A Picture’s Worth…
Digital Image Analysis

Dr. Neal Krawetz
Hacker Factor Solutions
www.hackerfactor.com
Contents

• Digital Image Analysis
  – Problem Space
  – Analysis Methods
• Accuracy and Limitations
• Case Study
• Conclusion

All images and screen shots are copyright by their respective owners and are included for academic discussion and research. This complies with the copyright law of the United States as defined and stipulated under Title 17 U. S. Code. The methods presented here are experimental.
Digital Image Analysis
Problem Space
Pictures Have Power

Space Shuttle Challenger

Iwo Jima, World War II

Not All Pictures Are Real

• Why not real?
  – Modified to influence opinions
  – Enhanced to convey a point
  – Designed to show techniques

• Implications
  – Legal: Child Pornography vs. Virtual Child Porn
  – Security: Image as Authentication
  – Media: Misleading Headlines
Images and the Law

- Pornography
  - Protected by the First Amendment

- Child Pornography
  - Child Pornography Prevention Act (1996)
  - Prevents use of children in sexually explicit materials
  - Does not distinguish real from fake

- Virtual Child Pornography
  - CPPA violated free speech rights
  - Distinction between “CP” and “VCP”
  - VCP does not use real children (it is regular “pornography”)
Images as Authentication

• How do you authenticate someone online?
  – Name, Address, Phone, Age
  – Collecting information from minors…
• Forging authentication
  – Yahoo!
  – Myspace
Yahoo! Impersonation

Date: Fri, 20 Apr 2007 12:38:13 -0700
Subject: Re: Abuse - Impersonation
Reply-To: Yahoo! Mail <mail-abuse@cc.yahoo-inc.com>

Hello,

Thank you for contacting Yahoo! Customer Care.

If you are an individual being impersonated by a Yahoo! Mail user, we will need a signed statement from you denying any involvement with the account, as well as a copy of the email (including full Internet headers) that is going out in your name. If you do not have an actual email message, please give us a detailed explanation of why you believe you are being impersonated. We will also need a copy of your photo ID.

If your company is being impersonated by a Yahoo! Mail account, we will need a signed statement on company letterhead denying any involvement with the account, as well as a copy of the email (including full Internet headers) that is going out with the company name.

You may fax your statement to us at:

(503) 615-3883
Defeating Yahoo Identification

- Fake Photo ID
  - Download template
    http://www.linkbase.org/make-fake-id/

- Photoshop

- Fax!
Images as Authentication

419eater.com
My Problem with MySpace

http://www.peacexpeace.org/elements/images/familysignguy.jfif
Fake Photos in the Media

- **Old School**
  - Staged
  - Mislabeled
  - *Not detectable*

- **Old-tech**
  - Negative splicing
  - Airbrushing
  - *May be detectable*

- **Hi-tech**
  - Digitally spliced
  - Digitally enhanced
  - “Shopped”
  - *Is detectable!*

Copyright 2007-2008 Hacker Factor
Old-School Fakes

10-Oct-1914: “I opened up the paper and what was my surprise to see a big spread picture of myself, lined up against that row of Melle cottages and being shot for the delectation of the British public.”

Adnan Hajj:
Beirut (Reuters)
22 July 2006
5 August 2006

http://www.greatwardifferent.com/Great_War/Belgium/Belgium_War_Reporters_01.htm
http://neveryetmelted.com/?cat=743
Old and New

• Problem
  – Photos are REAL
  – Only identified by close inspection or tracking source

• Combined with new methods

2002 Dust Storm 2004 Tsunami

http://www.snopes.com/photos/tsunami/sumatra.asp

Copyright 2007-2008 Hacker Factor
The Big Questions

• Distinguish “real” from computer graphics
• How to detect image manipulations
• How to pull out information from images
  – Real images: who, where, when, how
  – Digitally enhanced: what, how
  – Computer graphics: what, how
The Big Answers

- Observation
- Basic Image Enhancements
  - Color Tweaking
- Image Format Analysis
  - Meta Data Analysis
  - Quantization Table Fingerprinting
  - Estimated Compression Level
- Advanced Image Analysis
  - Error Level Analysis (ELA)
  - Principle Component Analysis (PCA)
  - Wavelet Transformations
  - Luminance Gradient (LG)
Observation
Warez Factory
Things to Look For

• Time
  – Clocks, calendars
  – Dated materials

• Location
  – Language
  – Region-specific technology
  – Currency and Electrical Outlets!

• Other
  – What’s on the computer screen?
  – Any other identifiable elements
Example: Buzz

- Andrea Bertaccini
  - www.tredistudio.com
  - “CG Choice Award” from CG Society, 2006

- Says based on NASA photo
  http://www.hq.nasa.gov/office/pao/History/ap11ann/kippsphotos/5903.jpg
Example: Buzz Compare
IMAX: Magnificent Desolation

• IMAX recreated moonwalk
  – http://www.imax.com/magnificentdesolation
  – Director: Tom Hanks

• Timeframe
  – Movie in 2005
  – Artist image in 2006
IMAX: Magnificent Desolation
What Happened?

• Artist likely:
  – Modeled position after NASA
  – Modeled spacesuit after IMAX
Format Analysis
Image Format Analysis

• Formats *are* information
  – Formats are data that contain data
  – Changes to image yield changes to format
• JPEG as an example
  – Most methods work with any image format
JPEG Feature Set

• Key Features of JPEG
  – Meta data
  – Quantization matrix for lossy compression
  – Lossy data format
  – Divide image into 8x8 cells
    • JPEG artifacts are usually visible 8x8 cells

• Feature Detection
  – Feature leads to manipulation detection
JPEG Meta Data

- **Information about image**
  - Camera type and settings
  - Date and time
- **Multiple images**
  - Varying quality
  - Useful for distinguishing cameras
- **Meta data problem:**
  - Modified or inaccurate
  - Applications do not update meta data!
  - Photoshop keeps camera info (even if picture changes)
  - Photoshop does not log Photoshop changes

```
$ exiftool IM001022.JPG
MIME Type : image/jpeg
JFIF Version : 1.1
Make : Hewlett-Packard
Camera Model Name : HP PhotoSmart 618
Orientation : Horizontal (normal)
X Resolution : 72
Y Resolution : 72
Resolution Unit : inches
YCbCr Positioning : Centered
Exposure Time : 1/125
F Number : 3.7
ISO : 100
Exif Version : 0210
Date/Time Original : 2007:05:28 09:19:49
Components Configuration : YCbCr
Compressed Bits Per Pixel : 1.6
Shutter Speed Value : 1/128
Aperture Value : 4.0
Exposure Compensation : 0
Max Aperture Value : 4.0
Subject Distance : 0.13 m
...
```
Quantization Fingerprinting

• Should compute optimal quantization tables
  – CPU intensive!
  – Slow user experience!

• Hard-coded quantization tables
  – Few systems actually generate Q tables
  – Digital cameras use different Q tables
    • Vary by make and model
    • Optimized for CCD, data size, manufacturer
    • Canon pictures look best on Canon printers (colors optimized)
  – Cannot just “copy over” Q tables

• Forensics
  – Match Q tables to application or camera
    • Media outlets: Pay attention!
Quantization Quality

• What if Q tables not known?
• JPEG uses a quality value
  – Save at 95%, 80%, 65%…
  – Quality corresponds with size
• Quality not saved in JPEG!
  – Fingerprint Q table? Know tool and quality
  – Unknown Q table? Need to determine quality
• Derive quality value!
Quantization Tables

- Q tables: compression and quality
- Two tables for YCrCb
  - 1 for luminance (Y)
  - 1 for both Cr and Cb
  - Optional:
    • 3 tables: Y, Cr, and Cb
- 64 elements
  - 1st element = DC
  - 63 elements = AC
    • Compression by frequency

<table>
<thead>
<tr>
<th>Table index=0 (luminance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 2 2 3 2 2 3 3</td>
</tr>
<tr>
<td>3 3 4 3 3 4 5 8</td>
</tr>
<tr>
<td>5 5 4 5 10 7 7</td>
</tr>
<tr>
<td>6 8 12 10 12 12 11 10</td>
</tr>
<tr>
<td>11 11 13 14 18 16 13 14</td>
</tr>
<tr>
<td>17 14 11 11 16 22 16 17</td>
</tr>
<tr>
<td>19 20 21 21 21 12 15 23</td>
</tr>
<tr>
<td>24 22 20 24 18 20 21 20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table index=1 (chrominance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 4 4 5 4 5 9 5</td>
</tr>
<tr>
<td>5 9 20 13 11 13 20 20</td>
</tr>
<tr>
<td>20 20 20 20 20 20 20 20</td>
</tr>
<tr>
<td>20 20 20 20 20 20 20 20</td>
</tr>
<tr>
<td>20 20 20 20 20 20 20 20</td>
</tr>
<tr>
<td>20 20 20 20 20 20 20 20</td>
</tr>
<tr>
<td>20 20 20 20 20 20 20 20</td>
</tr>
<tr>
<td>20 20 20 20 20 20 20 20</td>
</tr>
</tbody>
</table>
Example Derivation

• Average AC values
  – Table 0: 11.63
  – Table 1: 17.57

• Average Y, Cr, Cb
  \[(11.63 + 17.57 + 17.57) / 3 = 15.59\]

• Get RGB/YCrCb conversion
  \[||17.57 - 11.63|| = 5.94\]

• Combine to find quality
  \[100.0 - 15.59 + 5.94 = 90.35\%\]
  Call it 90%

See jpegquality.c
Quantifiable Problem

• Data loss is cumulative

• Resave problem:
  – Save an image at quality of 75%
  – Resave image at 90%
  – Image does not get better!
    • 90% of 75% = 67.5%
  – Quantization tables reflect 90%, not 75% or 67.5%

• How to detect image resaves?
  – Principal Component Analysis!
Principal Component Analysis

- PCA separates info
  - Computer vision
  - Data compression
- Identifies widest variance among points
  3D = 3 components
  PC1 = widest
  PC2 = next widest
  PC3 = narrowest
PCA Example
PCA Example
PC1 with Artifacts
PCA Mixing: 90% with 75%
PCA Mixing: 90% with 75%
Example: Back to the Moon
Buzz Aldrin Moon Walk

• “All the image are made in 3DS MAX and postprocessed in Combustion and Photoshop.”

• JPEG Q tables say:
  – Photoshop
  – 89% quality
Buzz Aldrin Moon Walk
Walk
Error Level Methodology

- JPEG is lossy format
- Each resave introduces more error
  - But “copy” does not
- Error rate not linear!
Error Level Analysis

• Each 8x8 cell should be at same quality level
• Changes to image change quality level for the 8x8 cell

Methodology
• Save image at 95%
  – Intentionally introduce known error rate
• Compare original and new 95% image
• Difference = error state
  – No difference = image local minima
  – Large difference = unstable 8x8 cell = original pixels!
Error Rate Example

Original

Resave #1, 75%

Resave #2, 75%
Modification Detection

Resave #1, 75%

Edited: Books, Dinosaur
The “Alf Kid”!
“Alf Kid” Error Level Analysis
Original “Alf Kid”?

Multiple resaves

Cropped
Crash Modifications
Crash Modifications
Wavelet Transformations

• Problem:
  – If quality is same, how can you find differences?
  – How to identify layers?

• Solution?
  – WAVELETS!
Wavelet Limitations

• Any signal can be approximated
• Some signals more difficult than others
  – Square waves or sharp color changes
  – Smooth, linear transitions
  – Extreme values (black or white)
• Some signals easier to approximate
  – “Natural” colors
  – Noisy images (e.g., CCDs)
Wavelet Image Analysis

• An 800x600 picture has 480,000 wavelets
  – Render only a few % to get general picture
    • Picture will appear blurry
  – Entire image should sharpen at same rate
• Image modification detection
  – Scaled images sharpen at different rates
  – Images from different focal lengths sharpen at different rates
  – Why? Images have different signal patterns
Wavelet Example
Watch her head
At 5% wavelets…
Analysis Limitations

- Small Images
  - Wavelets fail
- Scaled Images
- Low Quality
  - Image Corruption
  - GIF and limited-color images
- Wavelets and harmonics

- Mixing Media
  - From Photo to Magazine to JPEG…
- Extremely Talented Artists (rare)
Luminance Gradient

• Original plan: detect lights and direction
  – Turns out: the algorithm sucks.
  – Identifies general direction, but not specific

• The Power of LG
  – Distinguish CG from Real, manipulations
    • Vastly different light sources implies splicing
    • Edge and surface detection
      – Crisp edges, sharp curves, smooth gradients
      – Clean = computer graphics
LG Concept

• Lighting is never “even” (but coloring is)
  – Given an item with a uniform color…
  – Area closest to the light will be brightest
LG Algorithm

• Many many many variations
LG Arrows: Hands
LG Arrows: Hands
LG Arrows: Hands

Image source: DC3 Forensic Challenge, 2007
LG Arrows: Hands

Image source: DC3 Forensic Challenge, 2007
Copyright 2007-2008 Hacker Factor
LG Colorized Algorithm

• Recolor based on arrow direction

• Look for color transitions  
  – Colors make people look like death…
LG Coloring: Hands
LG Coloring: Hands
LG: More Hands…

Image source: http://i146.photobucket.com/albums/r253/pjbaker_2006/hands.jpg

Copyright 2007-2008 Hacker Factor
LG: More Hands…

Image source: http://i146.photobucket.com/albums/r253/pjbaker_2006/hands.jpg
LG: More Hands…

Image source: http://i146.photobucket.com/albums/r253/pjbaker_2006/hands.jpg

Copyright 2007-2008 Hacker Factor
Fun with Dubya

• President George W. Bush
  – January 2006
  – Visited the NSA (Fort Meade)
  – Photo from Newsweek & Washington Post

• Dshield and Nmap mailing lists
It’s a Trap!

http://images.insecure.org/nmap/images/wash-post-nsa.jpg
http://img378.imageshack.us/img378/5967/itsatrap7dz.jpg
It’s a Trap!

http://images.insecure.org/nmap/images/wash-post-nsa.jpg
http://img378.imageshack.us/img378/5967/itsatrap7dz.jpg
It’s a Trap! (ELA)
It’s a Trap! (PCA)
It’s a Trap! (LG)
Every Day with Rachael Ray

- FHM Magazine
  - October 2003
  - Later: Internet
- Adam Bates *claims* to have been looking at a cooking blog when he came across this picture. “Is this really her?”
Rachael Ray: ELA
Rachael Ray: LG

- Noise on picture
- Noise around apple
- No noise here
- Noise here
From the Photographer…

• Photographer:
  “This is Eric Cahan. I took the photos of rachael ray that appeared in FHM a few back [sic]. It’s just bad over retouching that FHM did but it’s her.”

• Photoshop: Magic wand and “Liquify”

• Other pictures not tampered as badly
Perfectly Imperfect

- Analysis Limitations
  - Image Quality
  - Algorithm Limitations
  - Humans…

- Overall Accuracy
Limitations: Quality

• Size Matters
  – Small images: Wavelets fail (use > 300x300)
  – Very small images: ELA, PCA, LG fail
• Scaled Images
• Low Quality
  – Image corruption (resaves)
  – Limited-color (e.g., GIF or monochrome)
• Image coloring
  – High contrast
  – Specular reflections and “washed out” areas
Limitations: Algorithm

- Wavelets and harmonics
- ELA and color selection
- Complex lighting and LG
- Mixing Media
  - From Photo to Magazine to JPEG…
  - Scanner, camera, video capture card
Limitations: Humans…

- **Human Interpretation**
  - Algorithms only highlight
  - Humans interpret

- **Extremely Talented Artists (rare)**
  - Most people already have the tools
  - Better tools can be purchased
  - Few people have the skill
Method Accuracy

• DoD Cyber Crime Center (DC3)
  – Blind test: “Real or CG?”
    • 51 images
    • 6 were “unknown” to the DC3

• Results
  – 86% accurate for known images
  – 0 false-positives (no “real” called “CG”)
  – 4 of 6 false-negatives were CG Society award winners
Case Study: Dr. Z

Dr. Ayman al-Zawahiri
#2 guy in Al Qaeda
He wore a black turban and white robe … he had a rifle behind his right shoulder that was leaning against a plain brown backdrop.

“… He wore a black turban and white robe … he had a rifle behind his right shoulder that was leaning against a plain brown backdrop.”
“He wore a black turban and white robe … he had a rifle behind his right shoulder that was leaning against a plain brown backdrop.”
USA Today Picture

28-Sept-2006

20-Dec-2006
USA Today Picture

IntelCenter

Copyright 2007-2008 Hacker Factor
What Else Added?

Last Things Added:
- Image Cropped
  - Observed, to 8x8 grid
- “IntelCenter”
- Subtitles & Logo
- Al-Zawahiri!
  - Outline = chroma key
- Banner text!

IntelCenter
And in the Original?
And in the Original?
About the Room…

• Background is independent of Dr. Z
• Claims it was computer generated
  – Possibly 3D Studio Max
• Can the room be recreated?
Recreating the Room?

lindsay|digital is a 3D design and visual effects studio in Pacific Grove, California run Spencer Lindsay, a 20-year veteran of digital design, game development and 3D design. He provides contract 3D modeling and Motion Graphics services.

http://www.lindsaydigital.com/blog/wordpress/?p=90
The Room!
The Room!

Some Words
Go Here
Side by Side
What About the Cannon?

- Cannon
  - Blurry, low-res
  - Style:
    - 18th Century British Naval Cannon
- Carriage
  - Stepped wood
  - No visible slats
  - Big wheel in front
- Trunnion inset in top
CG Cannon?

- http://www.turbosquid.com/FullPreview/Index.cfm/ID/253444
- Missing big tire
- 3D model: $39

- http://www.turbosquid.com/FullPreview/Index.cfm/ID/255325
- Looks much closer
- 3D model: $34.75
What About Other Videos?

27-July-2006
Zawahiri Video Speech Regarding Lebanon and Gaza
Analysis: Error Level and PC1
Analysis: PC3!
Wavelets 5%: 6 Layers!

Mohammad Atef
Very blurry

Blurry

Linear colors

Mohammad Atta
Almost crisp

Crisp

Ayman al-Zawahiri

Copyright 2007-2008 Hacker Factor
Mohammad Atta

Made in Layers

Identify any sources?
SITE Seeing

• *Saying* that there is a green screen is **not** the same as *seeing* the green screen

• SITE Institute ([www.siteinstitute.org](http://www.siteinstitute.org))
  – 22-Jan-2007: Intercepted Al Qaeda video!
  – 25-Jan-2007: Video released by Al Qaeda
Back in Black

Chroma-key screen

Screen border

they either hide the facts from you, or reveal them in embarrassment.
Lighting

they either hide the facts from you, or reveal them in embarrassment.

As-Sahab
Green Screen Fun

they either hide the facts from you, or reveal them in embarrassment.
Green Screen Fun

they either hide the facts from you
or seal them in embarrassment
Green Screen Fun

PC1

They either hide the facts from you, or reveal them in embarrassment.
Azzam al-Amriki

2-Sept-2006
Azzam al-Amriki

2-Sept-2006
Azzam al-Amriki

2-Sept-2006
Color Graph

As-Sahab Logo and Subtitles

Books
Case Study
Bin Laden’s Beard
The Big Gap

• 29-Oct-2004: “Graybeard”
  – 14 minute video
  – Only aired on Al Jazeera
  – Screen shots and low quality video available online

• 7-Sept-2007: “Blackbeard”
  – First video in nearly three years
  – Released online!
  – Lots and lots of oddities…
Graybeard Video

• 29-Oct-2004
  – Very low quality videos available online
  – Better quality only via screen shots
Blackbeard Video

• 7-Sept-2007
  – 26 minute video titled “The Solution”
  – Released online (677 Meg MPG)
7-Sept-2007 Video Timeline

- 26 minute video
  - Total: 3.5 minutes of animation
  - Current events only mentioned after audio splice and during frozen frames
Animation Oddity

• Animated segments are different!

Frame 1:56

Frame 12:47
Animation Oddity

• Animated segments are different!

Frame 1:56  
Frame 12:47
Animation Oddity: Wider

Frame 1:56

Frame 12:47
The Big Question:
“Is the Black Beard Real?”

Could it be digitally modified?
Image Analysis

2004

2007 - 2nd segment
Image Analysis

2004

2007 - 2nd segment
Image Analysis

2004

2007 - 2nd segment
What Can We Tell?

- No indication of digital manipulation
- Similar lighting
- Similar clothing
- Similar background
- Anything else?
  - Camera angle, aspect ratio, and coloring
Camera Setting

2004

2007 - 2nd segment
Camera Setting
Recolor Image

2004

2007 - 2nd segment
Analysis Summary

• No indication of digital modification
• Align on eyes
  – Same eyes, eyebrows, nose, mouth, hairline
  – Shoulder position matches
  – Hat is worn higher
  – Desk and papers align
  – SAME ASPECT RATIO
  – Overall color of 2004 likely altered during post-processing
• Implies
  – Extremely similar setting, lighting, camera setup
  – Same person but with a different beard
What About the Beard

- Option #1: All recorded in 2004
- Option #2: Recreation
Beard Option #1

• Option 1: All video recorded in 2004
  – Similar set, lighting
  – Similar camera position and aspect ratio
  – Similar clothing & hairline
• Implies:
  – Dyed or costume beard
    • Gray beard is larger with dark edges
    • Gray is likely fake
  – Multiple recordings hours or days apart
  – 2007 audio is dubbed
  – No Bin Laden since 2004
Beard Option #2

• Option 2: Recreation
  – Recreated lighting and set (including papers)
  – Recreated camera position and aspect ratio
  – Match clothing, hairline
  – And all 3 years later!

• Implies:
  – Significance to the set
    • But the set is plain…
    • Forgot beard, robe…
  – Bin Laden dyed beard
  – Beard shrank in length
Which Option is Right?

• Cannot tell from image analysis
• Practical view: Occam’s Razor
  – Simplest solution is likely correct.
  – Which is simpler?
    • Recording all video at once, and releasing over the years with audio-dubbed current events
    • Recreating the set, lighting and minutia but forgetting the big things
Conclusion
Need for Image Analysis

• Real versus Computer Generated
• If Modified, How?
• Uses
  – Media: Reality vs Fiction
  – Legal: Child Pornography vs VCP
  – Authentication: Real vs Doctored
Methods Covered

• Observation
• Basic Image Enhancements
  – Color Tweaking
• Image Format Analysis
  – Meta Data Analysis
  – Quantization Table Fingerprinting
  – Estimated Compression Level
• Advanced Image Analysis
  – Error Level Analysis
  – Principle Component Analysis
  – Wavelet Transformations
  – Luminance Gradient

Other types of analysis:
  Shadow Detection
  Mixture of Gaussians
  Minimum Variance Color Selection
  Minimum Variance Quantization
  K-Means
  Scale-Invariant Feature Transform
  Signal-to-Noise Ratio
  Color Filter Array Detection
  ... and the list goes on ...
Where to Start?

[Image of book covers]
Questions?

Shameless self-promotion.

Dr. Neal Krawetz
Hacker Factor Solutions
www.hackerfactor.com
Acknowledgements

Thanks to the following people:

- Cynthia Baron, Northeastern University
- Adam Bates, University of Colorado at Boulder
- Paul Whyman and Bob Gobeille
- Spencer Lindsay, lindsayDIGITAL
- Pedro Bueno, Internet Storm Center
- Hany Farid, Dartmouth College
- Mark Rasch, J.D.
- Department of Defense Cyber Crime Center