(\left(y-\frac{Q_{i,j}}{2}\right)\geq z\geq\left(y+\frac{Q_{i,j}}{2}\right)\right)$$

$$+\frac{1}{c_0}e^{-\lambda|n|}\left(1-e^{-\lambda Q_{i,j}/2}\right)\mathbb{1}\left(\frac{Q_{i,j}}{2}\geq z\geq\frac{Q_{i,j}}{2}\right)$$

Expression for $P(Z=z|Y=0)P(Y=0)$

$$=\frac{1}{\lambda}e^{-\lambda|z|}$$

Matches Laplace Distribution

$$\Rightarrow P(Z=z)=P(X=z)$$

Anti-Forensic Dither

- Bound on absolute difference between unquantized and anti-forensically modified coefficients
  - $|X - Z| \leq Q_{i,j}$
  - *Effectively limits visual distortion*
  - Maximal difference dictated by compression strength
- Efficient implementation
  - Only one parameter estimate per DCT subband
  - Need only two anti-forensic dither distributions
- Easily adaptable to other transform coders
  - Developed generalized framework
  - Implemented on DWT-based coders (SPIHT, JPEG2000, etc.)

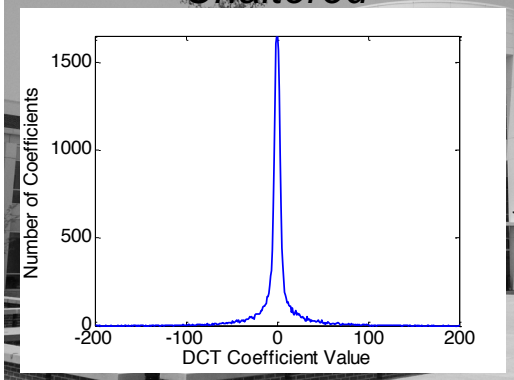
Anti-Forensic Performance

JPEG Compressed

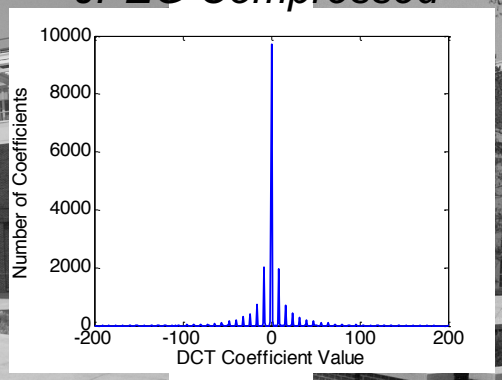
Anti-Forensically Modified

- *Tampering cannot be visually detected*
- No statistical traces of compression

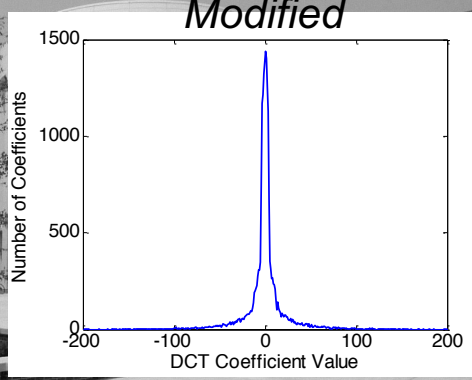
Unaltered



JPEG Compressed



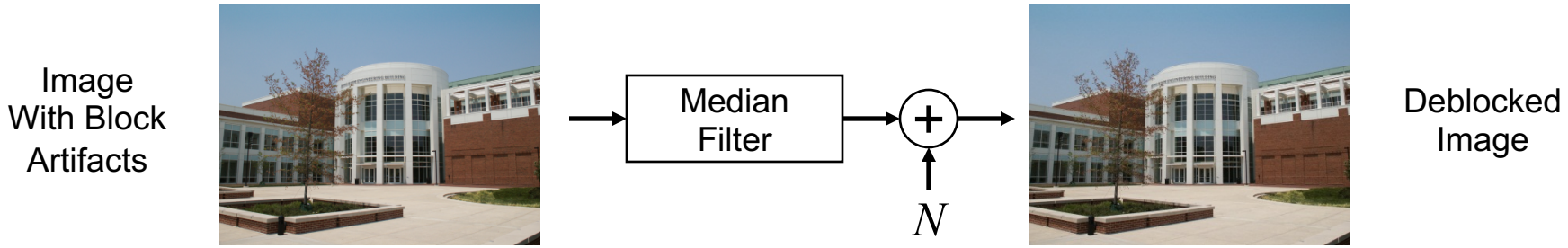
Anti-Forensically Modified



- Large-scale test: compressed then anti-foreniscally modifed 1338 images
  - *No evidence of compression detected*

Anti-Forensic JPEG Deblocking

- Existing deblocking algorithms are poorly suited for anti-forensics
- Proposed anti-forensic deblocking technique



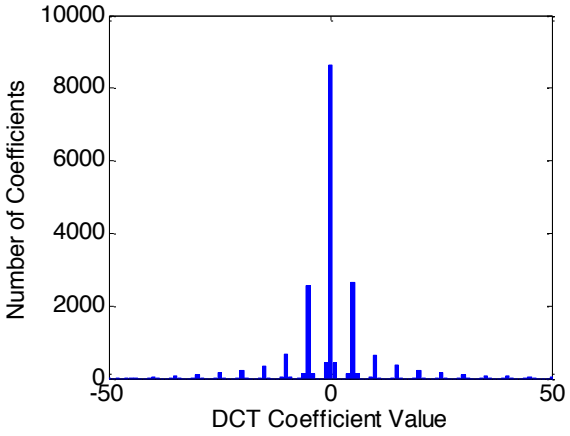
- Well suited for anti-forensic purposes
  - Removes blocking fingerprints
  - Does not alter DCT coefficient distribution

Quality Factor	Proposed Method			Zhai et al. (2008)	Liew & Yan (2004)
	s = 3 $\sigma^2 = 3$	s = 3 $\sigma^2 = 2$	s = 2 $\sigma^2 = 2$		
90	***0.0%***	***0.0%***	***0.0%***	70.1%	99.6%
70	***0.0%***	***0.0%***	14.8%	99.2%	99.6%
50	***0.0%***	0.9%	62.7%	98.8%	99.6%
30	3.3%	23.0%	93.4%	99.6%	98.8%
10	97.9%	97.9%	100.0%	100.0%	82.8%

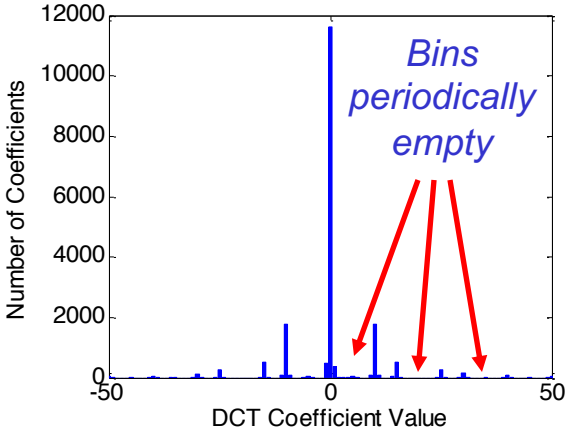
Undetectable Image Tampering

- Double compression can indicate an image was edited

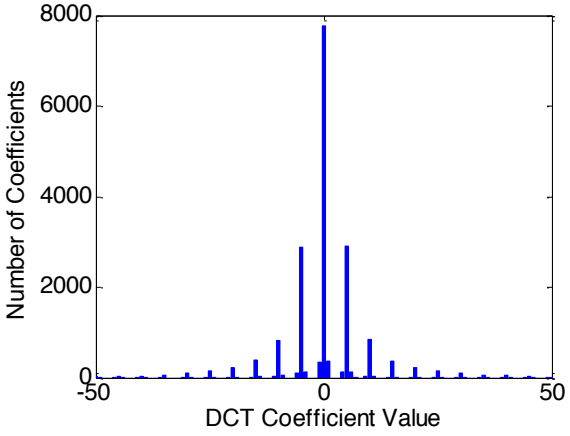
Single Compressed



Double Compressed



Double Compressed with Anti-Forensic Modification



- Anti-forensics can *hide double compression*

Undetectable Image Tampering

- Image origin linked to quantization table
- Anti-forensics can be used to *falsify image origin*

Falsified Origin	True Image Origin				
	Cam 1	Cam 2	Cam 3	Cam 4	Cam 5
Cam 1	-	100%	100%	100%	100%
Cam 2	100%	-	99%	100%	100%
Cam 3	100%	100%	-	100%	100%
Cam 4	100%	100%	100%	-	100%
Cam 5	100%	100%	100%	100%	-

- Anti-forensics can *hide localized compression history mismatches* that are evidence of *image forgery*

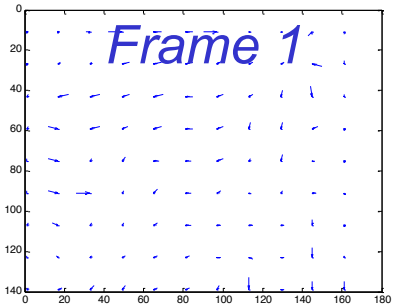
Digital Video Forensics

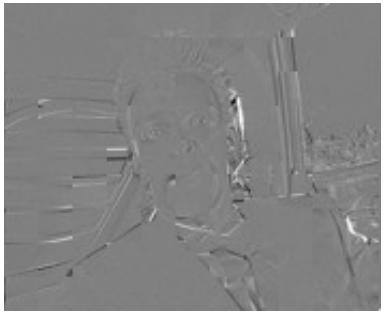
- Common video manipulation: *delete sequence of frames*
- *Frame deletion leaves behind fingerprints* [Wang & Farid 2006]
  - Detection via human inspection
- Developed a set of new video forensic and anti-forensic techniques

M. C. Stamm, W. S. Lin, and K. J. R. Liu, "Temporal Forensics and Anti-Forensics in Digital Videos", *IEEE Trans. on Information Forensics and Security*, vol. 7, no. 4, pp. 1315 - 1329, Aug. 2012.

Video Compression Overview

- Adjacent video frames are similar
- Frames are predicted from other frames

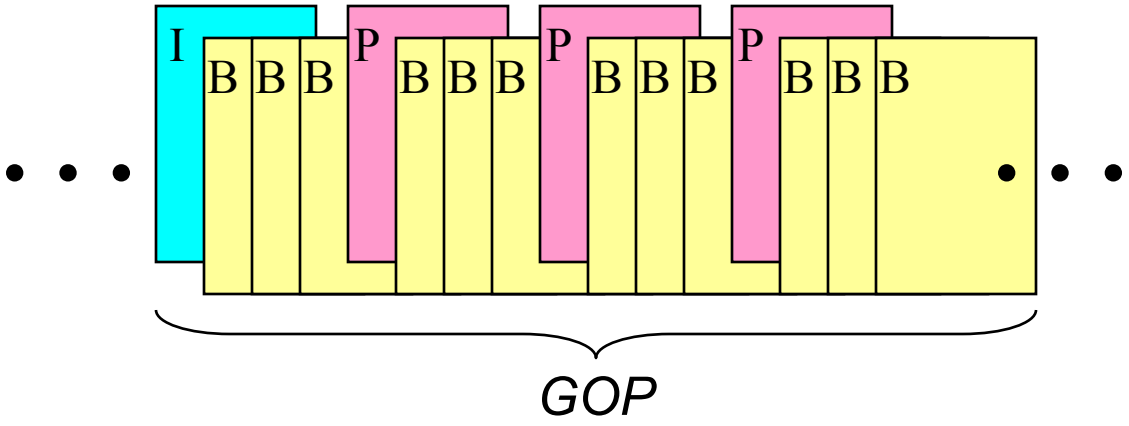




- Encoder stores motion vectors and prediction errors

Group of Pictures

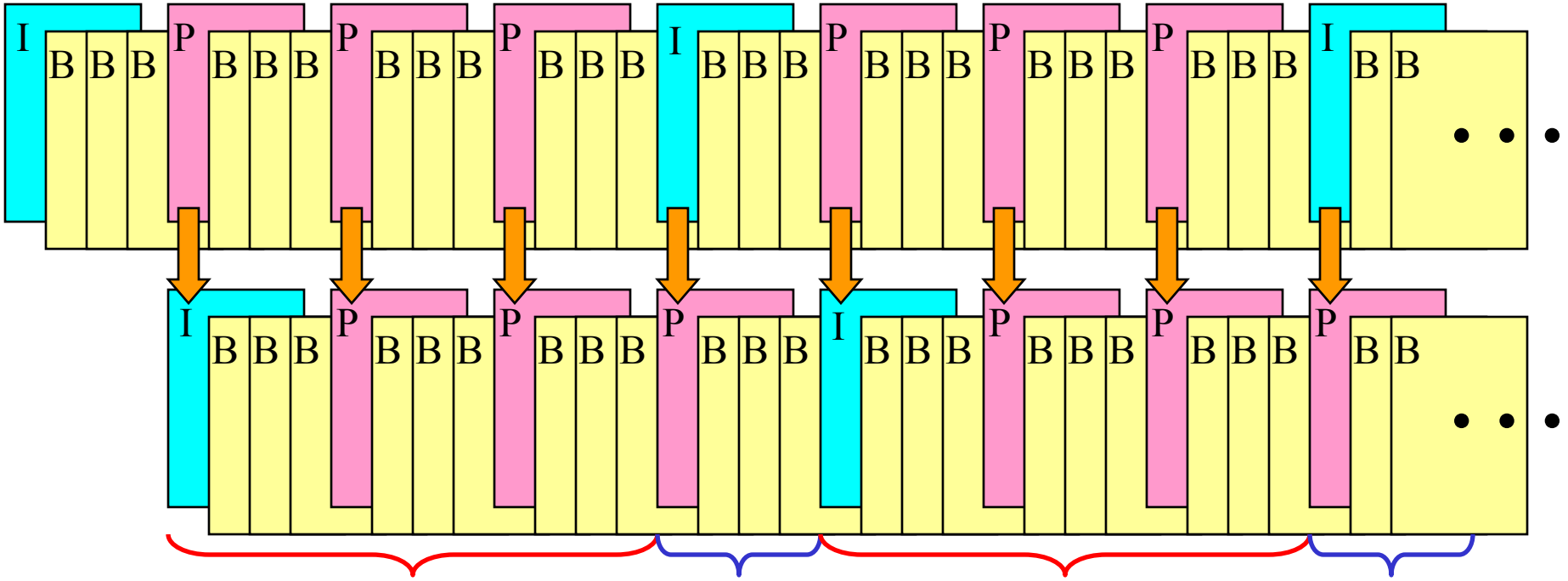
- Frames separated into several *Groups of Pictures* (GOP)
  - Prediction occurs only within each GOP



- Each GOP contains three basic frame types
  - Intra-frames (I-frames)
  - Predicted-frames (P-frames)
  - Bidirectional-frames (B-frames)

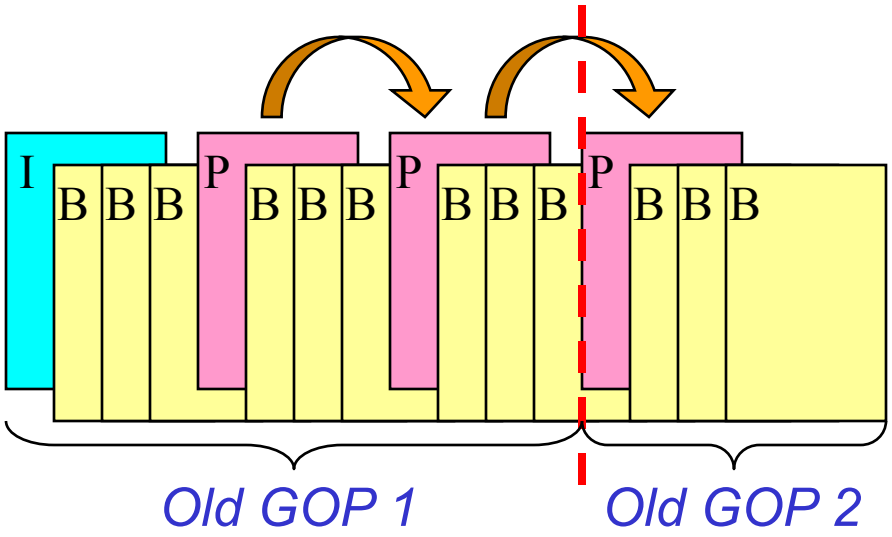
Frame Addition/Deletion Fingerprints

- Examine effect of deleting first 4 frames



- Frame deletion causes *shift in video sequence*
- New GOPs during recompression
- *Each new GOP contains frames from multiple different old GOPs*

Frame Addition/Deletion Fingerprints



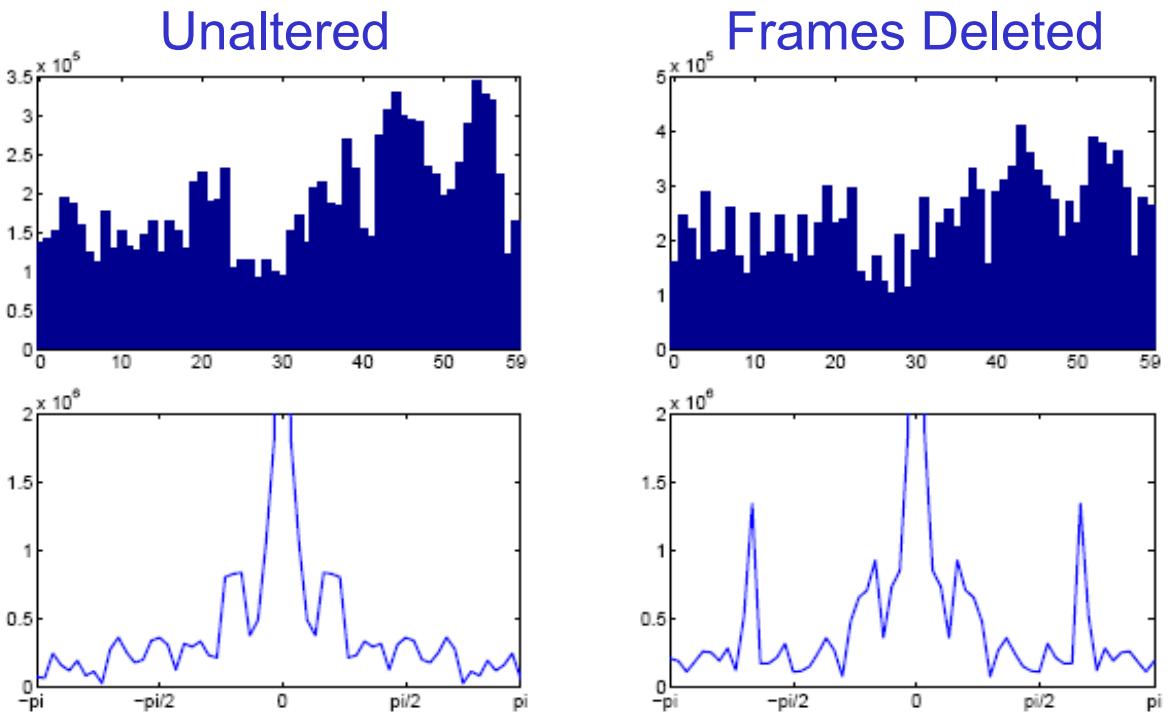
- Frame deletion affects P-frame prediction error
  - Frames predicted from an anchor within same old GOP have less prediction error
  - Frames predicted across old GOPs have greater prediction error
- Frame deletion & addition introduces *temporally distributed fingerprint* [Wang & Farid 2006]

Frame Addition/Deletion Fingerprints

- Define P-frame error signal as $e(n) = \sum_x \sum_y | p_{x,y}(n) |$

  P-frame index — *prediction error*

- Temporal fingerprint corresponds to pattern of increased $e(n)$ values



Automatic Detection

- Desire automatic frame deletion detection algorithm
  - Visual inspection prone to human error
  - Current technique only works with fixed GOP sequences

- Model $e(n)$ for videos with frame deletion

$$e_2(n) = e_1(n\text{-}n_D)(1+ s(n))$$

Video error sequence after frame deletion

Original video error sequence

Number of frames deleted

Frame deletion fingerprint

- Formulate as hypothesis testing scenario

$$\begin{cases} H_0 : e(n) = e_1(n) \\ H_1 : e(n) = e_2(n) = e_1(n) + e_1(n)s(n) \end{cases}$$

Automatic Detection

- Obtain estimate of $e_1(n)$ from $e(n)$

$$\hat{e}(n) = \text{med}\{e(n\text{-}1),\, e(n),\, e(n\text{+}1)\} \qquad \Rightarrow \hat{e}(n) = e(n) - \varepsilon(n)$$

estimation error

- Estimate the temporal fingerprint

$$\hat{s}(n) = \max(e(n) - \hat{e}(n), 0)$$

know that $s(n) \geq 0$

- Now detection can be reframed as

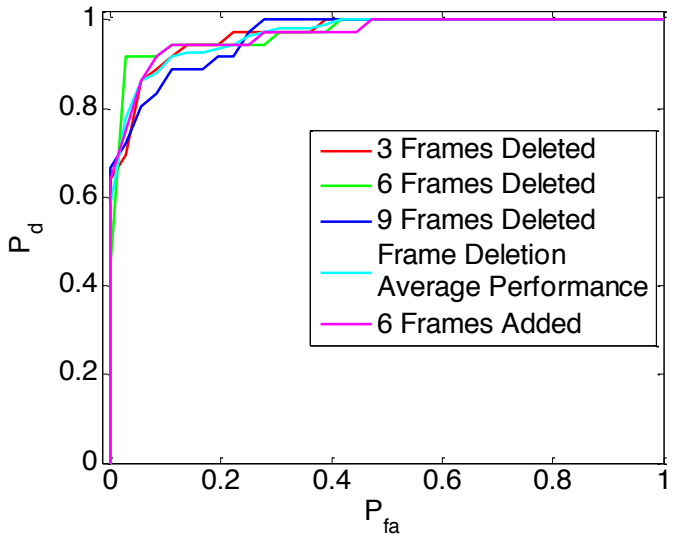
$$\begin{cases} H_0 : \hat{s}(n) = \max\left(\varepsilon(n), 0\right) \\ H_1 : \hat{s}(n) = \max\left(e(n)s(n) + \varepsilon(n), 0\right) \end{cases}$$

  - Fixed GOP structure ⇒ periodicity based detector
  - Variable GOP structure ⇒ energy based detector

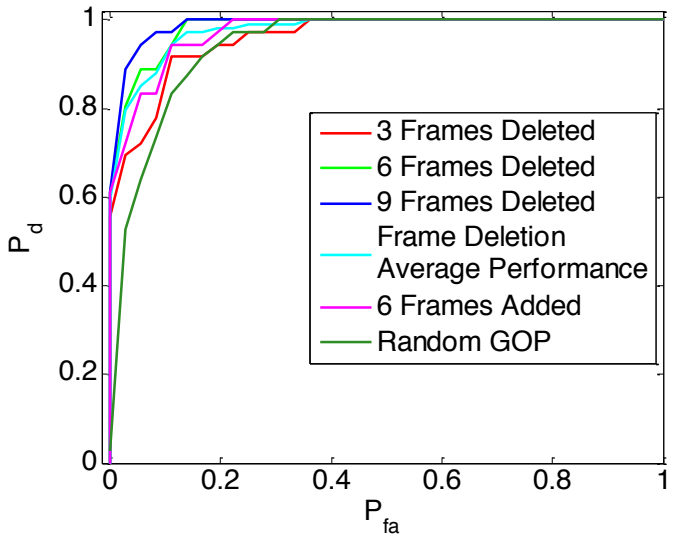
Simulations & Results

- Video compression/decompression simulated in Matlab
  - Fixed GOP structure *I B B P B B P B B P B B*
  - Randomized GOP structure
- Used 36 standard video QCIF sequences
  - Deleted frames from the beginning of each sequence

Periodicity Detector



Energy Detector



Frame Addition/Deletion Anti-Forensics

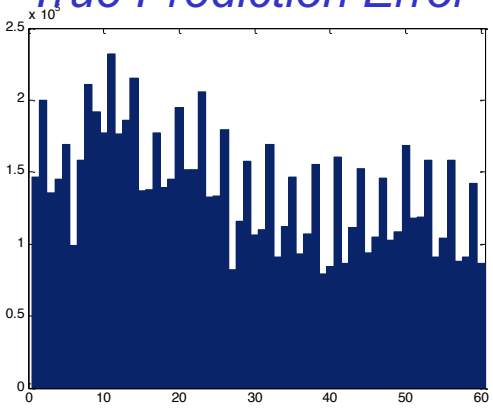
Anti-forensic problem:

- What can a forger do to hide frame deletion?
  - Must remove forensically detectable fingerprint
  - Can not degrade video quality
  - Must not alter decoder
- Each frame's prediction error can be increased
  - Use suboptimal motion vectors
  - Recalculate prediction error using these motion vectors
- Meets anti-forensic criteria
  - Modified video still decodable by standard decoder
  - Does not decrease quality of reconstructed video

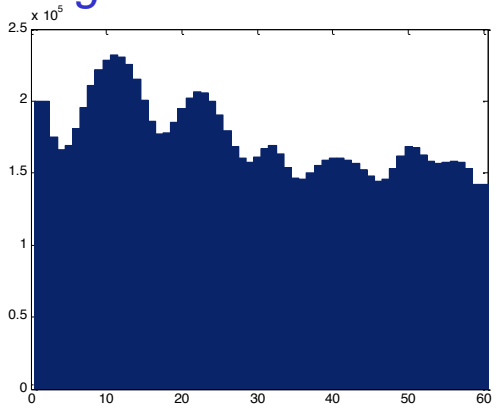
Anti-Forensic Technique

- Use model to create target error sequence without fingerprint
- Increase P-frame prediction error to match target value

True Prediction Error



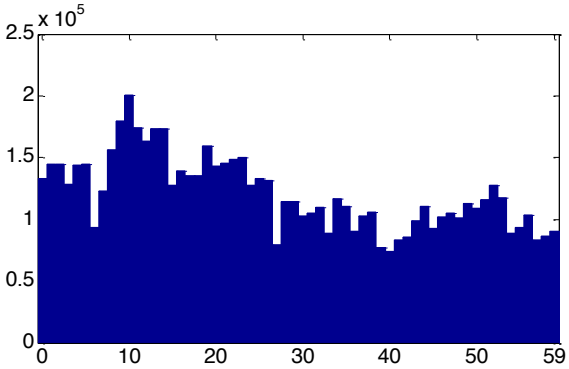
Target Prediction Error

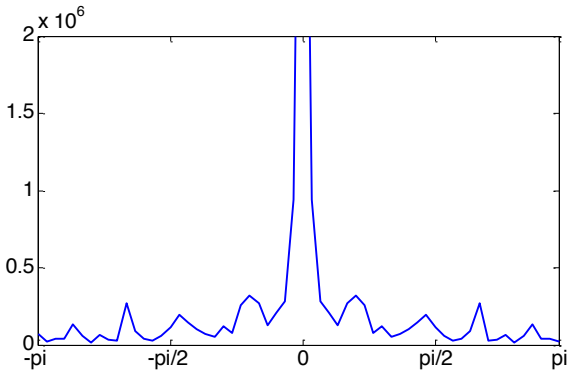


- Anti-forensic motion vector search procedure
  1. Find optimal set of motion vectors
  2. Fix anti-forensic motion vector search radius
  3. Find motion vectors within radius that maximize $e(n)$
  4. Check to see if current prediction error exceeds target error
     - If no, increase search radius and return to 2

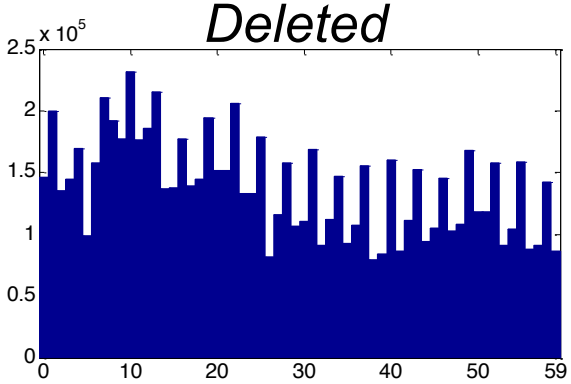
Anti-Forensic Results

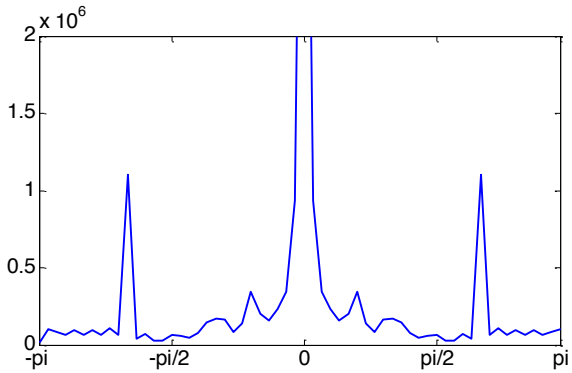
Unaltered



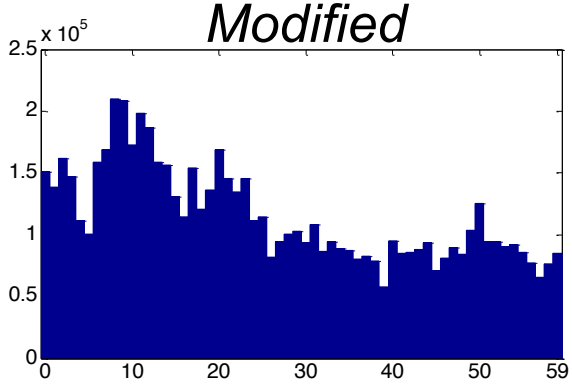


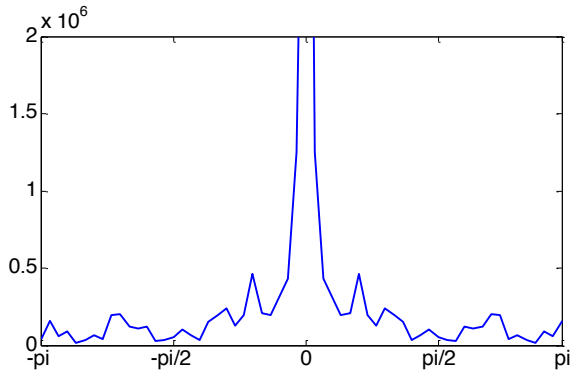
Frames Deleted





Anti-Forensically Modified





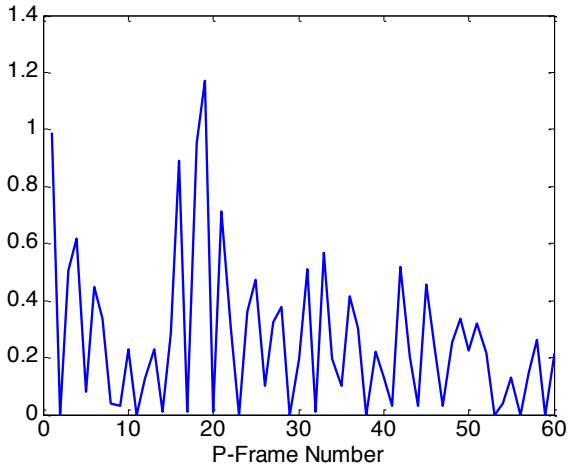
- In test on multiple videos, performance of forensic techniques *reduced to random decision*

Fingerprints From Anti-Forensic

- What can investigator do to combat anti-forensics?
- *Anti-forensics can leave its own fingerprints*
- Anti-forensics modifies motion vectors *but not true motion*
- Use motion to detect frame deletion anti-forensics

Anti-Forensic Processing Detection

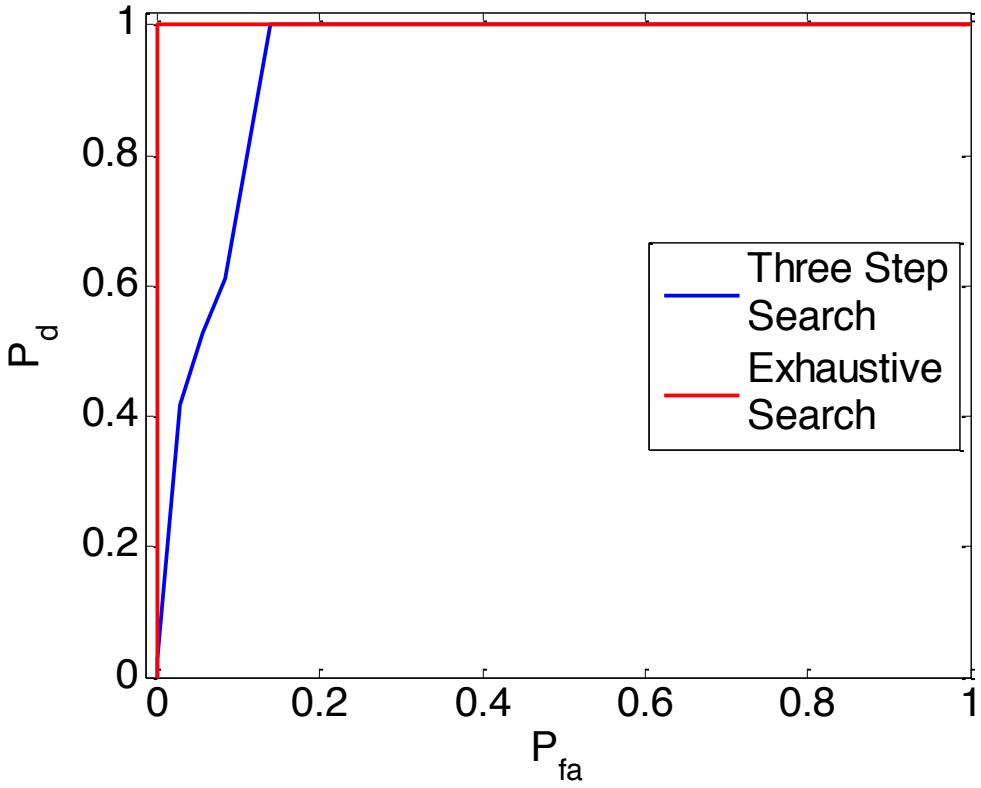
- Calculate mean Euclidean distance $d(n)$ between estimated and reported motion vectors



- $d(n)$ has periodic component for anti-forensically modified videos
- Create feature vector for detection
  - Mean value of $d(n)$
  - Measure of periodicity of $d(n)$

Anti-Forensic Detector Performance

- Frame deletion anti-forensics *can be reliably detected*



Forensic Manipulation Detection

Forensic Problem: Has a digital multimedia file ψ been manipulated using the operation $m(\cdot)$?

- Traditionally posed as hypothesis testing scenario

$$H_0 : \psi \neq m(\psi')$$
$$H_1 : \psi = m(\psi')$$

- Manipulation identified using detection rule δ_m
  - Measures strength of fingerprint m by calculating detection statistic
  - Compares detection statistic to a decision threshold

M. C. Stamm, W. S. Lin, and K. J. R. Liu, "Temporal Forensics and Anti-Forensics in Digital Videos", *IEEE Trans. on Information Forensics and Security*, vol. 7, no. 4, pp. 1315 - 1329, Aug. 2012.

Anti-Forensic Problem

Anti-Forensic Attack

- Forger uses anti-forensic technique α_m to fool δ_m

$$i.e.\ \ \delta_m(\alpha_m(\psi)) = H_0 \ \text{ if } \psi = m(\psi')$$

- What if α_m leaves behind its own fingerprints?
  - Detector δ_α can be designed to detect the use of anti-forensics

- This poses problem for forger:

  Should anti-forensics be used if it can also be detected?

Anti-Forensic Response

- Assume anti-forensic technique can be applied with strength $k \in [0,1]$
- Forger faced with trade-off
  - Wishes to minimize probability of forgery detection
  - Decrease in k leads to decrease P_d for δ_α but increase in P_d for δ_m
  - What is optimal k?
- Choice of k heavily depends on decision thresholds used by δ_m and δ_α

Forensic vs. Anti-Forensic Trade-Off

- Forensic investigator is not free to use any set of decision thresholds
  - Must meet false alarm constraint
  - *Both detectors contribute to false alarms*
- Allocation of false alarm levels depends on k, the strength of forger's the anti-forensic technique
- Both forger and forensic investigators optimal actions are *dependent on the actions of the other*
- Use game theory to find optimal set of strategies

Investigator vs. Forger Game

- Game theory is the study of strategic decision making
- *Nash Equilibrium* – set of strategies that no player has incentive to deviate from
- Players
  - Player 1 – Forensic Investigator
  - Player 2 – Forger
- Strategy sets
  - Player 1 – $\eta \in \left[0, P_{fa}^{Max}\right]$
    - η is P_{fa} level allocated to manipulation detector
    - Remaining false alarm level allocated to anti-forensics detector
  - Player 2 – $k \in [0,1]$

Utility Functions

- Utility of the forensic investigator is the probability of detecting a forgery

$$U_1(k,\eta) = P\left(\text{manipulation detected} \quad \text{anti-forensics detected}\right)$$

- Utility of forger includes
  - Probability of forgery being detected
  - Cost of distortion introduced into file by anti-forensics

$$U_2(k,\eta) = -U_1(k,\eta) - \gamma(\text{distortion from anti-forensics})$$

- Use these to find the Nash equilibrium strategies of the forger and forensic investigator

Performance at Nash Equilibrium

- Forensic investigator's ability to detect forgeries can be described using a *Nash equilibrium ROC curve*
- For each P_{fa}^{Tot}
  - Find the Nash equilibrium strategies (k^*, η^*)
  - Evaluate the probability of detecting a forgery = $U_1(k^*, \eta^*)$
- Ill-posed nature of forensics makes analytical evaluation of utilities difficult
  - Often expressions for relevant probabilities are not known
  - Utility functions must be evaluated numerically

Example: Video frame deletion forensics

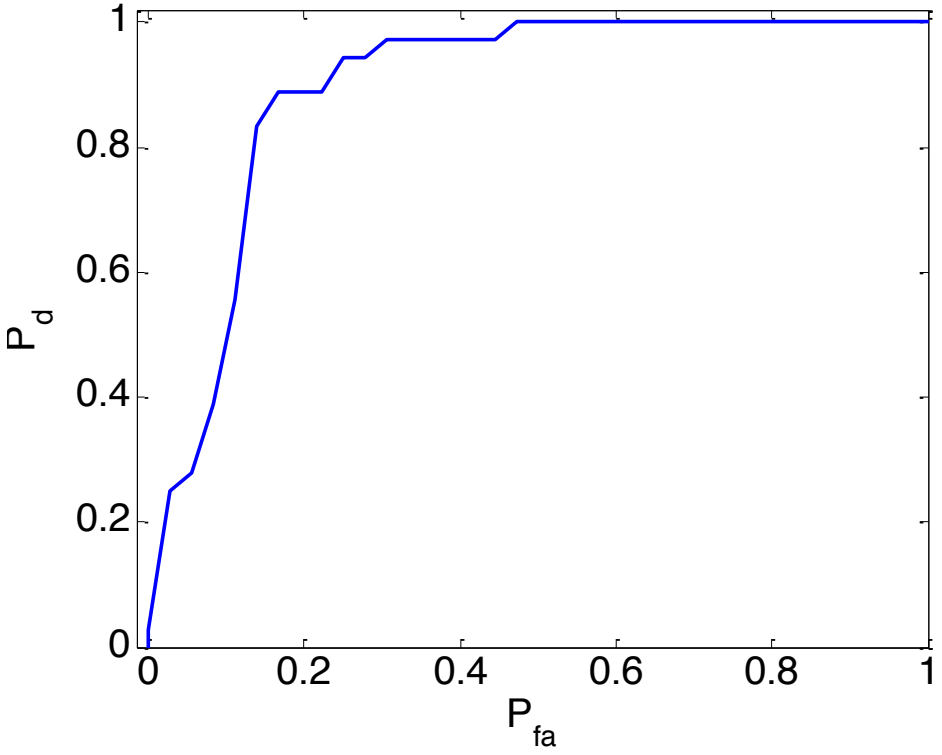
Nash Equilibrium ROC

- Apply game theoretic framework
  - $\gamma = 0 \Rightarrow U_1(k,\eta) = -U_2(k,\eta)$
  - $(k^*, \eta^*) = \arg\max_{\eta} \min_{k} U_1(k,\eta)$

Insights from NE ROC

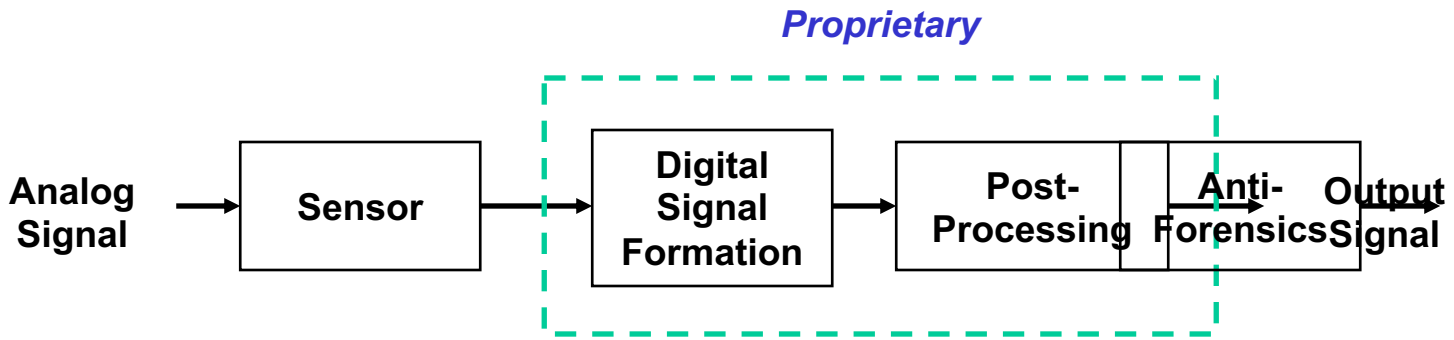
- At $P_{fa} < 10\%$
  - *P_d is worse than expected!*
  - *Forger has advantage*
- At $P_{fa} \geq 15\%$
  - Forgeries can be accurately detected
  - *Forensic investigator has advantage*

NE ROC Curve Using Additional Data



Information Protection Through Anti-Forensics

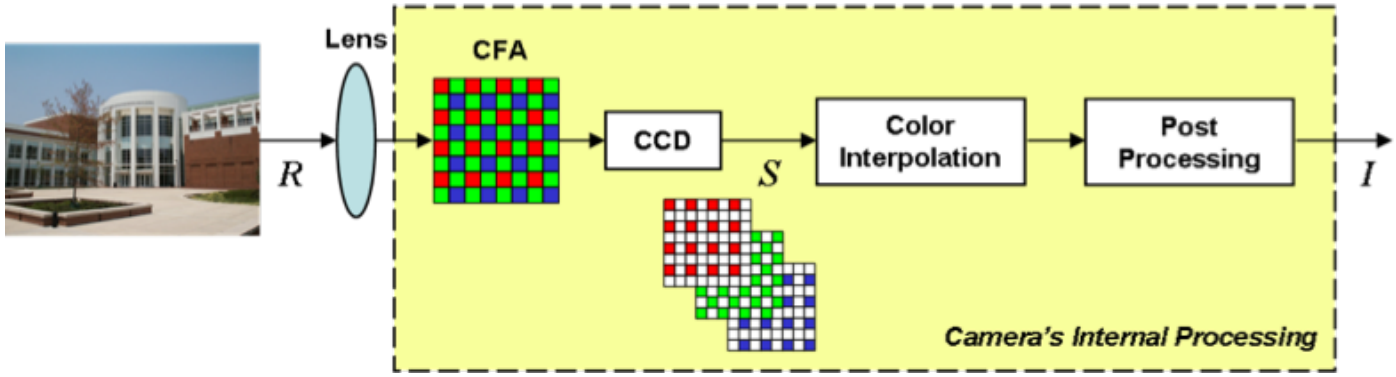
- Overlooked problem:
  - *Forensics can be used to perform reverse engineering*



- New use for anti-forensics
  - *Reverse engineering prevention*
- Anti-forensic module may be integrated into digital devices to prevent reverse engineering

Image Capture Process

- Specific application: digital cameras
- Image processing pipeline

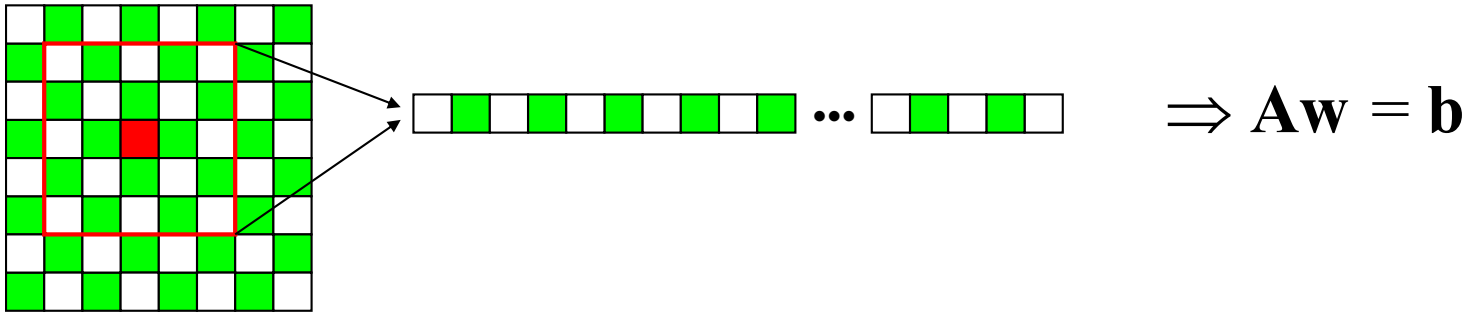


- Possible proprietary components
  - Color filter array pattern
  - Color interpolation technique
  - Image enhancing post-processing operations
- Forensically estimatable using *component forensics*

Component Forensics

- Guess color filter array pattern and resample image
- Obtain a linear estimate of the interpolation coefficients

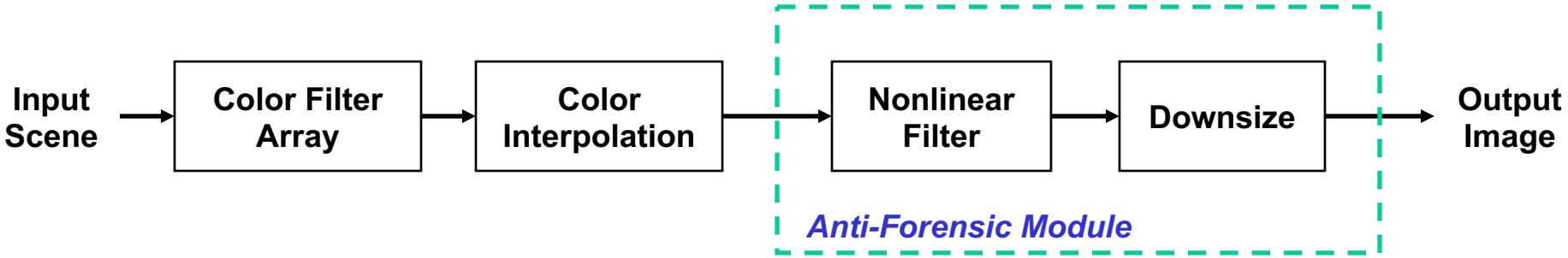
$$I(x,y,c) = \sum_{j\in\Omega}\sum_{k\in\Omega} w_{j,k} S_d(x,y,c)$$



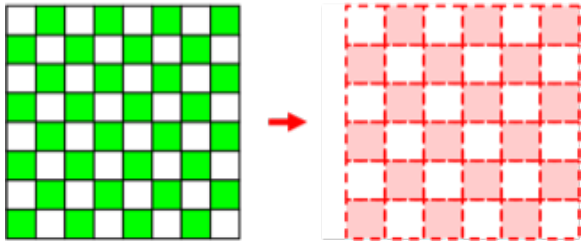
$$\Rightarrow \mathbf{Aw} = \mathbf{b}$$

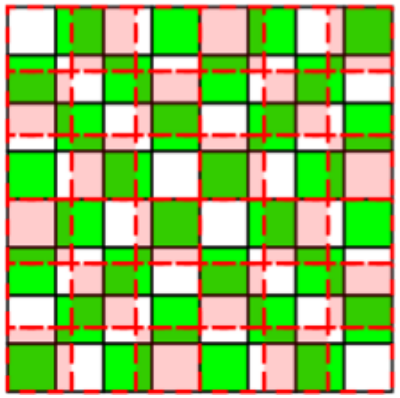
- Solve for $\mathbf{w}$ using least squares [Swaminathan et al. 2007]

Anti-Forensic Module



- Oversample then downsize
  - Disrupts CFA sampling pattern





- Nonlinear filter reduces linear dependencies between pixels
  - Example: median filter
  - Smooths image but preserves edges

Anti-Forensic Performance

- Tested performance using 5 color interpolation methods
  - Used prediction accuracy as performance criteria

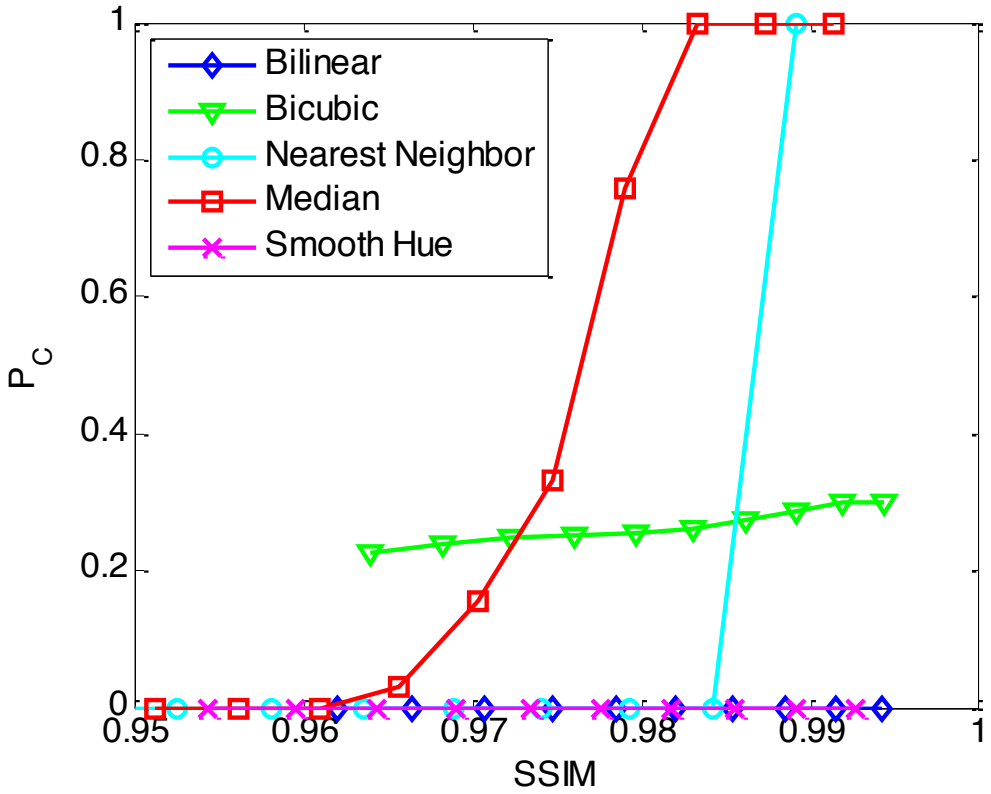
$$P_C^{(n)} = P(\mathcal{M} = M^{(n)} \mid \hat{\mathcal{M}} = M^{(n)})$$

  - Baseline forensic results: 100% accuracy

- Performance using anti-forensic module
  - *Only one method identified at rate above random decision*
  - This method identified at *2.5% above random decision*

Anti-Forensic Performance

- Can tradeoff between image quality and anti-forensic protection



Summary

- Editing operations leave fingerprints in digital multimedia content
- Forger can create anti-forensic techniques
- Anti-forensics can leave its own fingerprints
- Forger vs. forensic investigator dynamics can be analyzed using game theory
- Reverse engineering can be prevented using anti-forensics

Additional Work

- Contrast enhancement detection
- Median filtering detection
- Median filtering anti-forensics
- Order of operations forensics
- Detection of compressive sensing
- Rate-Distortion-Concealability tradeoff in anti-forensics

Future Directions

- Biometric identification systems are receiving increased attention
- Integration into everyday devices
- Potential for *tampering or falsification*

- Use forensic to *improve biometric security*

Future Directions

Short Term

- Challenge:
  - Several editing operations
  - Many forensic tests
  - *Need one forensic decision*
- Develop *forensic decision fusion techniques*
- Identify *optimal set of editing operations* for forger
- Examine adversarial dynamics using game theory

Digital Multimedia Forensics and Anti-Forensics

Matthew C. Stamm

Signals and Information Group

Department of Electrical and Computer Engineering

University of Maryland, College Park