A Picture's Worth…
Image Analysis and Forensics

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

Contents

- Digital Image Analysis
  - The Problem with Images
  - Authenticating Images
- Analysis Methods
- Case Study: Dr. Z
- Conclusion

Disclaimer

- 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

Pictures Have Power

Space Shuttle Challenger

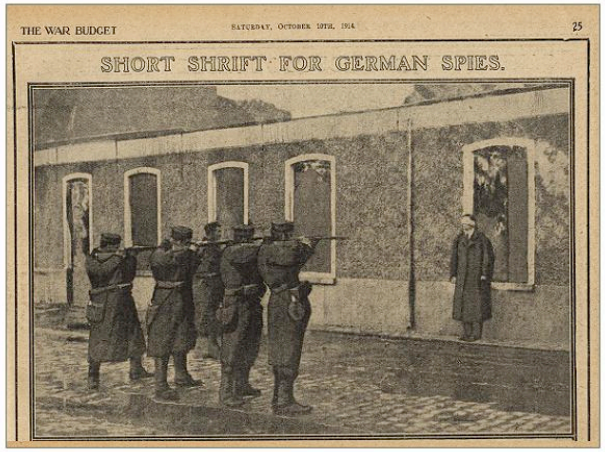
Iwo Jima, World War II

http://grin.hq.nasa.gov/IMAGES/SMALL/GPN-2004-00012.jpg
http://www.archives.gov/publications/prologue/2004/winter/top-images.html
http://funny-insurance.blogspot.com/2007/05/top-10-best-funny-photo-of-funny.html

Not All Pictures Are Real

- Why not real?
  - Modified to influence opinions
  - Enhanced to convey a point
  - Designed to show techniques
- Fake Photos
  - Old School:
    - Staged
    - Mislabeled
  - Hi-tech Methods:
    - Spliced
    - Airbrushed
    - Digitally Modified
    - “Shopped”
- Legal Implications
- Image as Authentication

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

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
  - Ashcroft v. Free Speech Coalition, 535 U.S. 234 (2002)
  - CPPA violated free speech rights
  - Distinction between “CP” and “VCP”
  - VCP does not use real children (it is regular “pornography”)

Images as Authentication

419eater.com

My Problem with MySpace

http://www.peacexpeace.org/elements/images/familysignguy.jfif

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
  - Principle Component Analysis
  - Wavelet Transformations

Observation

Warez Factory

Warez F

Scaled, sharpened, enhanced

Things to Look For

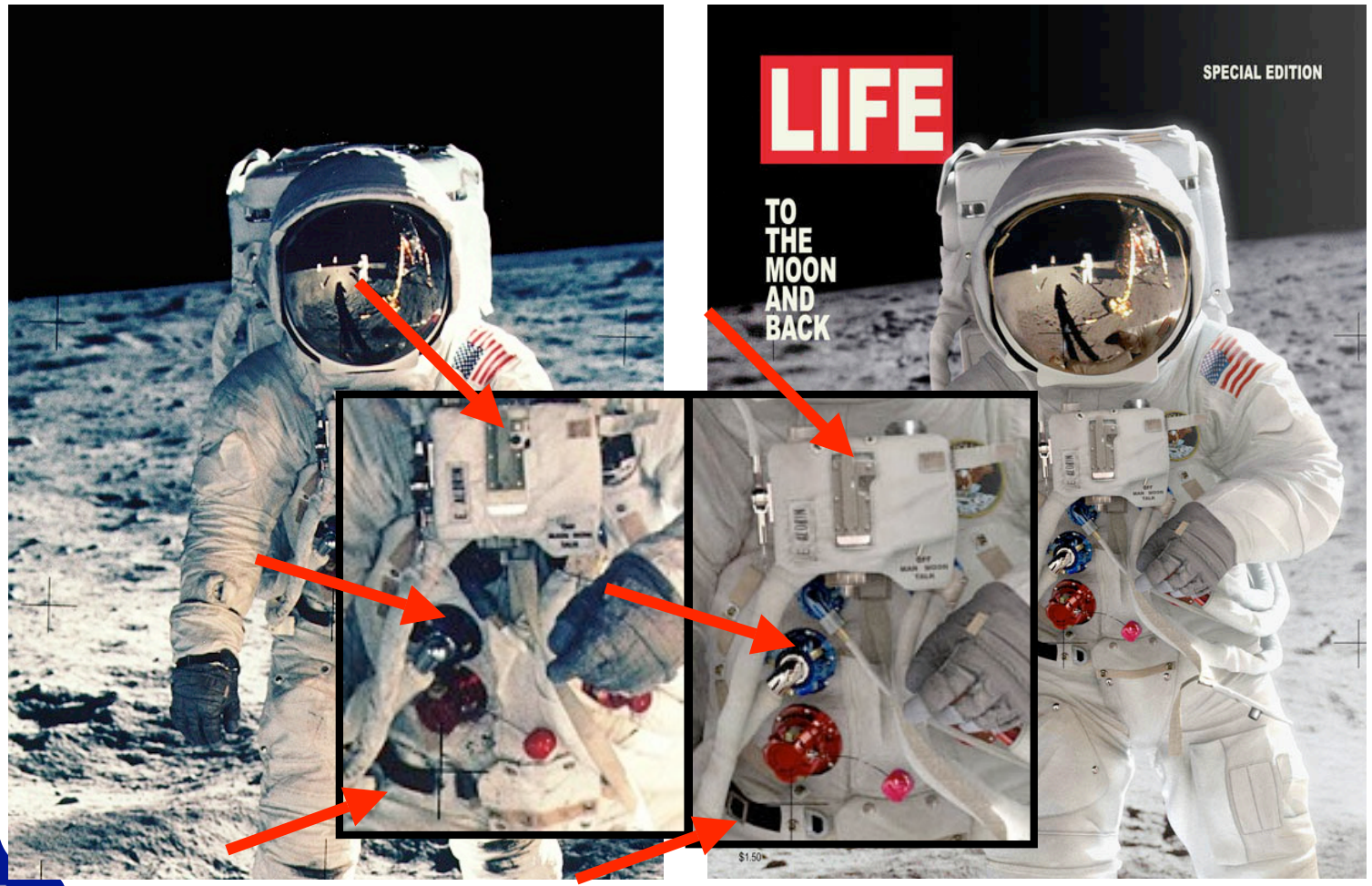
- Time
  - Clocks, calendars
  - Dated materials
- Location
  - Language
  - Region-specific technology
  - Currency and Electrical Outlets!
- Other
  - What's on the computer screen?
  - Any other identifiable elements

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

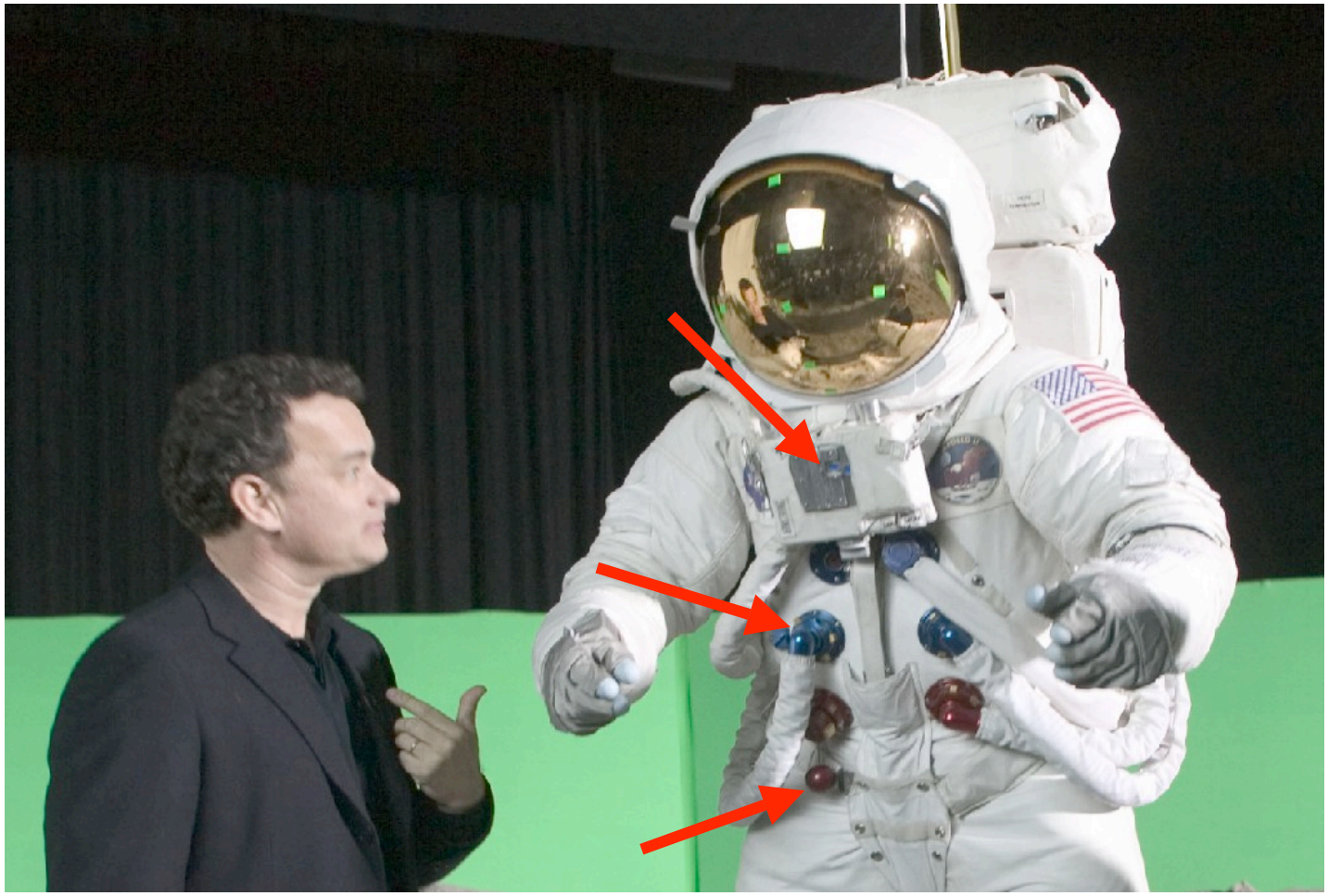
Example: Buzz Compare



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
Y Cb Cr 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

```
# Quantization table
#   Table index=0 (luminance)
    3   2   2   3   2   2   3   3
    3   3   4   3   3   4   5   8
    5   5   4   4   5  10   7   7
    6   8  12  10  12  12  11  10
   11  11  13  14  18  16  13  14
   17  14  11  11  16  22  16  17
   19  20  21  21  21  12  15  23
   24  22  20  24  18  20  21  20

# Quantization table
#   Table index=1 (chrominance)
    3   4   4   5   4   5   9   5
    5   9  20  13  11  13  20  20
   20  20  20  20  20  20  20  20
   20  20  20  20  20  20  20  20
   20  20  20  20  20  20  20  20
   20  20  20  20  20  20  20  20
   20  20  20  20  20  20  20  20
   20  20  20  20  20  20  20  20
```

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 convert
- Combine to find quality

  100.0 - 15.59 + 9.65 = 90.35%

  Call it 90%

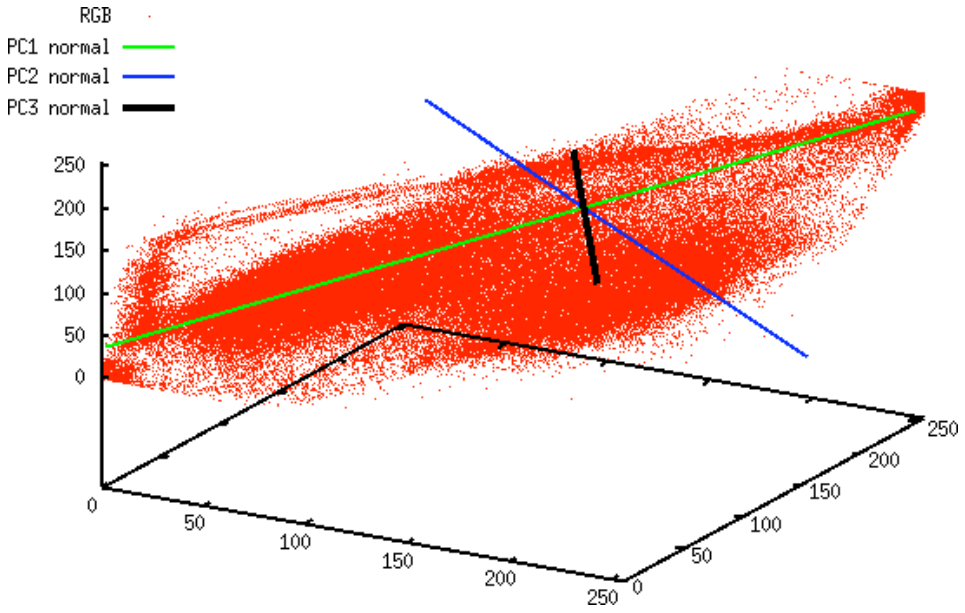
```
# Quantization table
#   Table index=0 (luminance)
    3   2   2   3   2   2   3   3
    3   3   4   3   3   4   5   8
    5   5   4   4   5  10   7   7
    6   8  12  10  12  12  11  10
   11  11  13  14  18  16  13  14
   17  14  11  11  16  22  16  17
   19  20  21  21  21  12  15  23
   24  22  20  24  18  20  21  20

# Quantization table
#   Table index=1 (chrominance)
    3   4   4   5   4   5   9   5
    5   9  20  13  11  13  20  20
   20  20  20  20  20  20  20  20
   20  20  20  20  20  20  20  20
   20  20  20  20  20  20  20  20
   20  20  20  20  20  20  20  20
   20  20  20  20  20  20  20  20
   20  20  20  20  20  20  20  20
```

Quantifiable Problem

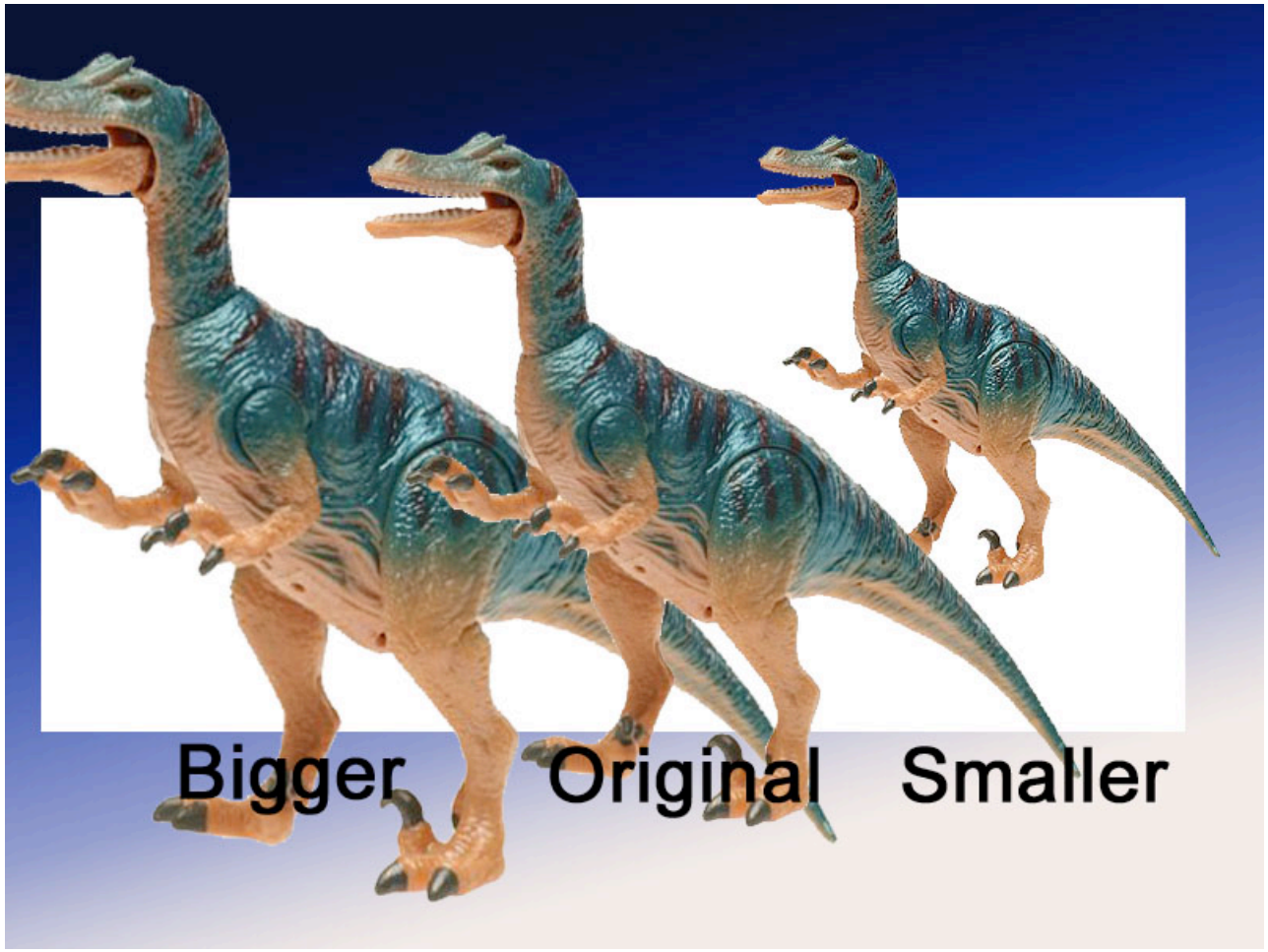
- Data loss is cumulative
- Resave problem:
  - Save an image at quality of 75%
  - Resave image at 90%
  - Image does *not* get better!
    - 90% of 75% = 67.5%
  - Quantization tables reflect 90%, not 75% or 67.5%
- How to detect image resaves?
  - Principal Component Analysis!

Principal Component Analysis

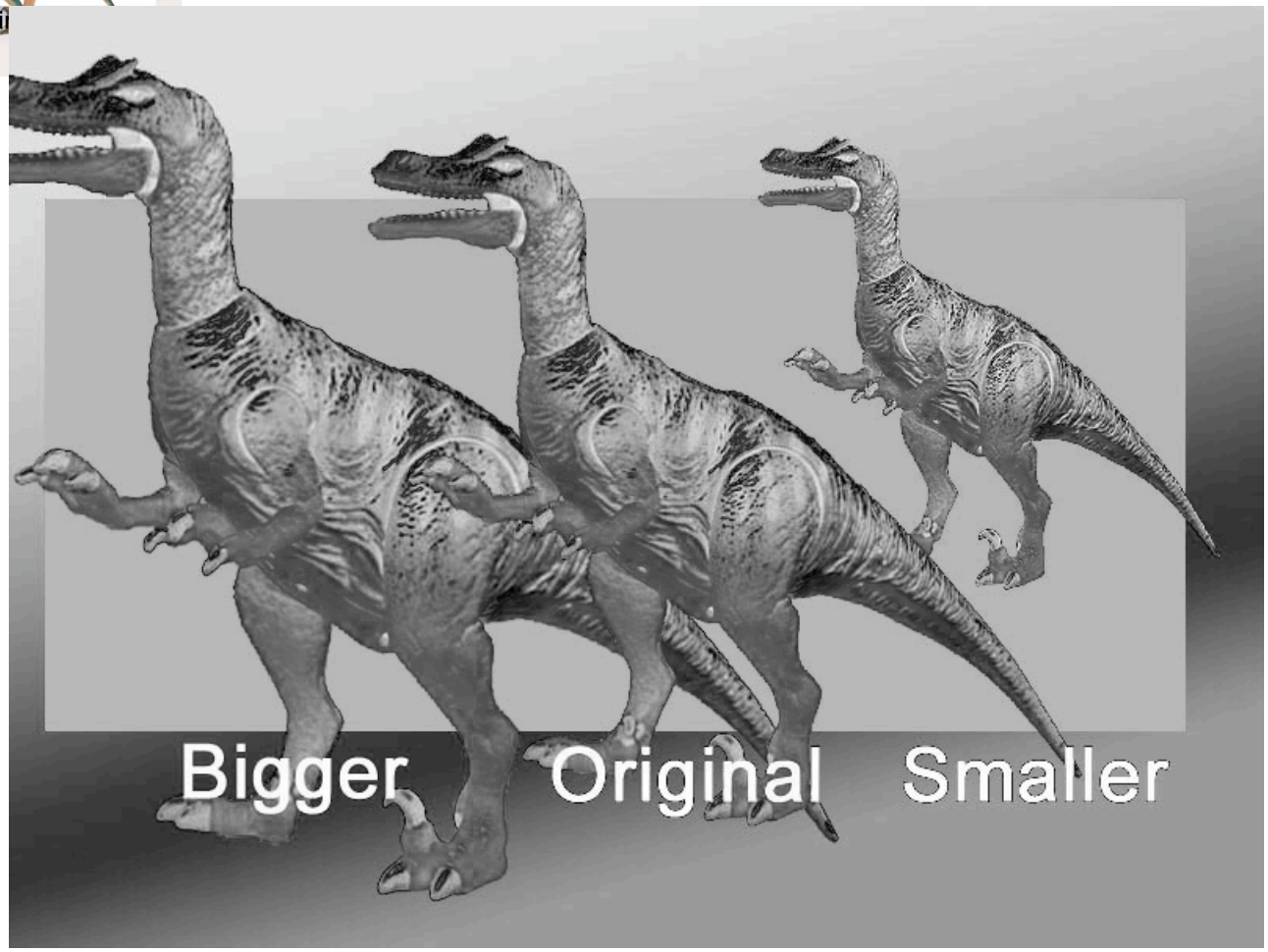
- PCA separates info
  - Computer vision
  - Data compression
- Identifies widest variance among points
  - 3D = 3 components
    - PC1 = widest
    - PC2 = next widest
    - PC3 = narrowest



PCA Example



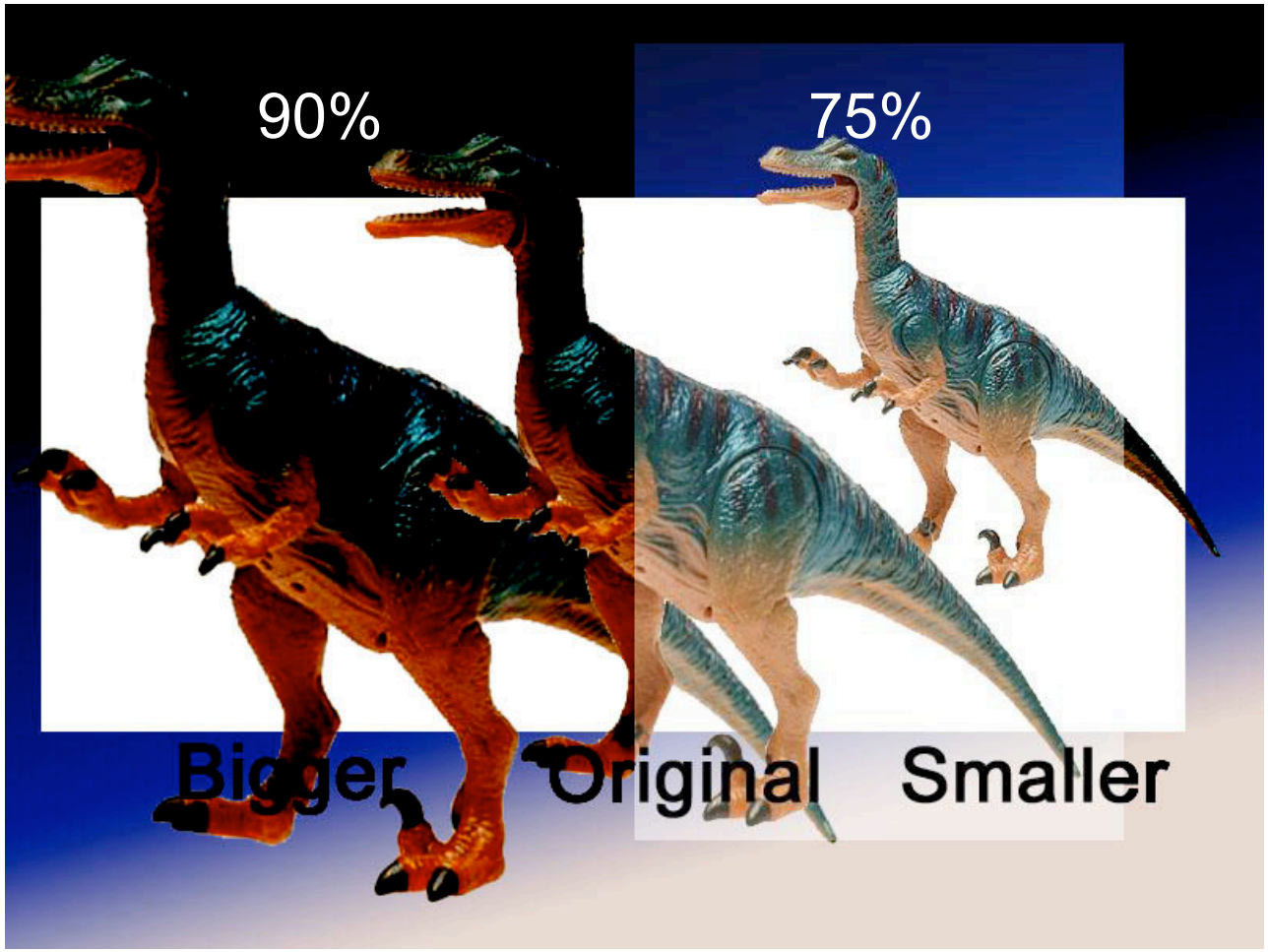
PCA Example



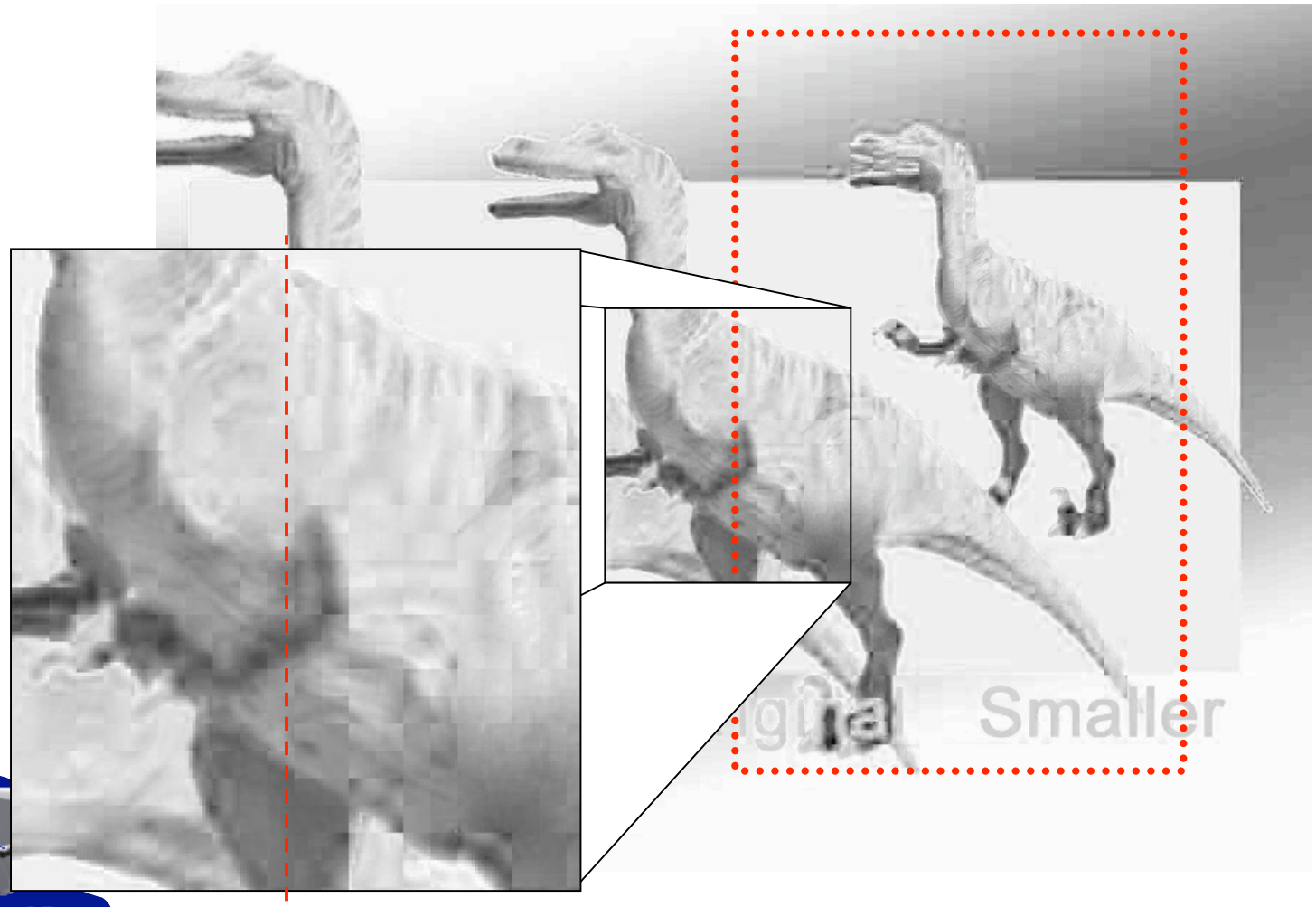
PC1 with Artifacts

95%
90%
80%
70%
60%
50%

PCA Mixing: 90% with 75%



PCA Mixing: 90% with 75%



Example: Back to the Moon

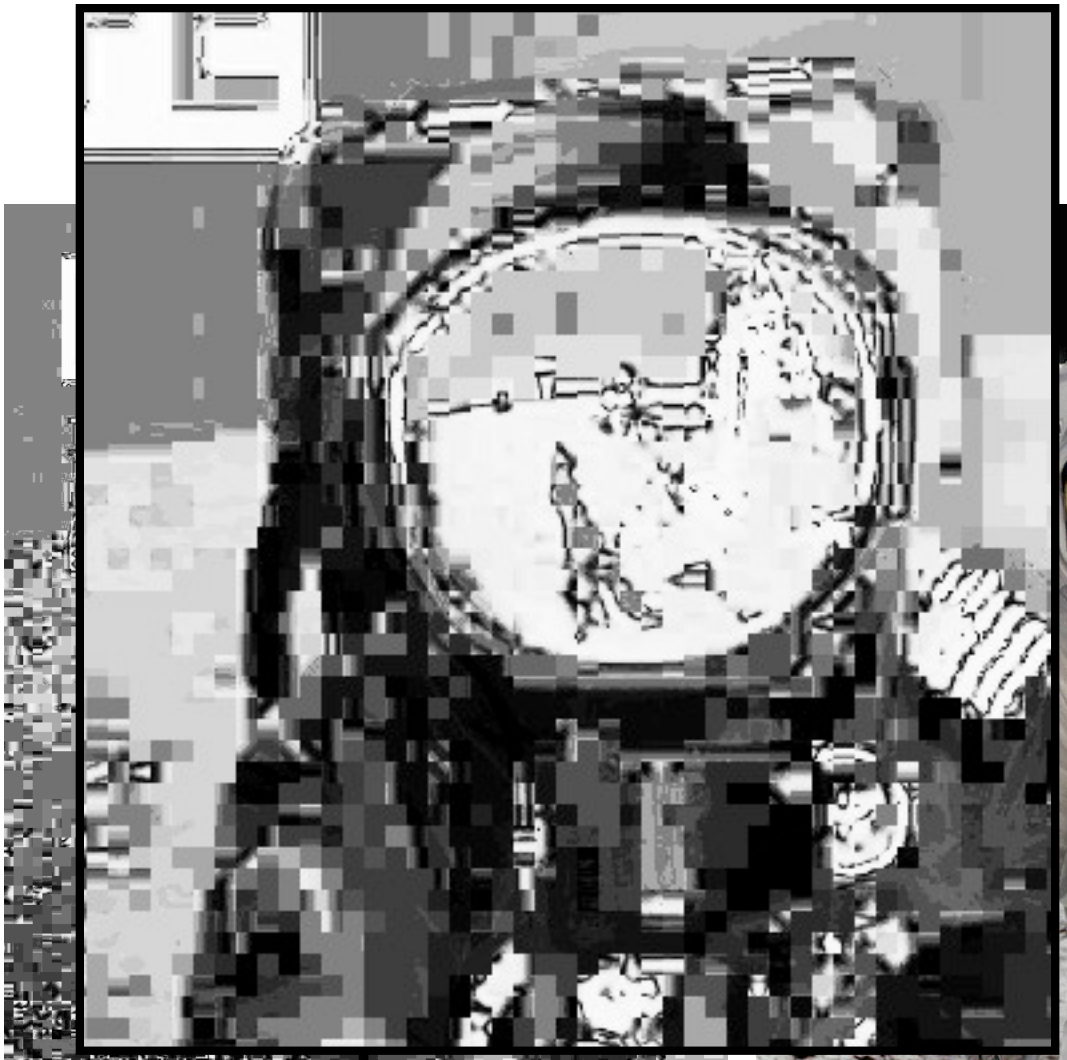
Buzz Aldrin Moon Walk

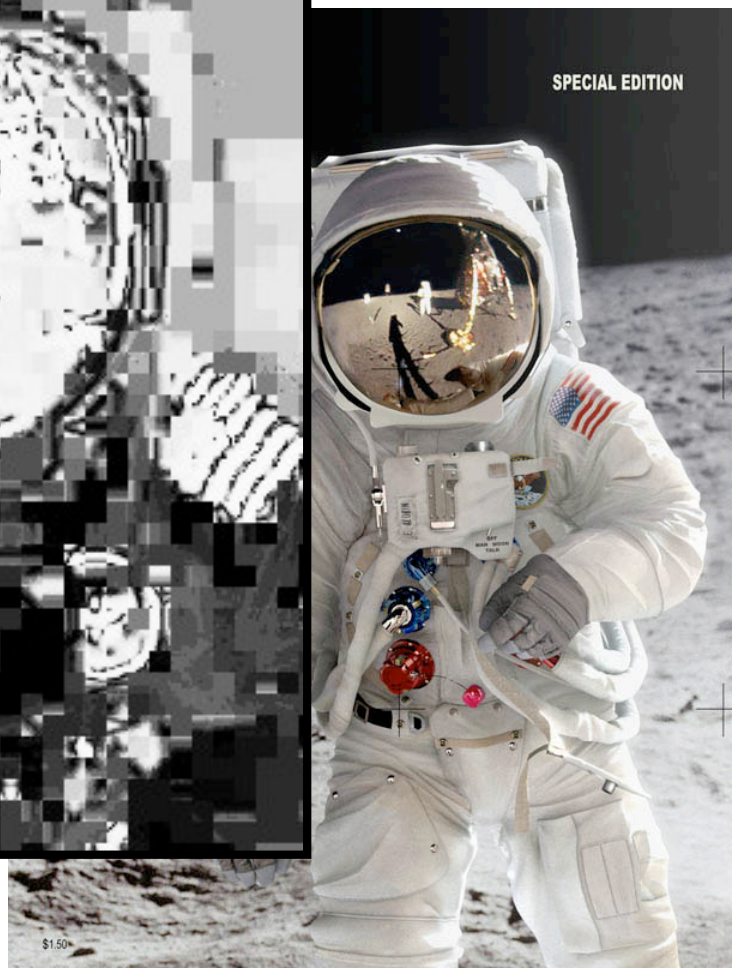
- "All the image are made in 3DS MAX and postprocessed in Combustion and Photoshop."

  http://forums.cgsociety.org/showthread.php?t=323480
- JPEG Q tables say:
  - Photoshop
  - 89% quality

Buzz Aldrin Moon Walk

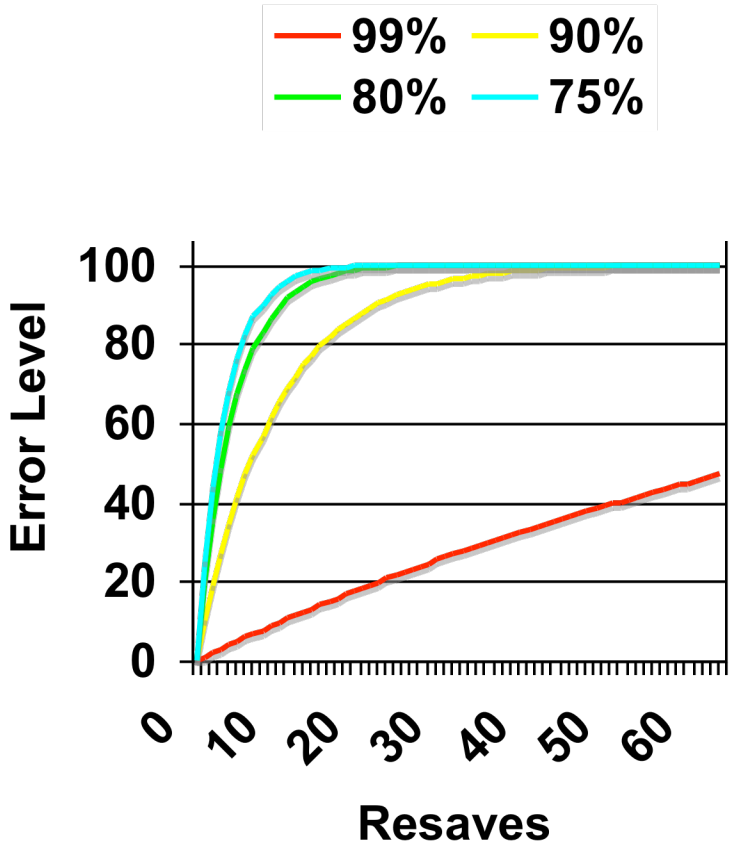
Walk





Error Level Methodology

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



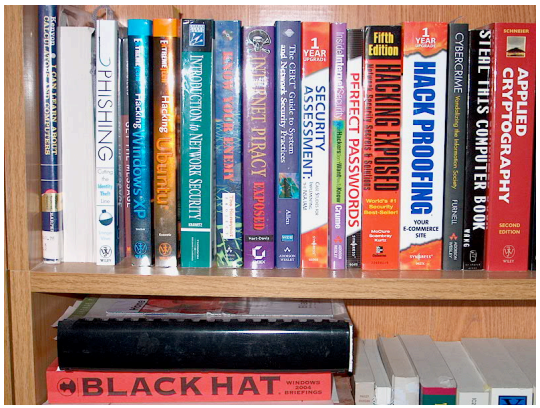
Error Level Analysis

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

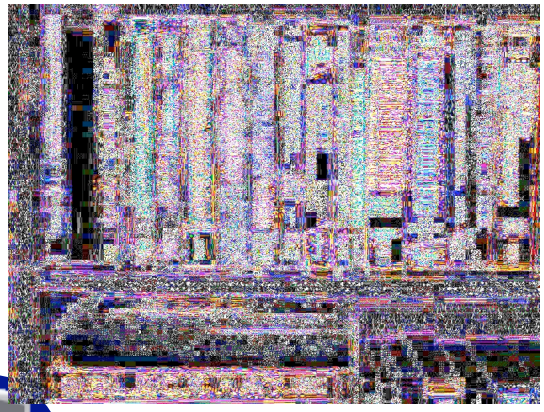
Methodology

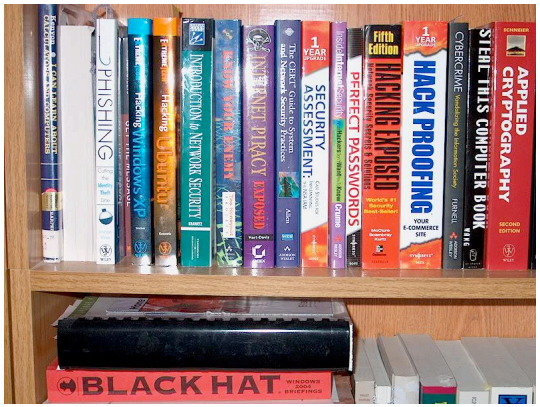
- Save image at 95%
  - Intentionally introduce known error rate
- Compare original and new 95% image
- Difference = error state
  - No difference = image local minima
  - Large difference = unstable 8x8 cell = original pixels!

Error Rate Example

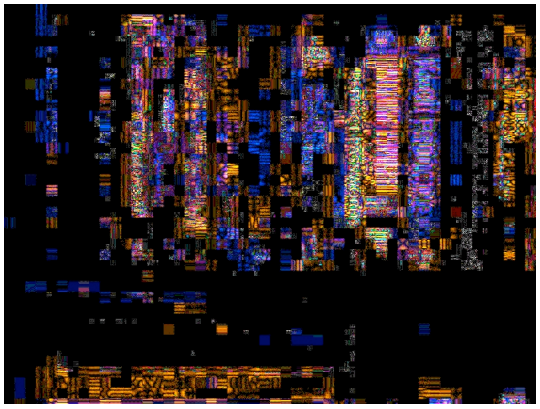


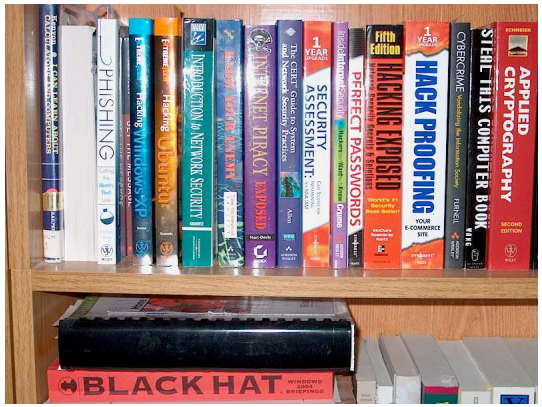
Original



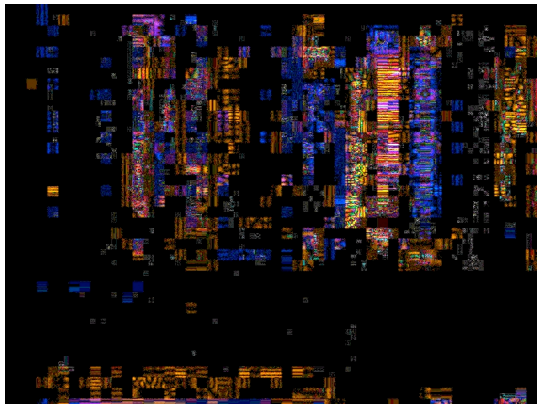


Resave #1, 75%

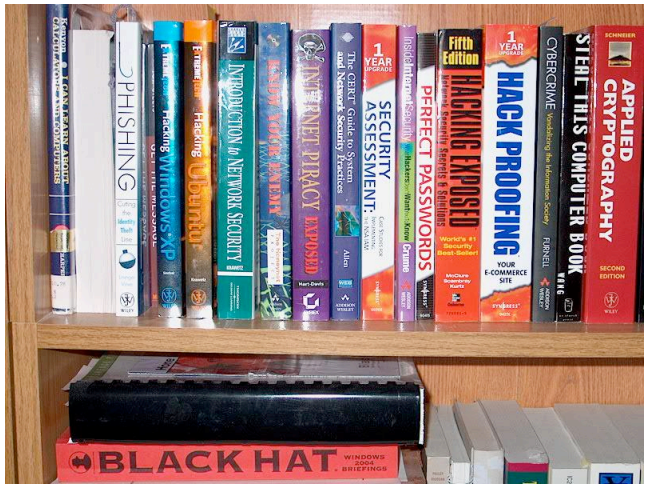




Resave #2, 75%

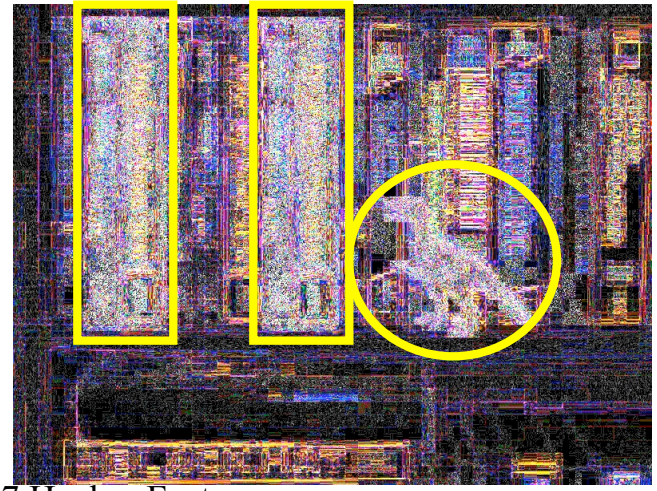


Modification Detection



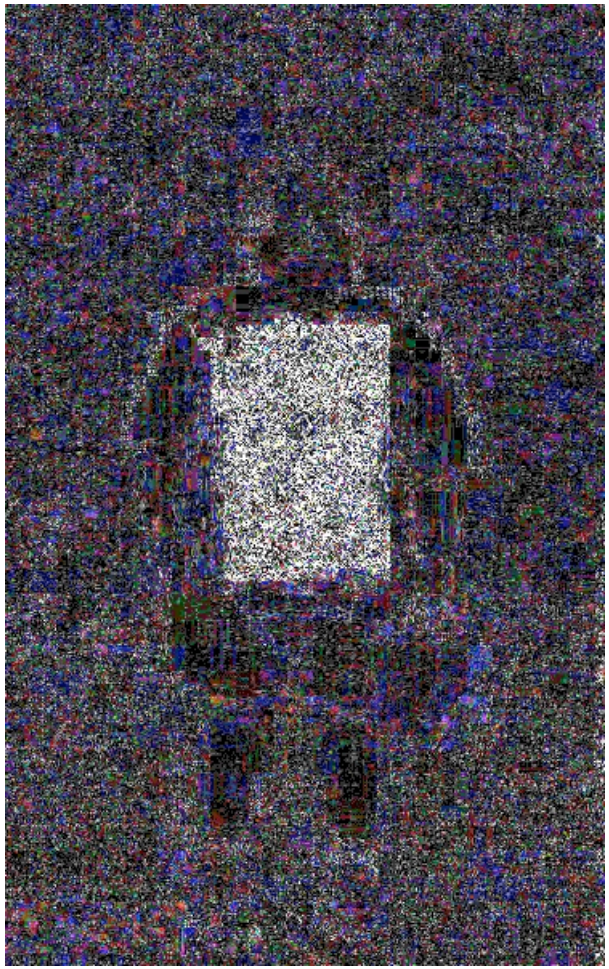
Resave #1, 75%

Edited: Books, Dinosaur

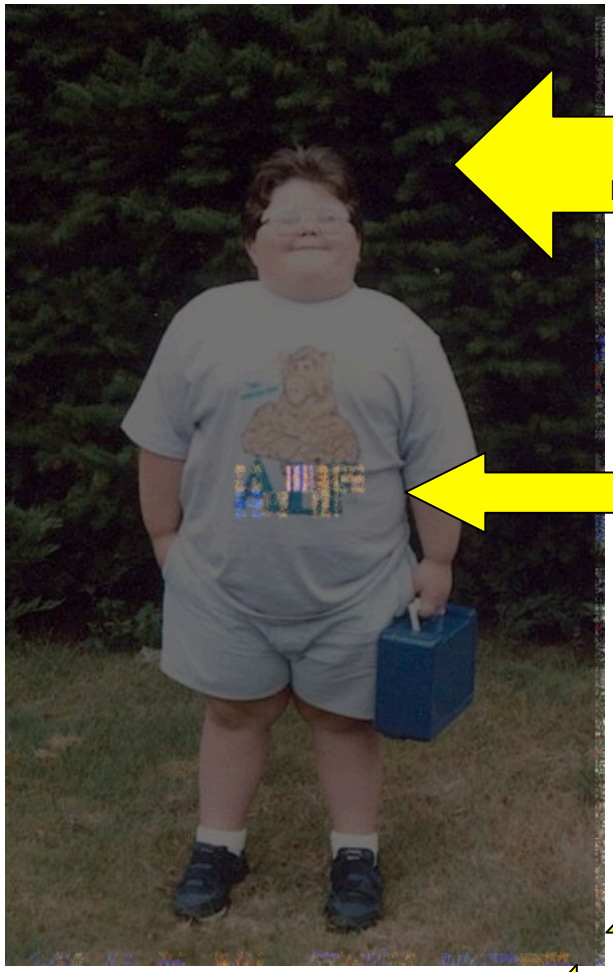


The “Alf Kid”!

“Alf Kid” Error Level Analysis



Original “Alf Kid”?



Multiple resaves

?

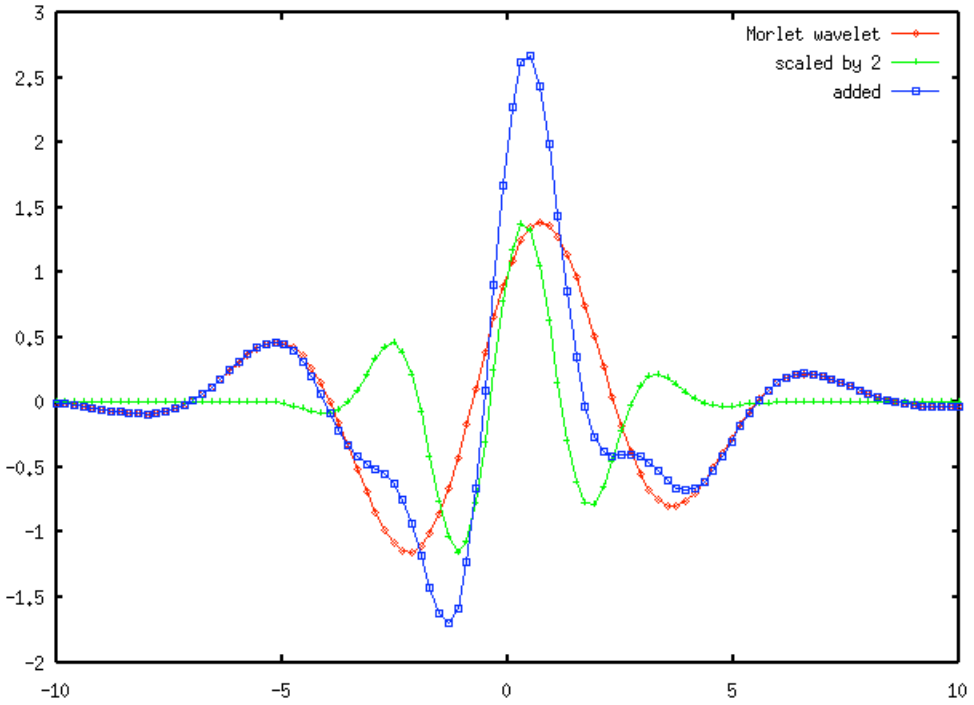
Cropped

Crash Modifications

Crash Modifications

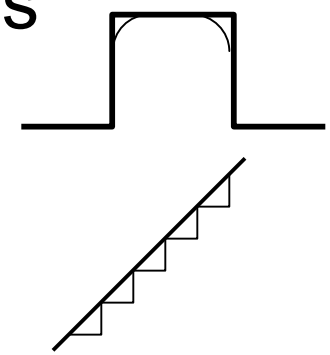
Wavelet Transformations

- 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

Original
1%
2%
3%
5%
8%
10%
20%
30%
40%


Harper's
BAZAAR
QUICK & EASY IDEAS
THAT FLATTER YOU
Fabulous At Every Age
HILLARY'S NEW LOOK FOR THE PRIMARIES
DECEMBER 2007
Harper's
BAZAAR
HILLARY CLINTON
QUICK & EASY IDEAS
THAT FLATTER YOU!
BEST HAIR & MAKEUP BUYS & TIPS
Fabulous
HILLARY'S NEW LOOK FOR THE PRIMARIES
WHY BILL IS BACK

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)

Case Study: Dr. Z

Dr. Ayman al-Zawahiri

#2 guy in Al Qaeda

USA Today

USA TODAY Home News Travel Money Sports Life Tech

World Inside News

Al-Zawahri: U.S. is talking to wrong people in Iraq

Updated 12/20/2006 8:13 AM ET

Enlarge AP

The last al-Qaeda video distributed by U.S. government contractor IntelCenter shows Ayman al-Zawahiri in a video, which was released in September 2006.

CAIRO (AP) -
al-Zawahri, t
negotiating
video broad
talking to his

"I want to tell
you are tryin
withdrawal,
and your atte
on the tape,
bulletins.

"It seems that you will go through a painful journey of failed negotiations until you will be forced to return to negotiate with the real powers," he said, without identifying these powers.

The video — which bore the logo of al-Qaeda's media production house, al-Sahab — was the 15th time this year that al-Zawahri has sent out a statement. In Wednesday's tape, he appeared exactly as in previous videos that have been authenticated by CIA analysts. He wore a black turban and white robe and pointed his finger at the camera for emphasis. As usual, 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

"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

What Else Added?

IntelCenter

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

And in the Original?

Original

Error

PCA

Wavelet

5%

What About Other Videos?

27-July-2006

Zawahiri Video Speech Regarding Lebanon and Gaza

Analysis: Error Level and PC1

Error

PCA

Analysis: PC3!

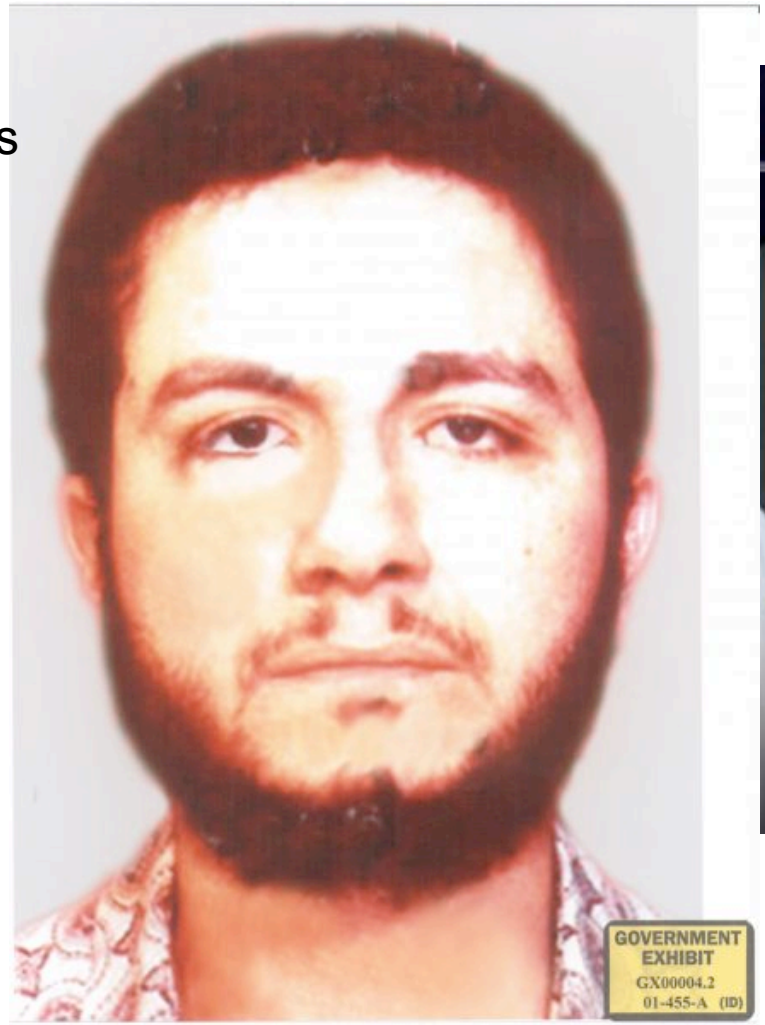
Wavelets 5%: 6 Layers!

Ayman al-Zawahiri

Mohammad Atta

Made in Layers

Identify any sources?



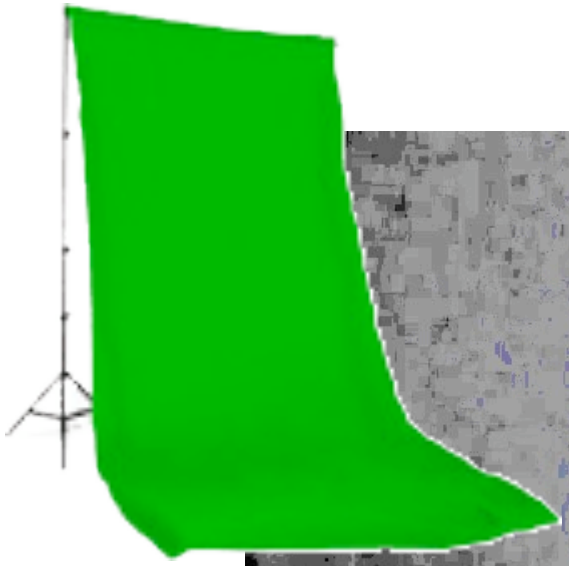
SITE Seeing

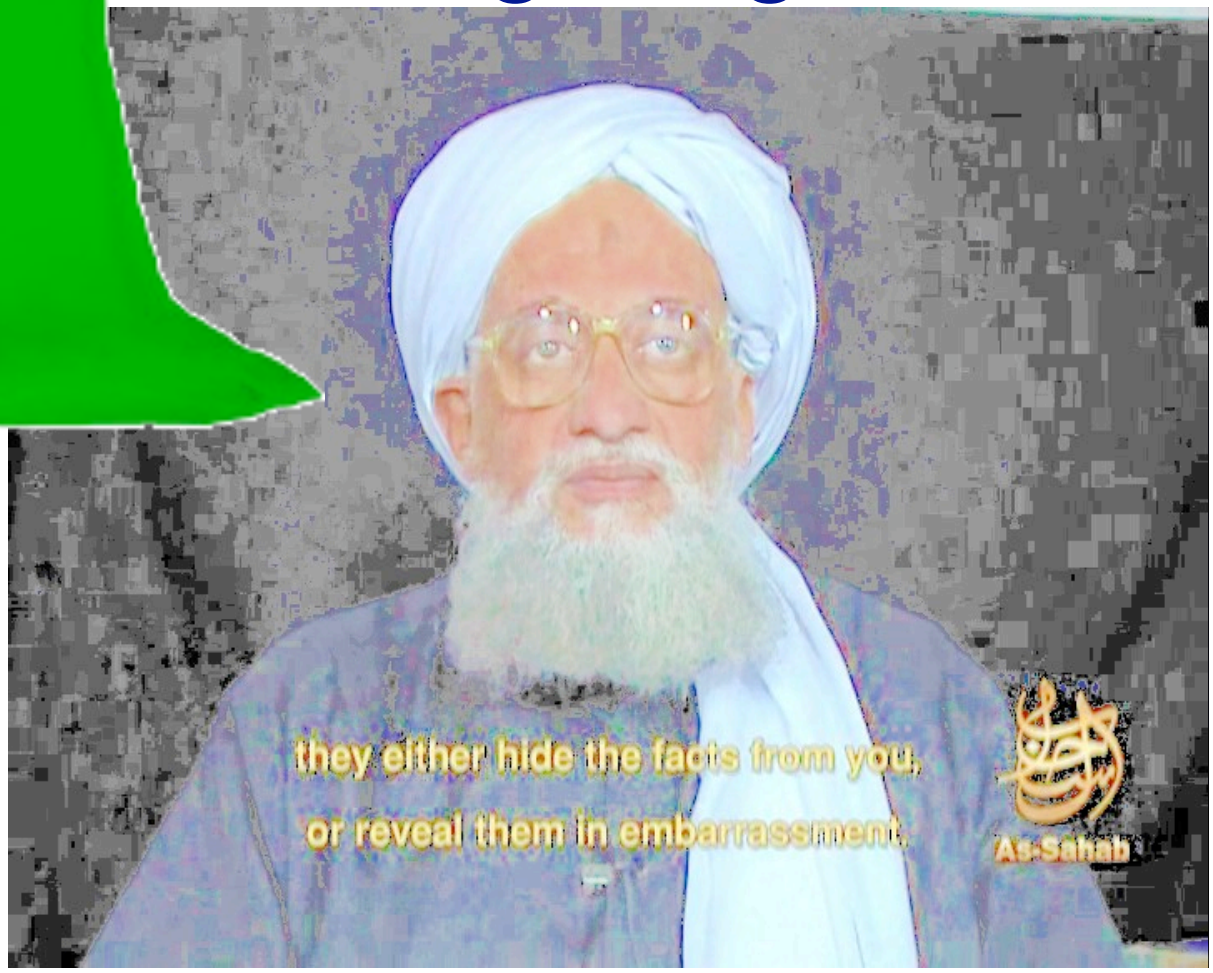
- *Saying* that there is a green screen is **not** the same as *seeing* the green screen

- SITE Institute (www.siteinstitute.org)
  - 22-Jan-2007: Intercepted Al Qaeda video!
  - 25-Jan-2007: Video released by Al Qaeda

Back in Black

Lighting





Green Screen Fun

Green Screen Fun

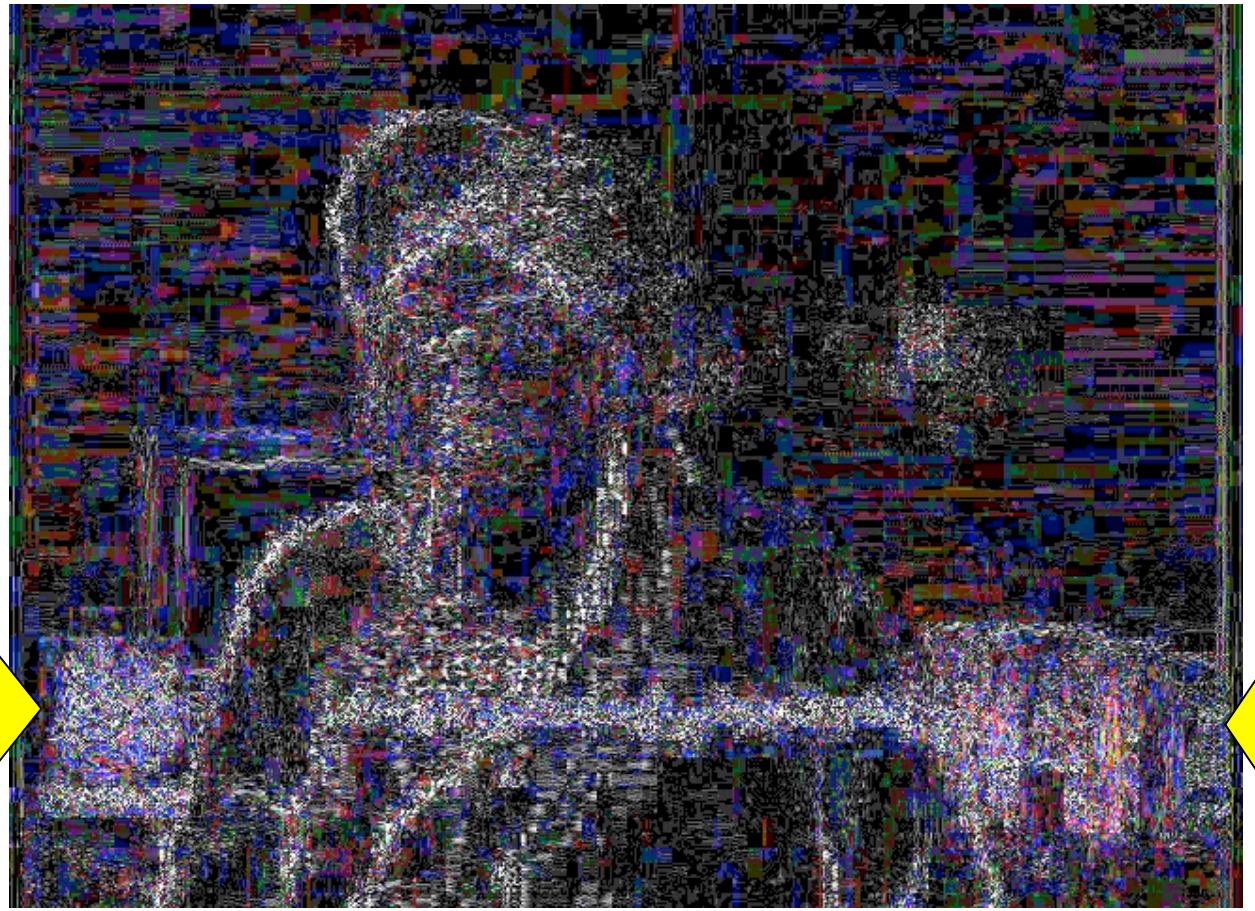
Green Screen Fun

PC1

Azzam al-Amriki

2-Sept-2006

Azzam al-Amriki



2-Sept-2006

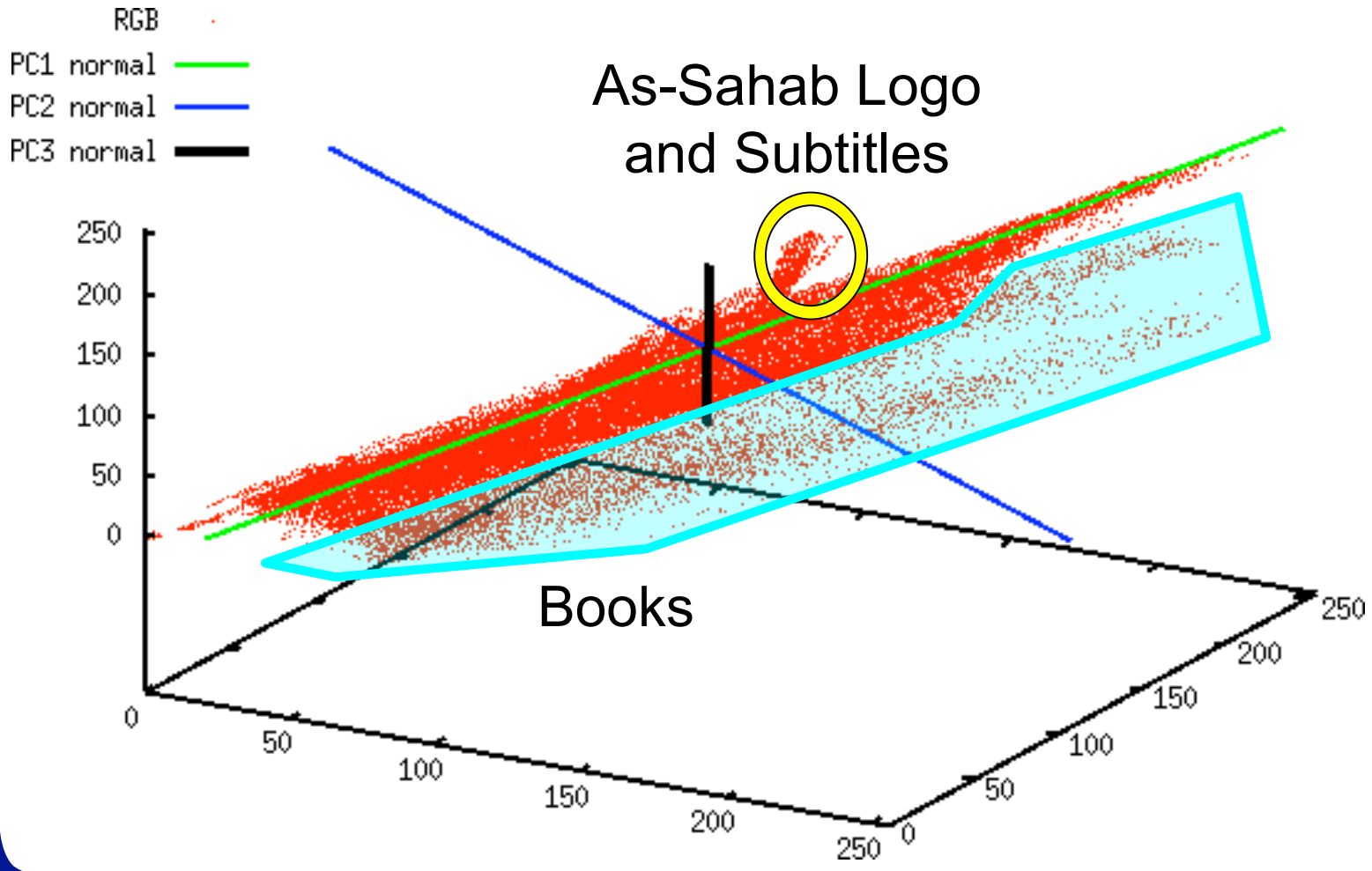
Azzam al-Amriki

Logo

Books?

2-Sept-2006

Color Graph



Conclusion

Need for Image Analysis

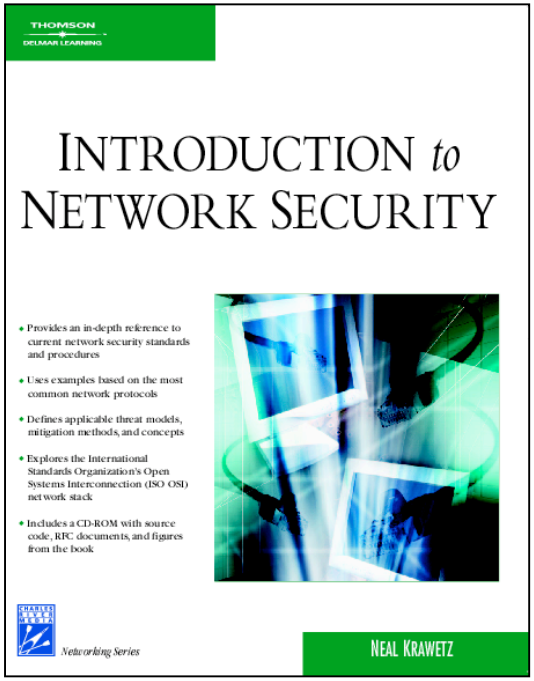
- Real versus Computer Generated
- If Modified, How?
- Uses
  - Media: Reality vs Fiction
  - Legal: Child Pornography vs VCP
  - Authentication: Real vs Doctored

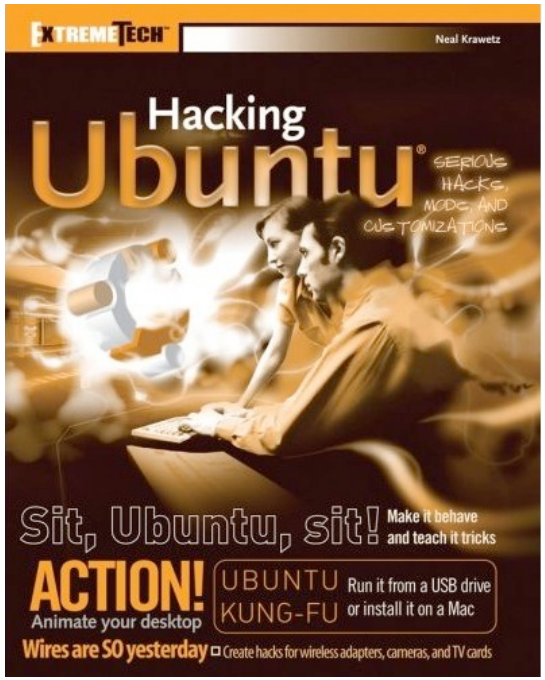
Methods Covered

- 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

Questions?

Shameless self-promotion.





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

Acknowledgements

Thanks to the following people:

Adam Bates, University of Colorado at Boulder

Cynthia Baron, Northeastern University

Paul Whyman and Bob Gobeille

Pedro Bueno, Internet Storm Center

Hany Farid, Dartmouth College

Mark Rasch, J.D.