A Picture's Worth...

Digital Image Analysis and Forensics

Version 2

Neal Krawetz, Ph.D.

Hacker Factor Solutions

www.hackerfactor.com

Table of Contents

A Picture's Worth...

Digital Image Analysis and Forensics[1]

Neal Krawetz, Ph.D., Hacker Factor Solutions

www.hackerfactor.com

Digital cameras and video software have made it easier than ever to create high quality pictures and movies. Services such as MySpace, Google Video, and Flickr make it trivial to distribute pictures, and many are picked up by the mass media. However, there is a problem: how can you tell if a video or picture is real? Is it computer generated or modified? In a world where pictures are more influencial than words, being able to distinguish fact from fiction in a systematic way is essential. This paper covers some common and not-so-common forensic methods for extracting information from digital images. This paper describes methods to distinguish real images from computer generated ones, and identify how pictures have been digitally manipulated.

1 Terminology

The following terms are used throughout this paper:

- **Computer generated (CG)**. An image created entirely with computer software. For example, every scene from the movie *Toy Story* is CG.
- **Digital photo**. A photograph from a digital camera or scanned image that has not been manipulated.
- **Digitally enhanced photo**. A digital photo that has been manipulated. This includes minor manipulations such as cropping and red eye reduction, to major re-coloring or digitally combining with other images.
- **Photoshopping**. Adobe Photoshop is a popular tool that can digitally enhance images. Images that have been modified using Photoshop or similar drawing tools (e.g., Gimp, Corel Draw, MS Paint) are described as being "photoshopped" or "shopped". The quality of the shopped image depends on both the tool and the artist. Many shopped images are obvious, while others can be very subtle.
- **Principal Component Analysis (PCA)**. An analysis approach based on data clustering.
- **Wavelet Transformations**. An analysis method based on signal decomposition.
- **Luminance Gradient (LG)**. An analysis method based on light intensity direction.

2 The Problem with Images

Images have power. Whether it is the space shuttle exploding during launch, man walking on the moon, or soldiers raising a flag on Iwo Jima during World War II, powerful images influence society. The advent of sophisticated digital imaging software and photo-realistic graphics allow artists to strengthen images or convey alternate meanings. Unfortunately, many altered pictures are presented as "real".

Prior to the digital age, powerful images were sometimes staged, edited through techniques like negative splicing and airbrushing, or simply mislabeled to convey an alternate meaning. Digital imaging software has removed the need for physical recreations. Images can be spliced together, graphically enhanced, or completely computer generated. Detecting these manipulated images can be difficult. Many identified manipulations have been followed by controversy. For example, *Newsweek* presented Martha Stewart on the March 2005 cover (Figure 1). The image

[1] All pictures are copyright by their respective owners and are includes for academic discussion and research. This complies with the copyright laws of United States as defined and stipulated under Title 17 U.S. Code.

actually showed Martha's head on someone else's body. This led to criticisms about *Newsweek*'s misrepresentation of the image.[2]

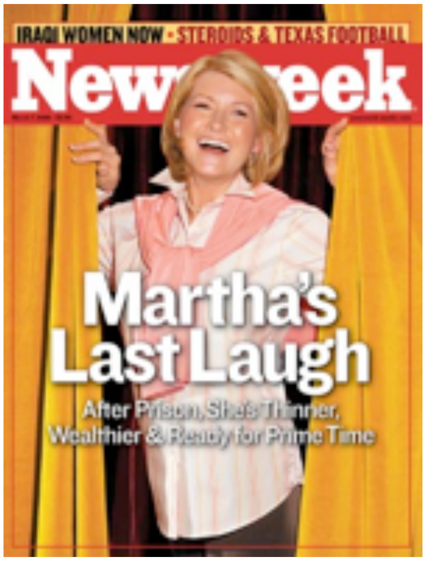

IRAQI WOMEN NOW • STEROIDS & TEXAS FOOTBALL
Newsweek
Martha's Last Laugh
After Prison, She's Thinner, Wealthier & Ready for Prime Time

Figure 1. The March 2005 cover of *Newsweek* shows Martha Stewart's head on a model's body.

Image manipulation can also lead to copyright issues. Many celebrities, including Sandra Bullock and Pamela Anderson, have had their heads placed on pictures of nude women in order to give the impression of nude celebrity pictures. In these cases, celebrities own their likeness and the misuse may infringe upon their copyright.

2.1 Child Pornography

There is a much darker side to digital imagery. In 1996, the United States passed the Child Pornography Prevention Act (CPPA). This law prevents the use of children in sexually explicit materials. However, it did not make a distinction between "real" children and computer generated or illustrations of children.[3] In 2002, the United States Supreme Court ruled that CPPA violated free speech rights.[4] In particular, if the child is not real, then no child was harmed and therefore the CPPA does not apply. This ruling created a distinction between "child pornography" (CP) and "virtual child pornography" (VCP).

- **Child pornography**. The use of real children in sexually explicit situations is prohibited by the CPPA.
- **Virtual child pornography**. Images that do not use real children are classified as "pornography" and are protected as free speech.

For law enforcement, the distinction between CP and VCP is extremely important when prosecuting pedophiles. A case based on VCP is much more difficult to justify than one that uses CP. As a result, the different aspects of image manipulation must be identified. For example, if an oil painting algorithm is applied to a real photo, then the image is digitally modified but still represents a real child. In contrast, a completely computer generated child is clearly VCP – if it can be identified as being computer generated. Other complex scenarios have not yet to be tested by the courts, including "Frankenstein Children," where pieces of different children are photoshopped together to form one or more children.

[2] http://money.cnn.com/2005/03/03/news/newsmakers/martha_photo/

[3] http://www.politechbot.com/docs/cppa.text.html

[4] http://supct.law.cornell.edu/supct/html/00-795.ZS.html

2.2 Digital Authentication

Many web sites require authentication in order to resolve complaints about impersonations. In some cases, photographs or copies of government IDs are requested. MySpace is one example of a site that requires photographic authentication. When reporting an online imposter or hijacked account, MySpace requires a digital image as authentication. The image should be of the account holder holding a sign that includes the MySpace account number.[5] However, MySpace has no apparent means to authenticate the image. As a result, MySpace has erroneously accepted the following types of forged images for removing or hijacking user profiles:

- **Not me**. Since MySpace has no means to tell if the picture is actually of the account holder, any person holding an appropriate sign can claim to be the account holder.
- **Fake text**. If the account holder has a picture of him holding any type of signboard or paper, then Photoshop can be used to replace the text.
- **Fake sign**. A picture of the valid account holder can be photoshopped so that he appears to be holding a sign.

Although there are advanced methods to detect forged images, doctored images such as those described above have been submitted to MySpace and have resulted in the termination of MySpace profiles.

3 Methods to Analyze Images

Image analysis addresses questions about the manipulation of an image:

- Is the image real, CG, or digitally enhanced?
- If the image is real: where was the picture taken, when, and how (e.g., camera model)?
- If the image is digitally enhanced: what was manipulated and how was the manipulation accomplished?
- If the image is CG: how was the image created?

There are four different approaches to analyzing images:

1. **Observation**. Many times forgeries or modified images can be identified through direct observation; no image analysis tools are required.
2. **Basic image enhancements**. Through common algorithms such as sharpening, blurring, scaling, and re-coloring, attributes within the image can be made more distinct.
3. **Image format analysis**. Changes to images alter the file format. In the case of JPEGs and other lossy image formats, changes to images can be detected.
4. **Advanced image analysis**. Signal analysis can detect manipulations. Approaches range from error level analysis to principal component analysis (PCA), wavelet transformations, and light direction detection.

3.1 Observational Analysis

The simplest analysis approach uses human observations to pull information out of the image. Inconsistencies usually suggest digitally enhanced or CG. Items within the image may be used to identify where and when the image was created. The main items to look for in an image:

- **Specular highlights and shadows**. Sharp highlights and shadows indicate light direction. When items are merged into one picture, they may not have the same lighting.
- **Color tones in anti-aliasing**. Images rarely have sharp, crisp edges. Instead, anti-aliasing techniques blur adjacent colors together. When an object is cut out of one picture and pasted into another, the edges may contain coloring that does not match the new background.

[5] http://www.myspace.com/index.cfm?fuseaction=misc.faq&Category=3&Question=26

- **Reflections**. When images are shopped, reflections may not be modified. For example, a removed object may still be present in a reflection. Similarly, an added object may be missing within a reflection.
- **Scale**. When images are combined, they may not be at the correct scale.
- **Roots**. People or objects spliced into an image may appear to be "floating" (not rooted to the ground).
- **Objects**. Common objects in the image may be regional or time-specific. For example, electrical outlets differ by continent. Text and currency in an image may identify a location. And clocks and calendars disclose time. Similarly, photos that show computer screens may display a specific application, operating system, or recognizable web site.
- **Duplication**. Items within the picture may be copied and duplicated in other locations around the image. The duplications may be unmodified, scaled, rotated, flipped, or otherwise manipulated.

Unfortunately, many items cannot be clearly recognized without enhancing the image. Basic image enhancement methods can clarify elements within a picture for easier identification.

3.2 Basic Image Enhancements

Most photo editing tools contain basic image enhancement functionality. These can be used to clarify pictures (or regions within pictures). Common enhancement functions include:

- **Brightness and contrast**. These algorithms can make a dark area lighter, revealing hidden objects, or tone down the brightness from a washed-out image.
- **Color adjustment**. Lighting can dramatically impact the color scheme within an image. Many tools permit changing the temperature (frequency range) of an image or adjusting the individual color components.
- **Invert**. Inverting portions of an image (negative image) can reveal information obscured due to similar coloring.
- **Sharpen and blur**. Items that are not in focus, or blurred due to motion, may be corrected with these functions.
- **Normalization and histograms**. Advanced tools allow the viewing and modification of color ranges. For example, if an image appears too uniformly colored, then a normalized image will create a greater color range.
- **Scale**. With some image formats, objects can be zoomed in before incurring distortion. For example, a high-resolution JPEG may be zoomed in as much as 200% before the image becomes too distorted. Similarly, very large images may be shrunk for easier viewing.

3.2.1 Example: Warez Factory

A picture of an unauthorized CD duplication facility (Figure 2) was recently presented in a forum dedicated to warez. The picture shows CD duplicators, an inkpad for affixing "authentic" marks, and even CD case slipcovers. However, it is unclear how long this picture had been circulating the Internet. The question becomes: when and where was this picture taken?

The only distinct text in the room comes from the *Tarzan & Jane* movie poster and the bottle of isopropyl alcohol. By sharpening these images and enhancing the contrast, the text becomes readable (Figure 3 and Figure 4). In particular, the text is in Spanish, so the location is likely Mexico, Venezuela, Argentina, or any other Spanish-speaking country. The bottle of alcohol includes the text "Madrileno" (Madrid), suggesting Spain. The movie poster for *Tarzan & Jane* denotes a release-to-DVD date of 26 Marzo (March 26), but does not specify the year. The movie was released to DVD in Spain on 26-March-2003. Considering that DVD promotional posters are usually released a few months before the video becomes available, and are hard to find items shortly after the release, the photo's date range can be estimated: January – May 2003. Although there are many items in the room, no other items have been identified with specific years beyond 2003.

In addition to the date and location, features of the room are identifiable. For example, the telephone connector on the wall (Figure 5) is a $3 part[6] for creating a new phone outlet. This suggests that the room was not originally phone-ready with an RJ11 jack. Since no phone is plugged into the outlet, it could have been installed specifically for the computers.

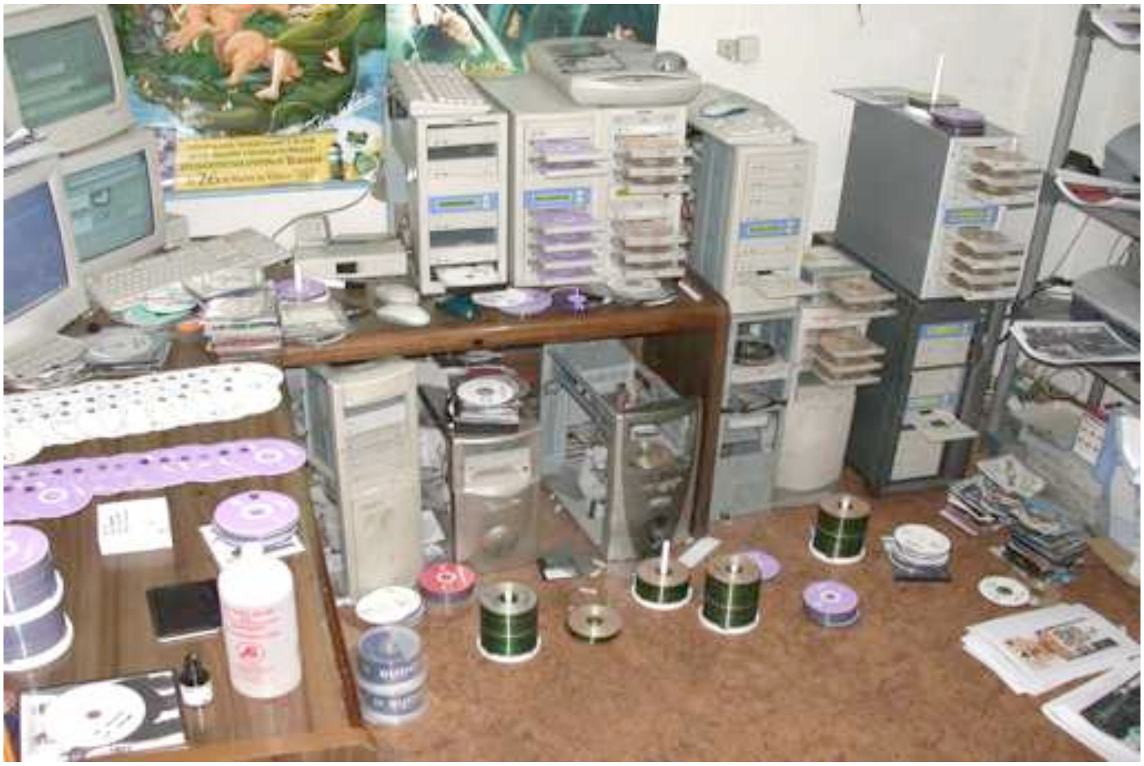


Figure 2. An unauthorized CD duplication factory. This images comes from http://www.bork.ca/pics/?incoming/warez_factory.jpg, retrieved on 23-May-2007.

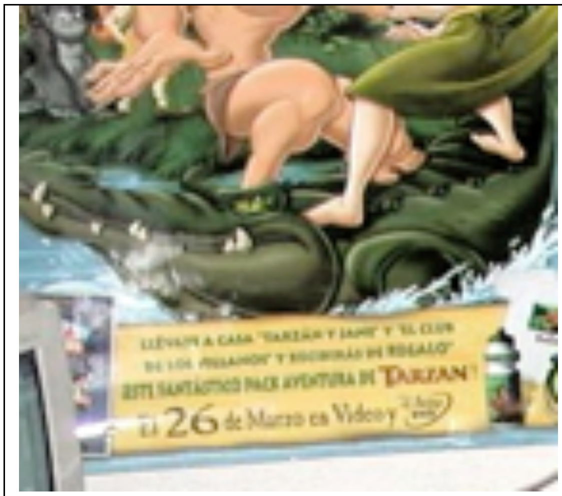


Figure 3. *Tarzan & Jane* poster announcing the DVD release on "26 de Marzo" (March 26). This picture has been sharpened.

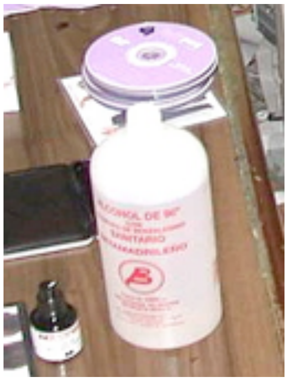


Figure 4. Isopropyl alcohol bottle with a Spanish label. This image has been scaled, sharpened, and contrast adjusted.

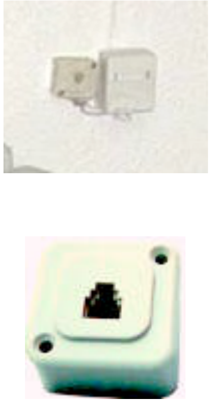


Figure 5. Telephone outlet from the wall (above) and electronic store part (below).

[6] http://www.phonecoinc.com/topic.asp?map=4&horh=home&gorl=list&group=main&category=Acc&topic=01009

3.2.2 Example: Moonwalk

In 2006, Andrea Bertaccini was awarded the "CG Choice Award" from the CG Society for the rendering of Buzz Aldrin's famous moonwalk[7] (Figure 6). According to the artist, the picture was based the original NASA photo[8]. However, details within the picture suggest additional resources.

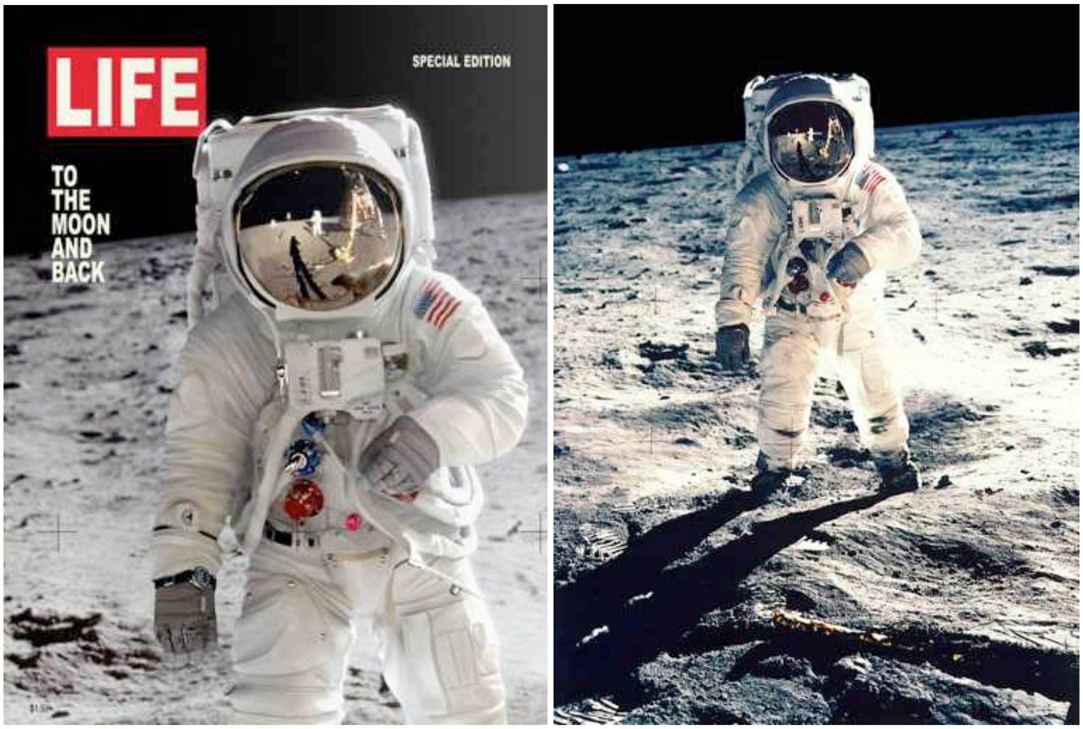


Figure 6. Image by Andrea Bertaccini (www.tredistudio.com, left) and original NASA photo (right).

A comparison of the two photos shows a significant number of differences. While the lack of moon dust in the artist's image is expected, other discrepancies are interesting:

- **Puffy**. The CG spacesuit appears "puffed out" while the real suit is wrinkled.
- **Fibers**. The real spacesuit has distinct fibers visible on the arms, similar to a wool sweater. The CG image is missing the fibers.
- **Grounding screw**. The torso box (Figure 7) in the original shows a dark grounding screw in the center. However, the artist uses a light-colored screw.
- **Connectors**. The red and blue torso connectors in the artist's image show six very reflective screws. These screws are not visible in the real photo. The actual spacesuit did have these screws, but they were not reflective.
- **Belt**. The artist's belt has metal clips. The NASA photo shows no metal clips on the belt.

These discrepancies may be due to the source of the image. The artist's image was released in 2006. In 2005, IMAX produced a movie titled *Magnificent Desolation* in which they recreated the famous moonwalk.[9] A behind-the-scenes picture from this movie (Figure 8) shows a spacesuit similar to Bertaccini's picture: the spacesuit appears puffed out (not wrinkled), there are no visible fibers on the arms, the grounding screw is light-colored, the torso

[7] http://forums.cgsociety.org/showthread.php?t=323480

[8] http://www.hq.nasa.gov/office/pao/History/ap11ann/kippsphotos/5903.jpg

[9] http://www.imax.com/magnificentdesolation

connectors have six very reflective screws, and the belt has a metal clip. While Bertaccini's positioning of the figure may have been based on the NASA photo, the spacesuit appears to have been based on the IMAX recreation.

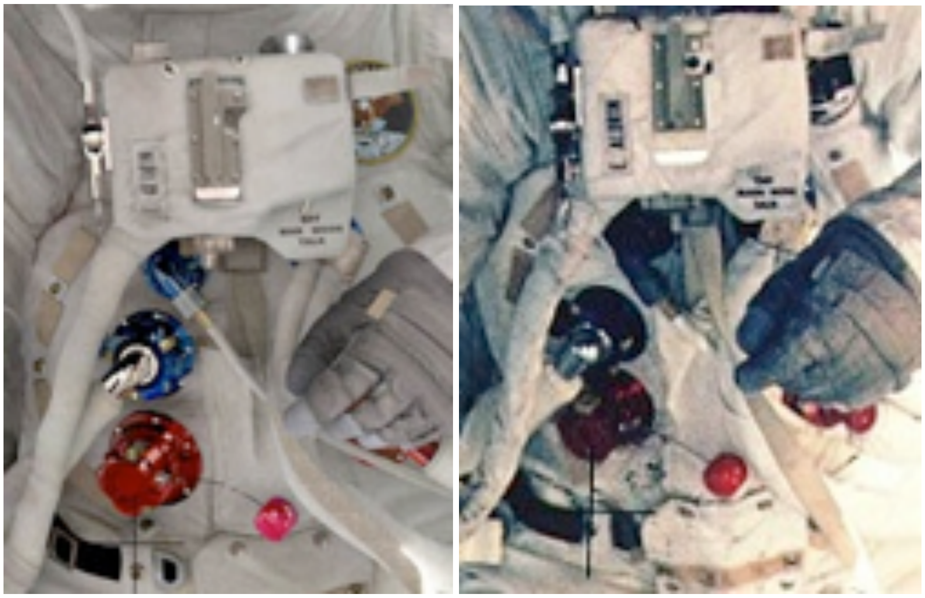


Figure 7. Close-up of the torso details on the recreation and original photos.

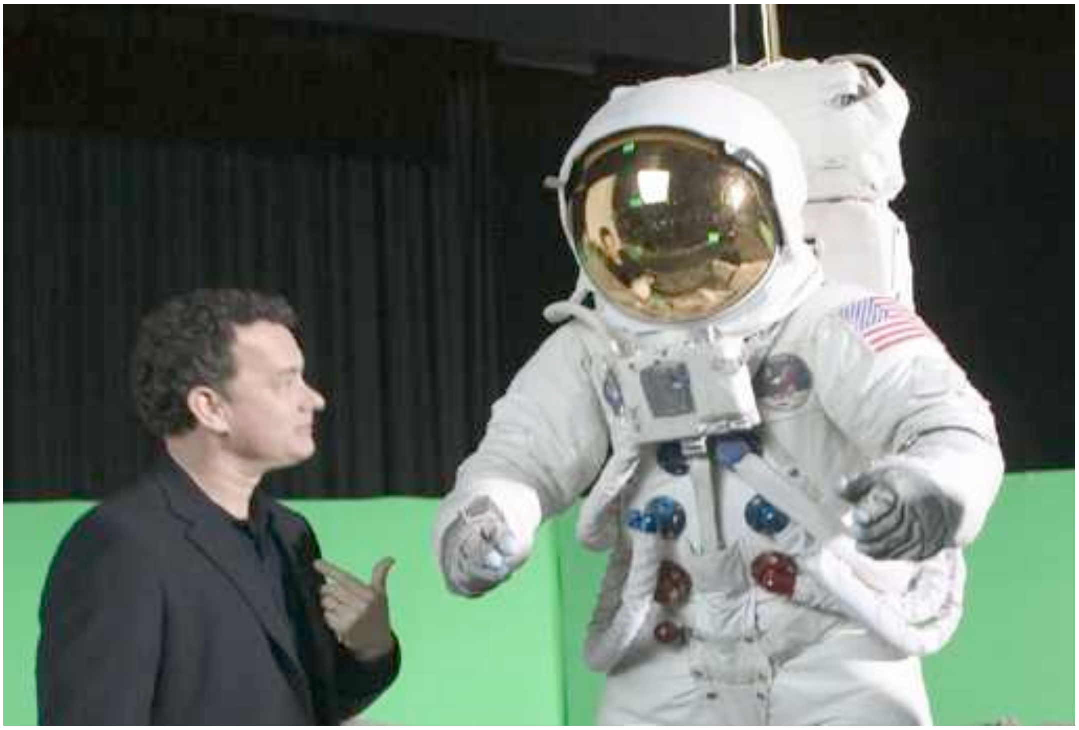


Figure 8. Behind-the-scenes photo from the IMAX movie *Magnificent Desolation*.

3.3 Image Format Analysis

Images can be stored in a variety of formats. Some, such as RAW, only contain pixel data, while other formats contain a wealth of information. In many cases, such as JPEG, GIF, PNG, and TIFF, the format can be as informative as the image. Changes to the image yield changes to the file format. Although this paper focuses on the JPEG format, the methods can be applied to other complex formats including GIF, PNG, and TIFF.

JPEG files contain a well-defined feature set.[10] Changes to the image will modify the feature set. Thus, if the features indicate manipulation then the manipulation can be identified. The feature set for JPEG includes meta data, quantization tables for image compression, lossy data compression, and subdivided image processing using 8x8 pixel cells.[11]

3.3.1 Meta Data Analysis

Most JPEGs include a significant amount of meta data that describes the source of the image. For example, a JPEG from a digital camera usually includes the camera type, resolution, focus settings, and other features (Figure 9).

```
$ 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
...
```

Figure 9. Sample meta data from a digital camera photo.

Although meta data provides a significant amount of information, it has a some limitations. First, the meta data can be edited. Although unlikely, false information about the camera type and settings can be placed within the JPEG. More likely are bad default settings. For example, most digital cameras do not account for time zones or daylight saving time, and may have clocks that drift. As a result, the time may not be accurate.

More common than intentional meta data modifications is misleading meta information. For example, a photo from a digital camera can be opened in Photoshop and manipulated. When the image is saved, it retains the camera's meta information, even though it may no longer be applicable. In addition, Photoshop does not update the meta

[10] Information about the JPEG file format can be found at <http://www.exif.org/specifications.html> and <http://www.obrador.com/essentialjpeg/HeaderInfo.htm>.

[11] Gregory K. Wallace, "The JPEG Still Picture Compression Standard", *Communications of the ACM*, April 1991 (vol. 34 no. 4), pp. 30-44.

information to record changes. As a result, meta data can be useful when it is accurate, but the data is not provably accurate.

3.3.2 JPEG Quantization Fingerprinting

Quantization fingerprinting, or ballistics[12], provides a method to detect images that do not match the specified meta data. The JPEG algorithm uses a set of quantization matrices to control image compression and quality. For JPEGs, images are converted from RGB to YCrCb. One quantization matrix handles the luminance (Y) and a second matrix handles the chrominance for both red (Cr) and blue (Cb).[13]

Ideally, the quantization tables should be generated and optimized for each image. However, computing these matrices is a time-consuming process; most digital cameras do not have the CPU power, and most applications do not want to impede the user experience by spending twenty seconds on computations each time the image needs to be saved. To simplify this process, virtually all graphical applications and digital cameras use hard-coded quantization tables.[14] These default tables are usually optimized for the data size, color spectrum, digital camera CCD properties, and manufacturer needs. For example, a photo taken with a Canon digital camera usually prints better on a Canon printer because the colors are optimized for the manufacturer. And a digital camera with three quality settings (low, medium, and high) usually has three hard-coded quantization tables.

Since the pre-computed quantization tables are manufacturer specific, they are usually distinct between applications and camera models.[15] If the quantization table can be identified, then the tool that saved the JPEG is identifiable. More importantly, if the quantization table does not match the camera information specified in the meta data, then the image can be identified as having been resaved or modified.

3.3.3 JPEG Quality Detection

When saving a JPEG image, most tools allow the selection of the image quality. In general, lower quality results in a smaller image. For example, an image saved at 90% implies roughly 10% data loss – where the pixel colors do not perfectly match the original. While a quality of 99% will result in virtually no data loss, it will generate very large files. In contrast, 75% might be good enough to convey the meaning while creating significantly smaller files.

Although a known quantization table allows the identification of the tool as well as the quality, the quantization table may not always match a known application or camera. In this situation, the quality of the JPEG must be approximated.

Each quantization table contains 64 bytes. The first byte is the DC and acts as a scalar value. The remaining 63 bytes are the AC and define compression by frequency. The algorithm developed by Hacker Factor Solutions for approximating the quality of a JPEG is as follows:

[12] Hany Farid, "Digital Image Ballistics from JPEG Quantization" Dartmouth College, TR2006-583, 2006. Available online at <http://www.ists.dartmouth.edu/library/204.pdf>.

[13] JPEGs usually have two quantization matrices. However, some applications and digital cameras create three. If there are two matrices, then one is Y and the other is used for both Cr and Cb. If there are three, then Y, Cr, and Cb each have a matrix.

[14] Many applications interpolate between static stables when an intermediate quality is requested.

[15] http://www.impulseadventure.com/photo/jpeg-quantization.html

1. Compute the average AC value for each quantization table. For example, if the tables are:

```
# 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
```

   Then the average AC value for Table 0 is 11.63 and Table 1 is 17.57.

2. The average compression value is computed across all tables. For example, since Table 0 is used once (Y) and Table 1 is used twice (Cr and Cb), the average becomes (11.63 + 17.57 + 17.57)/3 = 15.59.

3. Images are rendered using RGB, but the tables represent YCrCb. The conversion is as follows:

$$R = Y + (R-Y) = Y + Cr$$

$$G = Y - 0.51(R-Y) - 0.186(B-Y) = Y - 0.51Cr -0.186Cb$$

$$B = Y + (B-Y) = Y + Cb$$

   Since the significant ratio is 0.51, the conversion rate is determined by the difference between tables:

$$D = \|Y-Cr\| * (1.0 - 0.51) + \|Y-Cb\| * (1.0 - 0.51)$$

$$D = \|11.63 - 17.57\| * 0.49 + \|11.63 - 17.57\| * 0.49 = 5.82$$

4. The conversion rate is incorporated to form the estimated quality: 100 - 15.59 + 5.82 = 90.23. Since the JPEG algorithm uses integers instead of floating point values, this can be rounded to the nearest integer. This example image was saved with a quality level of 90%.

This algorithm accurately determines the image quality from images saved using 'xv', Gimp, ffmpeg, libjpeg, and most other tools. However, Photoshop seems to compute the percentage differently; JPEGs saved using Photoshop do not appear to match the quality specified in the user interface. For example, using Adobe Photoshop, an image saved at "80%" (using "Save for Web") has quantization tables equivalent to 91%.

3.4 Advanced Image Analysis

Image format analysis can confirm meta data inaccuracies and detect the last tool that modified an image. However, format analysis does not evaluate the image itself. Methods such as principal component analysis, error level analysis, wavelet transformations, and luminance gradient permit the identification of specific image manipulations.

3.4.1 Principal Component Analysis

JPEG uses a lossy compression algorithm; the image rendered from a JPEG file is not a perfect copy of the original image. Each time a JPEG image is resaved by a graphics editor, the image loses quality – even if the editing tool made no picture changes. This leads to a problem with quantization table analysis: if an image is saved at 75%, loaded into a drawing program, and resaved at 90%, then the quantization tables will reflect 90% while the image quality is 67.5% (90% of 75%).

Errors within a JPEG appear as blocky artifacts and color distortions. The blocky artifacts appear on the 8x8 pixel boundaries used by the JPEG algorithm. In many cases, the JPEG artifacts are too subtle for the human eye to detect. However, principal component analysis (PCA) can identify these JPEG artifacts.

3.4.1.1 Understanding PCA

PCA is used for clustering data points.[16] Each principal component defines a plane across the data set. The first principal component (PC1) identifies a plane with the widest variance across the data. In effect, the average distance from every point to the PC1 plane is maximized. The second principal component (PC2) identifies the second widest variance with respect to PC1. Since this is a three-dimensional plot, there are three principal components. The final component (PC3) identifies the smallest variance; the average distance from each point to PC3 plane is smallest.[17]

PCA analysis is commonly used for information reduction problems such as clustering, robotic vision, and data compression. For example, PC1 is associated with the greatest variance in the data, while PC3 contains the least. For lossy data compression, the values found in PC3 can be removed with the least amount of impact. Each principal component emphasizes different sections of information.

For image analysis, PCA is used to identify the color spectrum within the image. Consider an entire image plot based on the pixel colors: (R,G,B) is mapped to (x,y,z) (Figure 10). Most images have a narrow range of colors that appear as a large cluster when plotted. PC1 identifies the widest range across the color set. When two images are spliced together from different color sets, they usually end up forming two distinct clusters. With PCA, areas within the picture that come from different clusters will have noticeably different values.

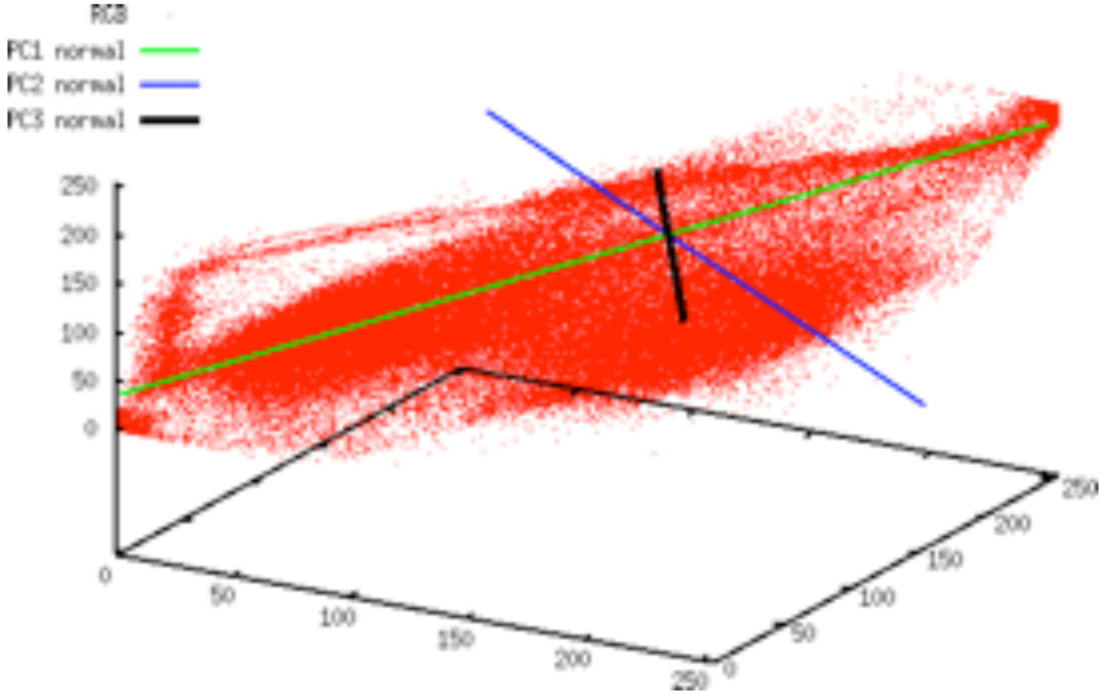


Figure 10. Sample scatter plot of an image and principal components.

3.4.1.2 PCA Visualization Options

There are two ways to render the distance from each pixel to the principal component (Figure 11). First, the distance from each point to the plane can be measured, showing the maximum variance across the data set. In general, PC1 generates very crisp gray-scale pictures; it contains the largest amount of information. In contrast, PC3 usually appears with even coloring because all points are similar distances from the PC3 plane. Although PC3 defines the least amount of information, it is usually best at identifying JPEG artifacts.

[16] Jonathon Shlens, "A Tutorial on Principal Component Analysis". Salk Institute for Biological Studies, 2005. Available online at <http://www.snl.salk.edu/~shlens/pub/notes/pca.pdf>.

[17] PC3 may not be the absolute minimum possible variance. The principal components are orthogonal to each other. PC1 is the maximum variance, and PC3 is the minimum variance with respect to PC1 and PC2.

Distances can also be measured from a line normal to the plane that passes through the center of the data set. For PC1, the average distance from each point to the line defines the narrowest variance; this is the minimum variance across the maximum amount of information. Rendering with the PC1 line usually appears to have uniform coloring. In contrast, the line associated with PC3 shows a very crisp picture; it is the maximum variance across the minimum amount of information. Although images based on the distances from the PC1 line and PC3 plane may look similar, the PC1 line contains more information. Unless specified, all PCA analysis in this paper is performed using the line normal to the principal component's plane.

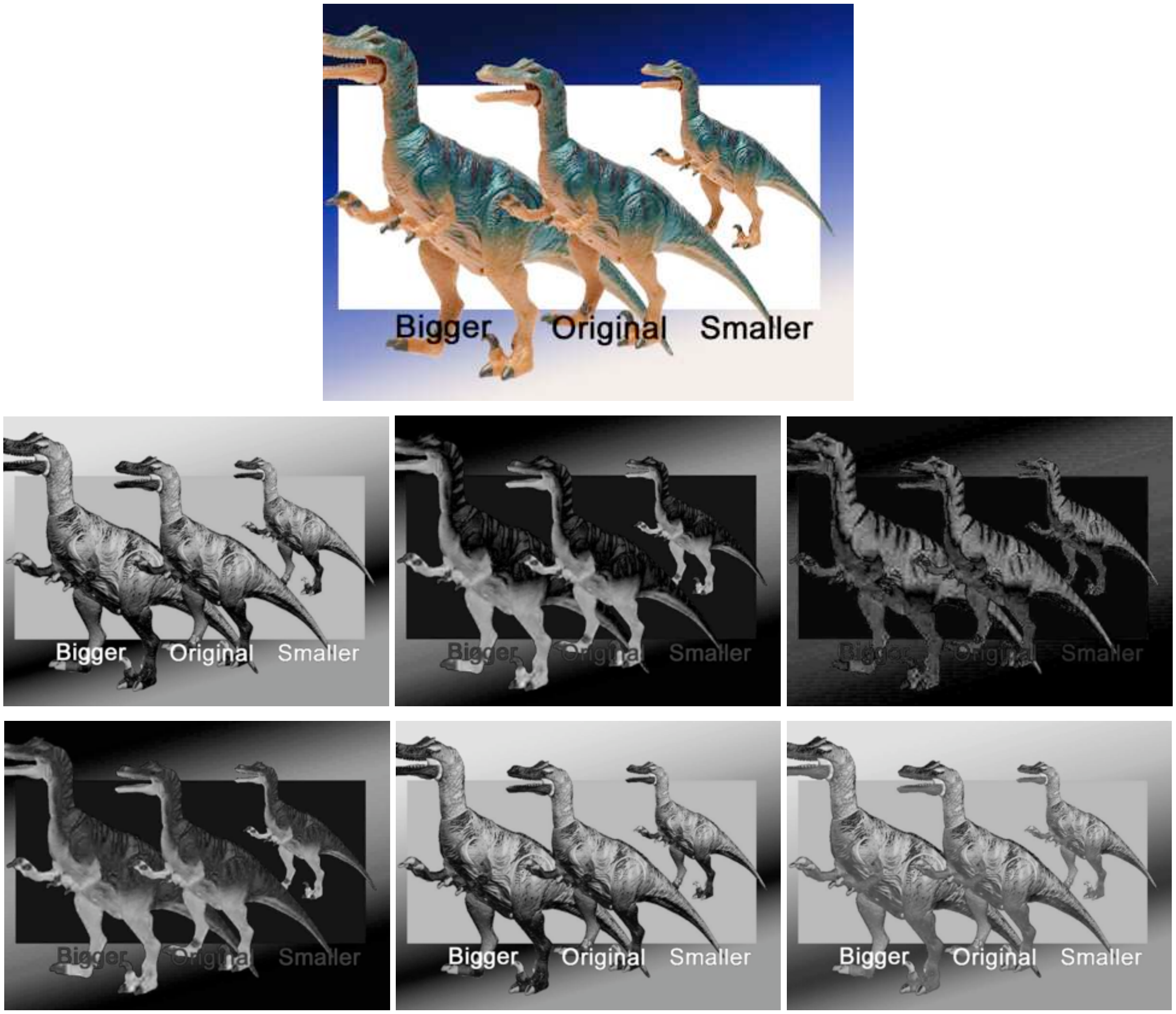


Figure 11. Top row: initial image (top); middle row: PC1, PC2, and PC3 with color based on distance from the plane; bottom row: PC1, PC2, and PC3 with color based on distance from the line normal to the plane.

3.4.1.3 JPEG Artifact Detection with PCA

JPEG artifacts are usually visible when rendering with either the PC1 line or PC3 plane. These artifacts appear as rectangular chunks in the background and distortions around the figures (Figure 12). However, the PC3 plane usually shows artifacts from only one JPEG resave. When two pictures of different qualities are combined, they bring with them different JPEG artifacts. Rendering based on the PC1 line highlights these differences, allowing an observer to identify a spliced image (Figure 13).

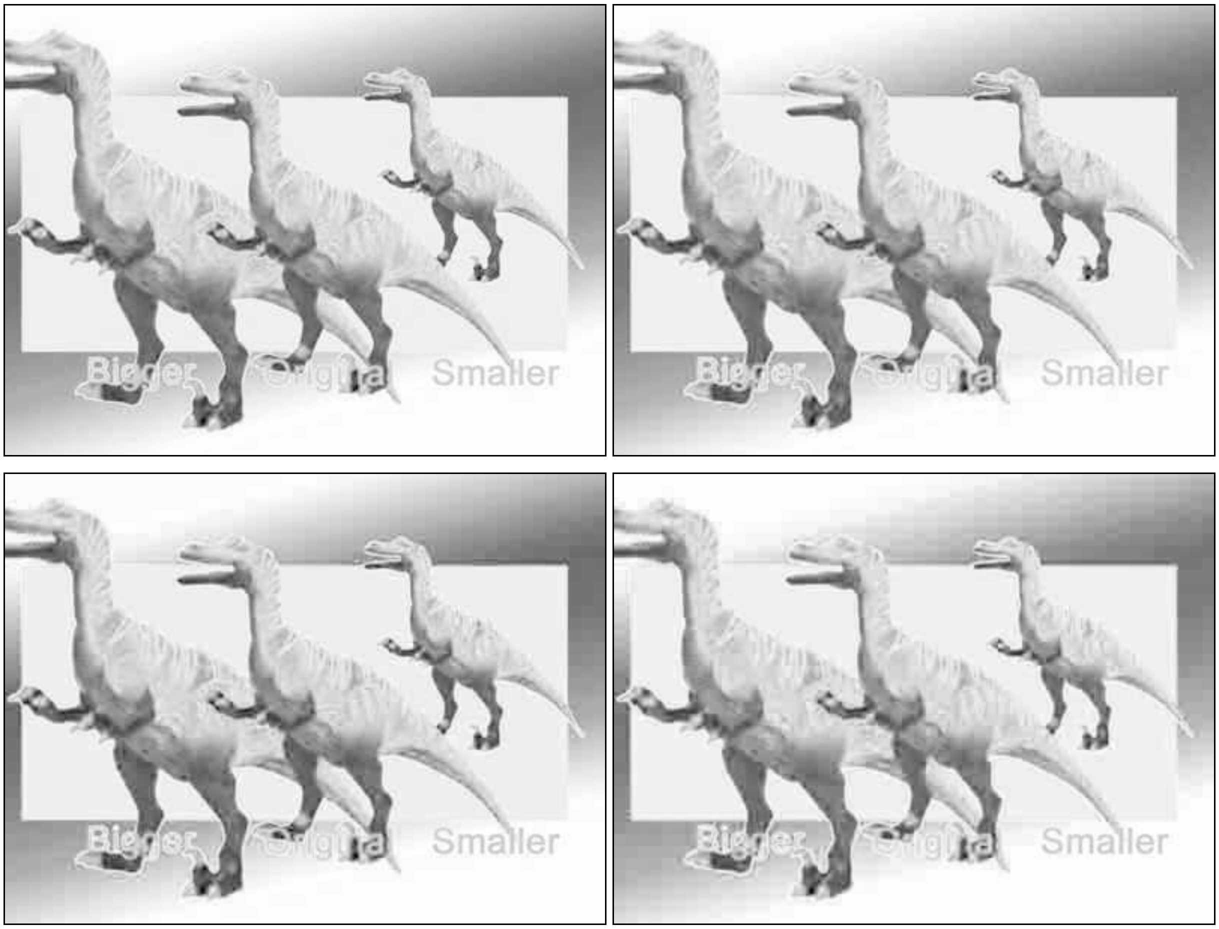


Figure 12. JPEG artifacts shown using the PC1 line. Images are at 90%, 80%, 70%, and 50% and inverted for clarity.

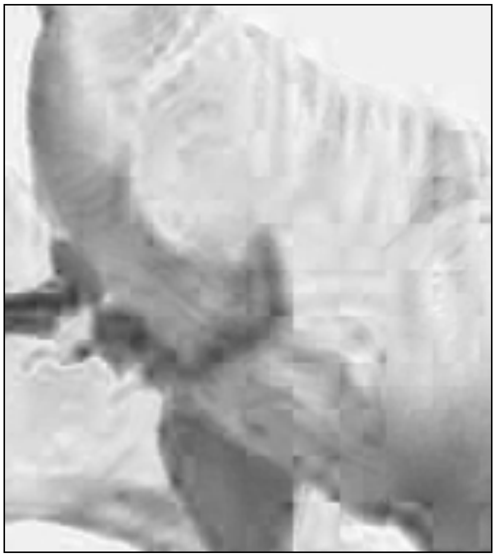


Figure 13. PC1 of a figured created by splicing two images. The image on the left was at 90% and the right was 75%. The blocky JPEG artifacts are more distinct from the 75% image.

3.4.1.4 PCA Example: Moonwalk

As an example, consider the moonwalk picture discussed earlier (Section 3.2.2). The artist stated that the image was created using 3DS MAX and post-processed using Combustion and Photoshop.[18] The quantization matrix matches Photoshop's "high (8)" quality, equivalent to a JPEG saved at 89%. However, using the PC1 line shows a significant number of artifacts that resemble a quality around 40% (Figure 14). This suggests that the image was saved multiple times.

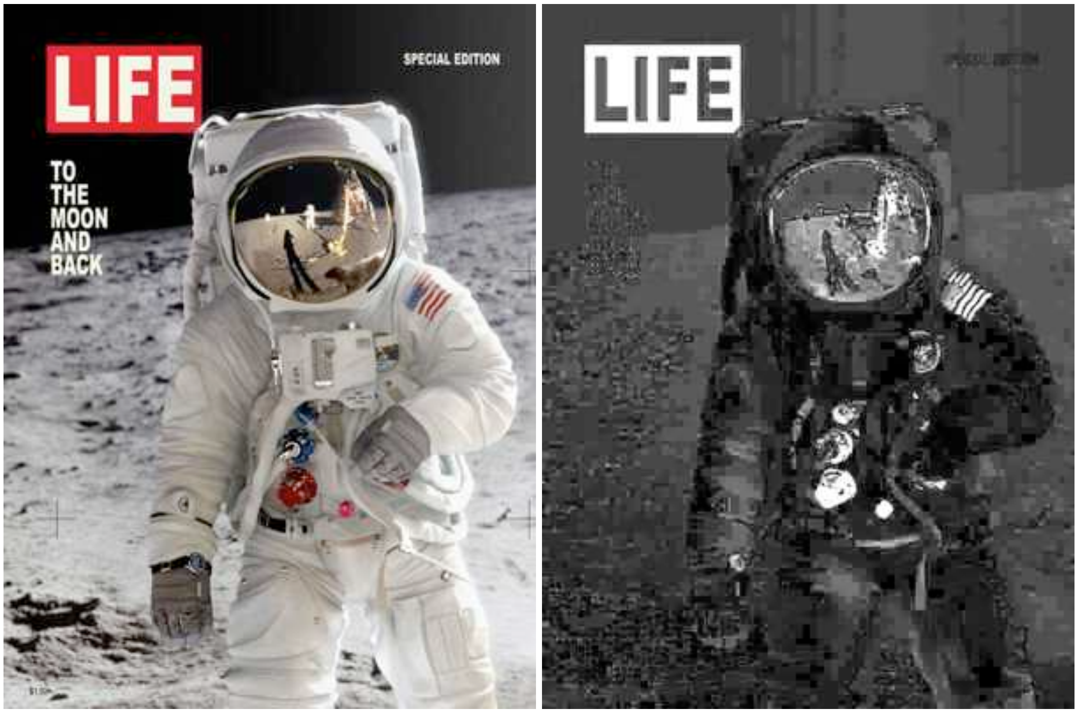


Figure 14. Moonwalk image and PC1.

In addition to the large number of resaves, the spacesuit shows more artifacts than the background and helmet reflection, supporting the artist's description that the background and helmet reflection are bitmaps that were added after the astronaut was rendered. PC1 also identifies the red and blue connectors, red "LIFE" background, and American flag as having the wrong color scheme for this image (white indicates far from the PC1 line). These are areas that were likely enhanced by the artist after the initial rendering.

3.4.2 Error Level Analysis

JPEG is a lossy format, but the amount of error introduced by each resave is not linear. A 90% image resaved at 90% is equivalent to a one-time save of 81%. Similarly, saving an image at 75% and then resaving it at 90% (75%→90%) will generate virtually the same image as 90%→75%, or saved once at 67.5%.[19] The amount of error is limited to the 8x8 cells used by the JPEG algorithm; after roughly 64 resaves, there is virtually no change. However, when an image is modified, the 8x8 cells containing the modification are no longer at the same error level as the rest of the unmodified image.

Error level analysis (ELA) works by intentionally resaving the image at a known quality level, such as 95%, and then computing the difference between the images. If there is virtually no change, then the cell has reached its local minima for error at that quality level. However, if there is a large amount of change, then the pixels are not at their local minima and are effectively "original". Figure 15 shows an original image, the image resaved at 75%, and

[18] http://forums.cgsociety.com/showthread.php?t=323480

[19] Because the JPEG algorithm operates on integers instead of floating point values, the 75%→90% image will be nearly identical to 90%→75%, but may not be a perfect match.

resaved again at 75%. The 95% ELA for each of the three images shows the areas containing original pixels. Nearly all pixels in the original image are not at their local minima. The first resave (75%) shows large areas where the pixels have reached their local minima. The second resave introduces more areas that have reached their local minima for error.

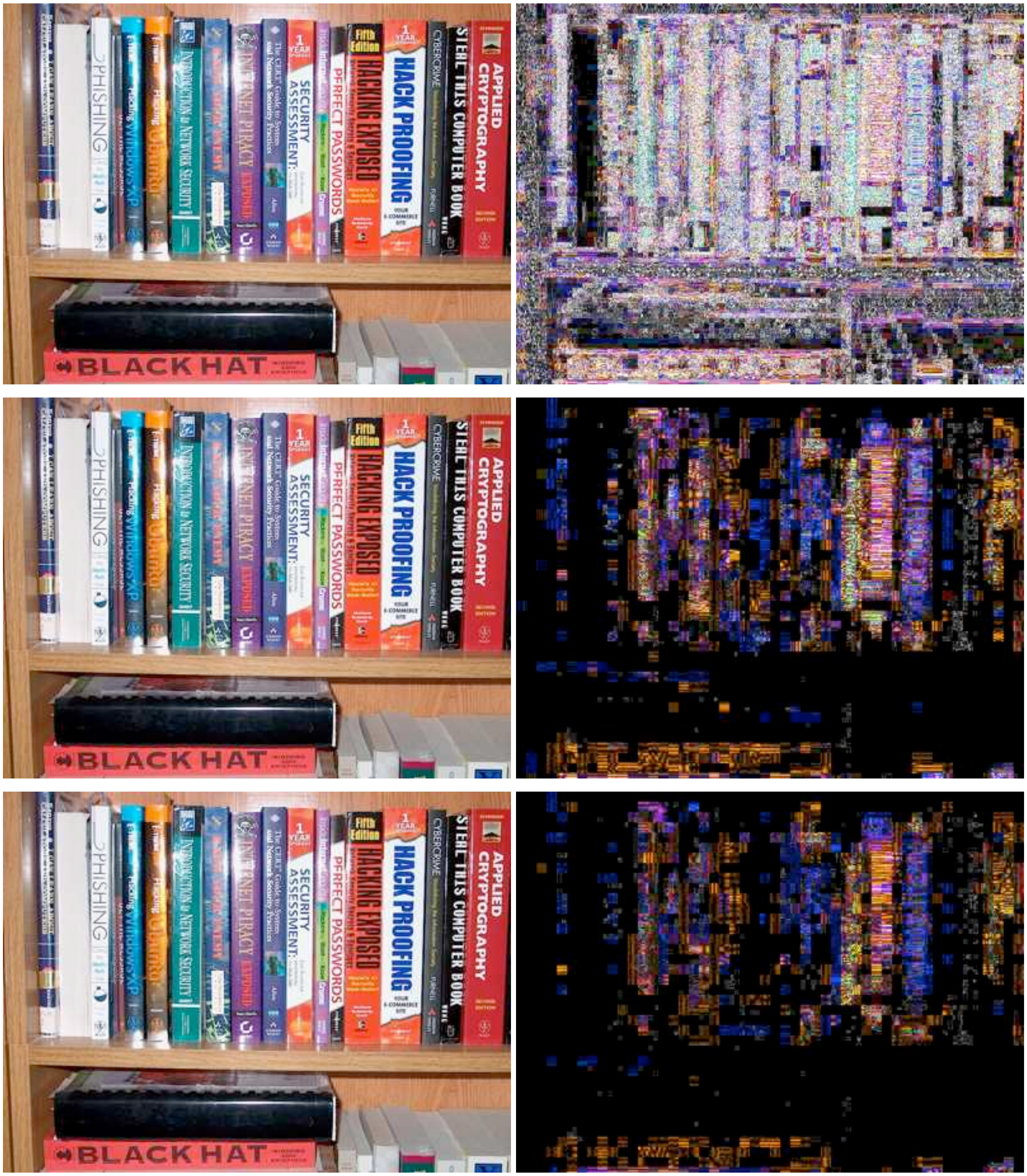


Figure 15. An original photo from a digital camera, resaved at 75%, and then resaved again at 75%. Each picture is associated with the 95% ELA for the image.

Any modification to the picture will alter the image such that stable areas (no additional error) become unstable. Figure 16 shows a modified image using Photoshop. The modified picture was based on the first 75% resave. Books on the shelf were duplicated and a toy dinosaur was added to the shelf. The 95% ELA identifies the changes since they are areas that are no longer at their minimal error level. Additional areas of the picture show slightly more volatility because Photoshop merged information from multiple layers, effectively modifying many of the pixels.

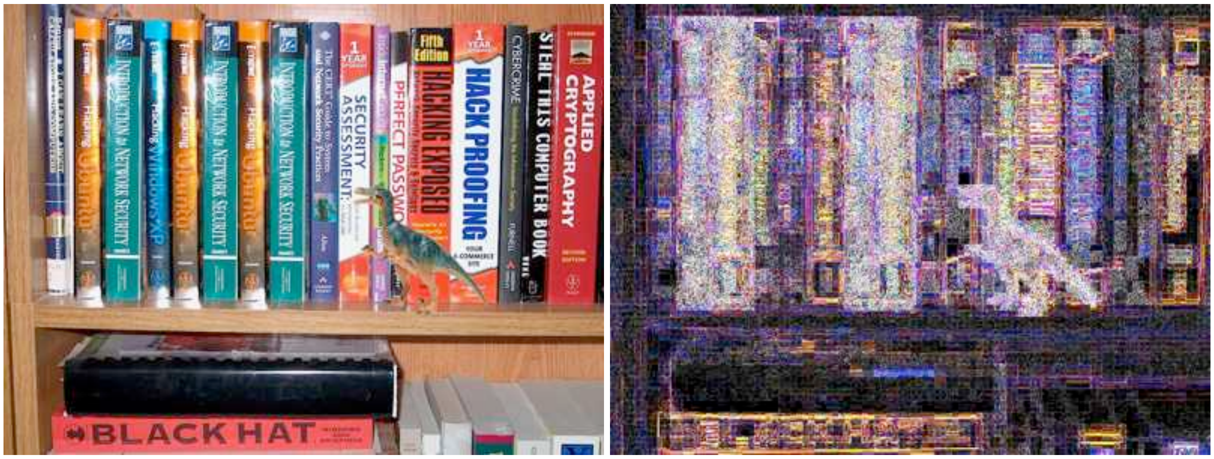


Figure 16. The first resaved image (75%) was modified. The 95% ELA identifies the modified areas: books were copied on the shelf and a dinosaur was added.

3.4.2.1 Example: The Alf Kid

The "Alf Kid" or "Fat Alf Kid" (Figure 17) is arguably one of the most photoshopped people on the Internet. Usually artists alter his shirt or place him in humorous situations (Figure 18). ELA can identify the last modifications made to his image. Ironically, the "original" picture that is used by most artists has been repeatedly resaved and shopped; the last change was the image being cropped – denoted by a high ELA values along the bottom and right margins – and the letters "ALF" being added to his shirt. In actuality, the original photo may not have even had the Alf character on his shirt; large number of resaves has resulted in the loss of that information.

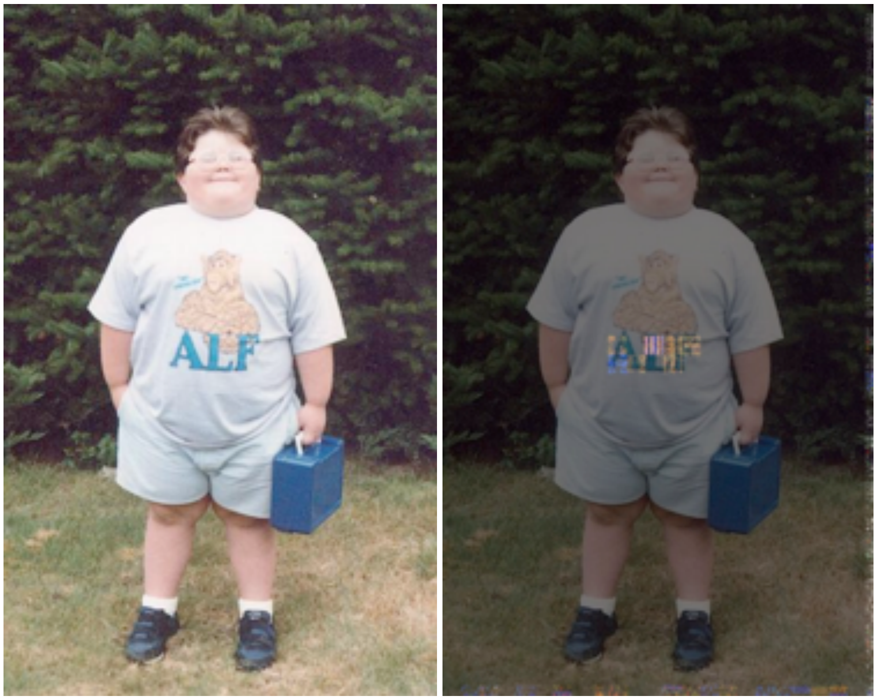


Figure 17. The Alf Kid "original" image and overlay with the 95% ELA.

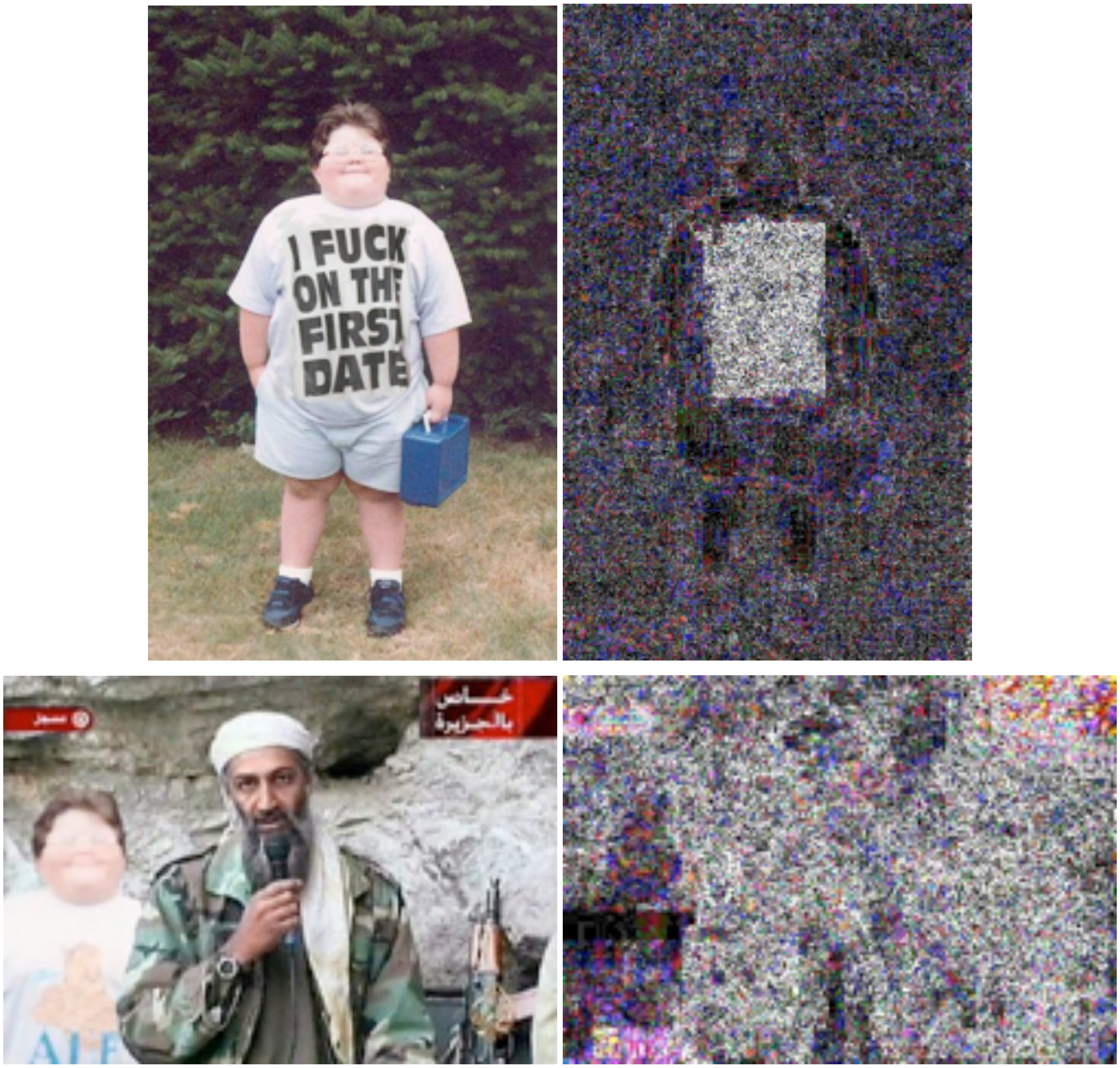


Figure 18. The Alf Kid with his shirt modified and next to Osama bin Laden. The 95% ELA identifies the shirt change and shows that the Alf Kid has a lower error level than the rest of the Osama bin Laden picture.

3.4.2.2 Example: WTC Crash

Shortly after September 11, 2001, a picture surfaced of a tourist standing on the roof of the World Trade Center with an airplane heading for the building (Figure 19). As expected, this image created a firestorm of controversy before being declared a fraud. The 95% ELA identifies the last changes made to this image: the date stamp was added, the United Airlines stripe was placed on the nose of the airplane, and minor modifications were made to the person. Even though the airplane was added to this picture, it has been resaved enough times to obscure that information from ELA.

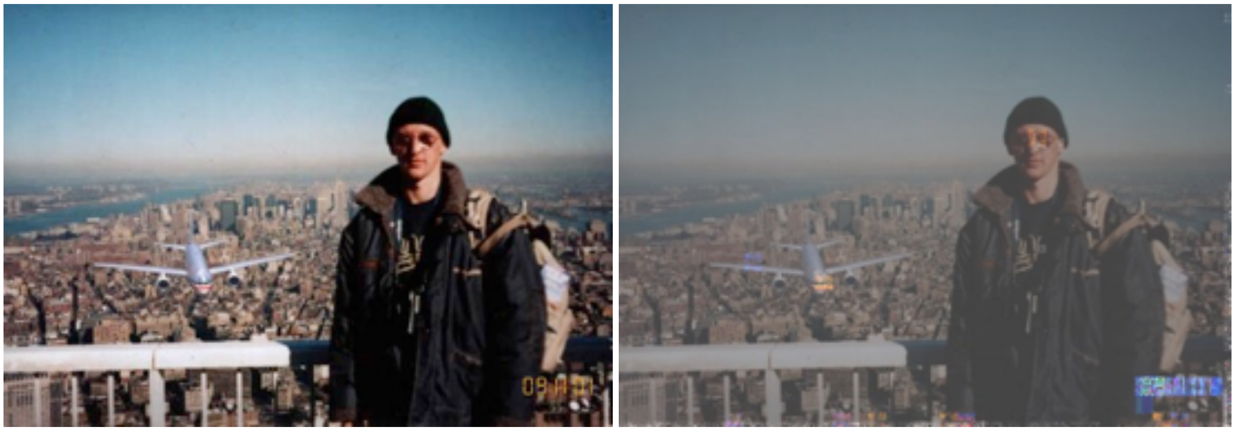


Figure 19. The crash picture and 95% ELA overlay.

3.4.3 Wavelet Transformations

While ELA is useful for identifying recent changes relative to the number of resaves, resaving a picture many times or using a very low quality JPEG can obscure ELA results. However, changes to pictures can still be identified through the use of wavelet transformations.

Wavelets are used for signal decomposition.[20] A single wavelet is a known and well-defined signal. This signal can be scaled and added to itself in order to create more complicated signals. Any real signal can be decomposed into a set of wavelets that, when combined, approximate the signal (Figure 20).

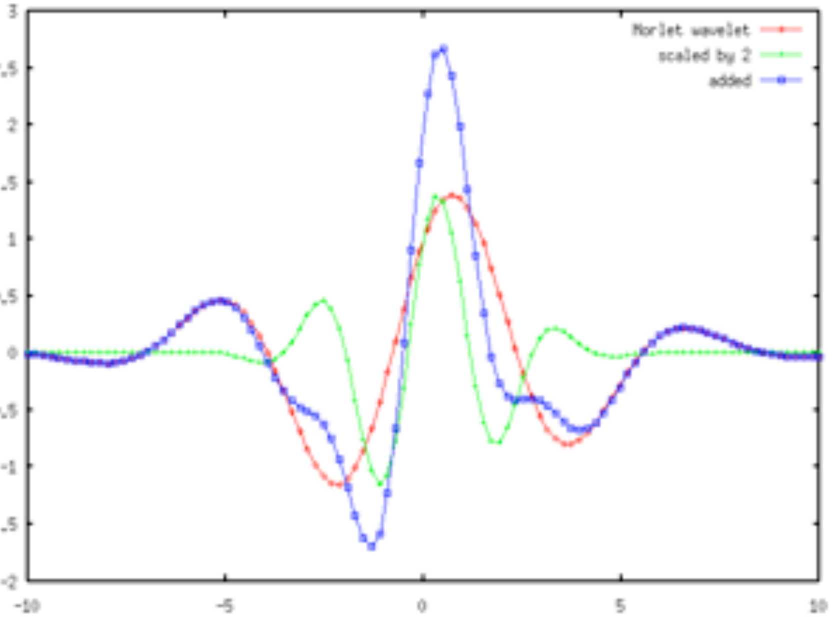


Figure 20. A Morlet wavelet (red), scaled (green), and combined to form a more complex signal (blue).

Although wavelets can approximate any signal, some signal types are more difficult to approximate. Square waves, or areas with sharp color changes, are difficult to approximate. Although the flat area of the square wave can be approximated quickly, the sharp corners may require many wavelets to properly fit the signal. Similarly, linear transitions are approximated by a series of stepped square waves. In addition, extreme values (black and white) are

[20] Amara Graps, "An Introduction to Wavelets". *IEEE Computational Science and Engineering*, Summer 1995, vol. 2, num 2. Available online at <http://www.amara.com/ftpstuff/IEEEwavelet.pdf>.

difficult to approximate. In contrast, wavelets are very good at approximating "natural" colors and noisy images, such as those generated by digital cameras.

In the case of digital photos, the picture is the signal and wavelets approximate the image. Rendering an 800x600 pixel image requires up to 480,000 wavelets per color channel to perfectly recreate the picture. However, if only a small percentage of the wavelets are used, then the main attributes of the picture become visible, even if they are blurry. As more wavelets are included in the rendering, the image sharpens. And even more wavelets fine-tune the sharpened colors.

This property of wavelets – from blurry to sharp to correct colors – can be used to identify image manipulations. In particular, the entire image should sharpen at the same rate. If the picture components are scaled or merged from different focal lengths, then the components will sharpen at different rates.

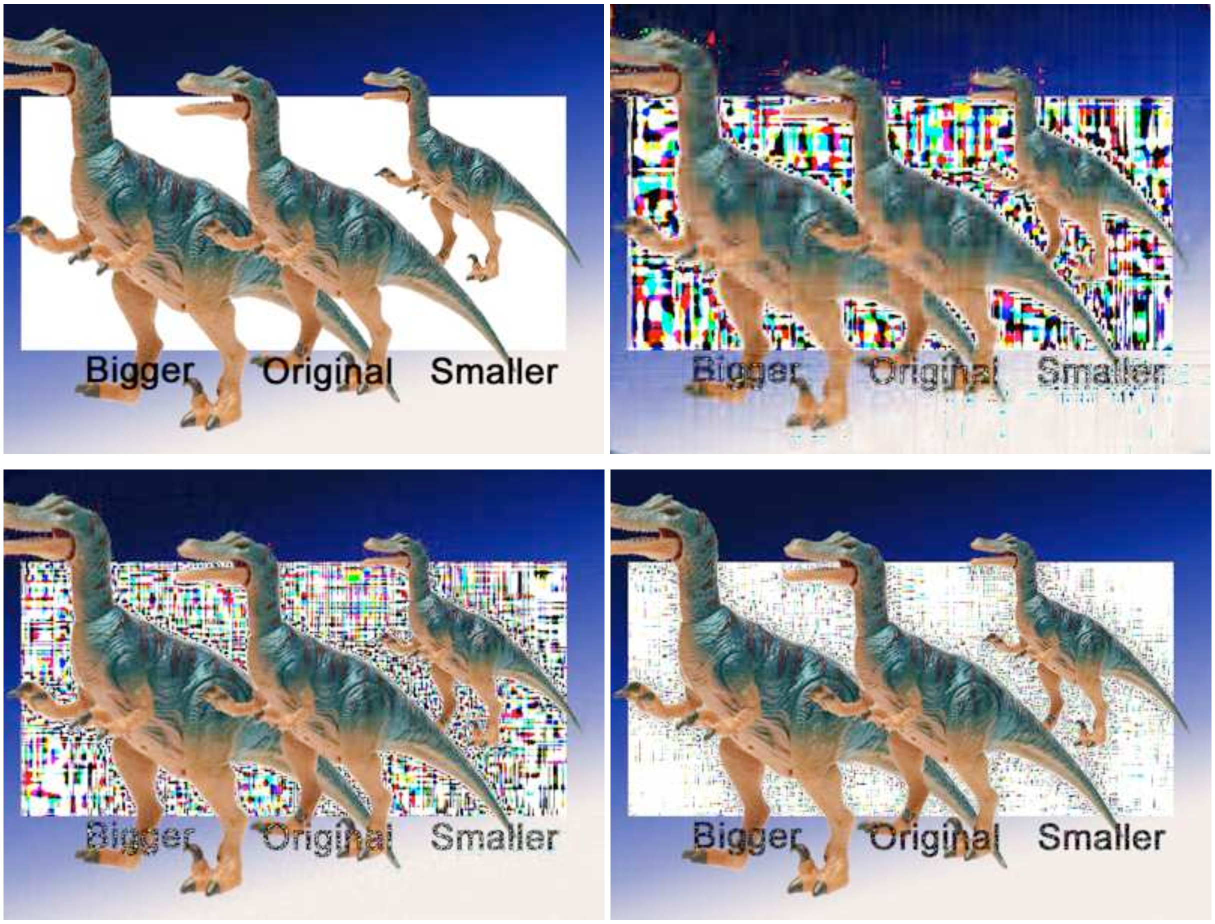


Figure 21. A dinosaur scaled larger and smaller (top-left), then rendered using 1% of available wavelets (top-right), 5% (bottom-left), and 20% (bottom-right).

Consider the dinosaur example in Figure 21. This picture was from one image of a dinosaur.[21] Using Photoshop, the image was scaled bigger and smaller. At 1% of the rendered wavelets, the entire image appears blurry. The extreme colors (white background and black text) are not rendered properly. At 2%, the shoulder and hip on the small dinosaur becomes crisp, while the others remain blurry. At 3%, the small dinosaur becomes crisp – additional wavelets fine-tune the colors but not the sharpness. The original dinosaur becomes sharp at 5%, while the big one becomes crisp at 8%. After 8%, additional wavelets fine-tune the colors but not the sharpness.

[21] http://images.amazon.com/images/P/B00004L8M7.01.LZZZZZZZ.jpg

3.4.3.1 Wavelet Example: Hillary

Celebrities and politicians are frequently photoshopped into fictional situations. Consider Figure 22. This image was created by "redcard" as part of an image manipulation contest.[22] Rendering the image with 5% of the available wavelets shows a crisp torso and near-crisp arms and legs. However, the face remains fuzzy. The fuzziness ends just below the chin. The wavelet analysis suggests that the head is from a picture of Senator Hillary Clinton, the neck and torso comes from a second source, and the arms and legs may be from a third source.

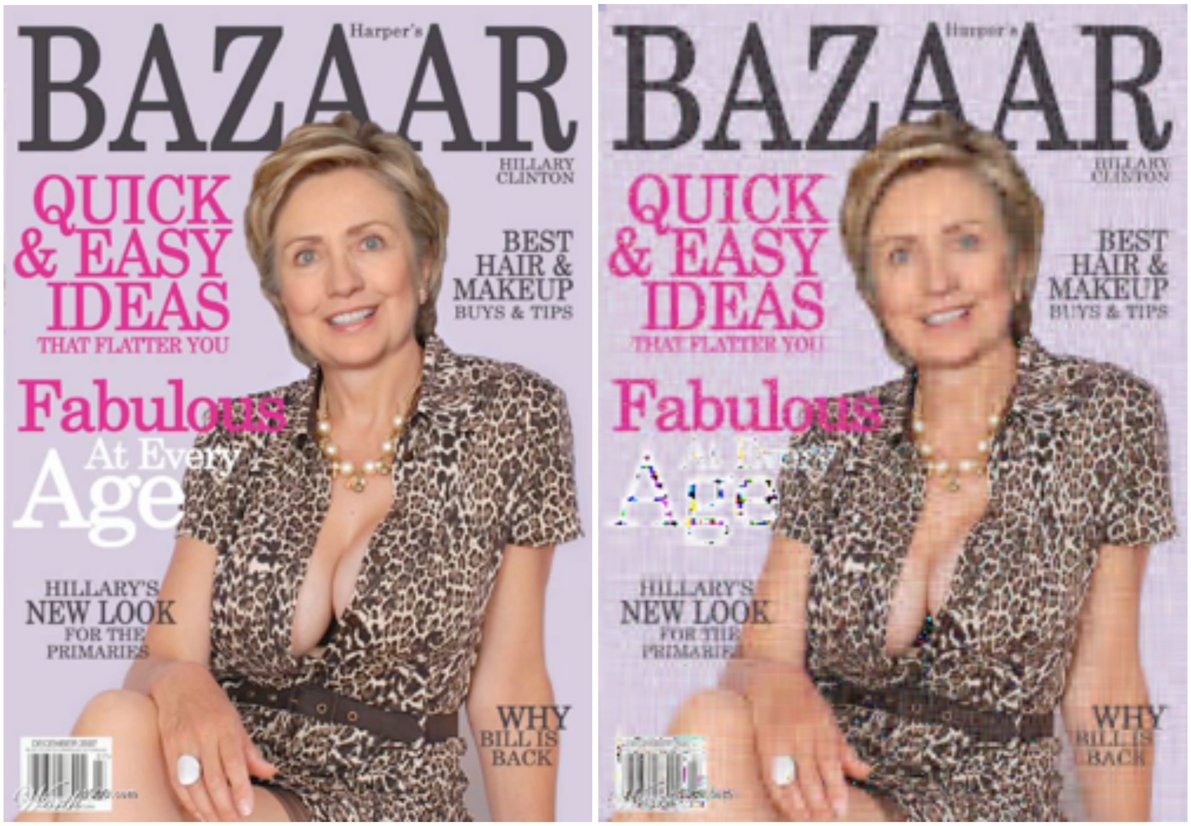


Figure 22. Photoshopped image with Hillary Clinton's head (left) and resolved using 5% of wavelets (right).

3.4.4 Luminance Gradient

Light rarely hits an object with uniform intensity. Instead, sections of the object that are closer to the light source will appear brighter. There are many algorithms for measuring the direction of light intensity, or luminance gradient (LG).[23,24] One of the simplest algorithms identifies light direction based on adjacent pixels. For example (Figure 23), LG can be determined based on a 3x3 pixel square. The computation is done in two steps. First, a circle is drawn around the center pixel. This ensures that diagonal pixels do not assert more influence than the four adjacent pixels. Second, a vector is computed based on the brightest color direction, scaled by the amount of the pixel included in the circle. The result is a set of arrows (vectors) pointing toward the brightest local light source.

[22] http://www.worth1000.com/emailthis.asp?entry=341612

[23] Fattal, R., Lischinski, D.L, and Werman, M., "Gradient Domain High Dynamic Range Compression", Proc. ACM SIGGRAPH 2002. Available online at http://www.cs.huji.ac.il/~danix/hdr/hdrc.pdf.

[24] Johnson, M.K., Farid, H., "Exposing Digital Forgeries in Complex Lighting Environments", IEEE Transactions on Information Forensics and Security, 2(3):450-461, 2007. Available online at http://www.cs.dartmouth.edu/farid/publications/tifs07a.pdf.

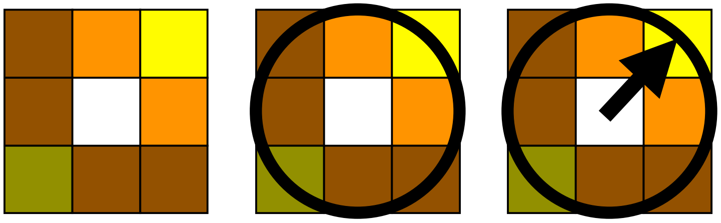


Figure 23. Luminance gradient example using a 3x3 pixel grid. The circle ensures that diagonal colors do not exert too much influence on the light direction.

The example images in Figure 24 illustrate the use of arrows for detecting computer-generated images. Digital capture devices, such as cameras, do not take clean pictures; there is always noise in the image. This noise alters the direction of the arrows. In the real pictures (Figure 24 left and right), the arrows generally point in the same direction, but there is a significant amount of noise that creates random permutations along adjacent arrows. Even the high intensity light source (Figure 24 right) has adjacent arrows without congruency. In contrast, the computer-generated hands (Figure 24 middle) have very uniform arrows along the fingers and smooth waves across the background.

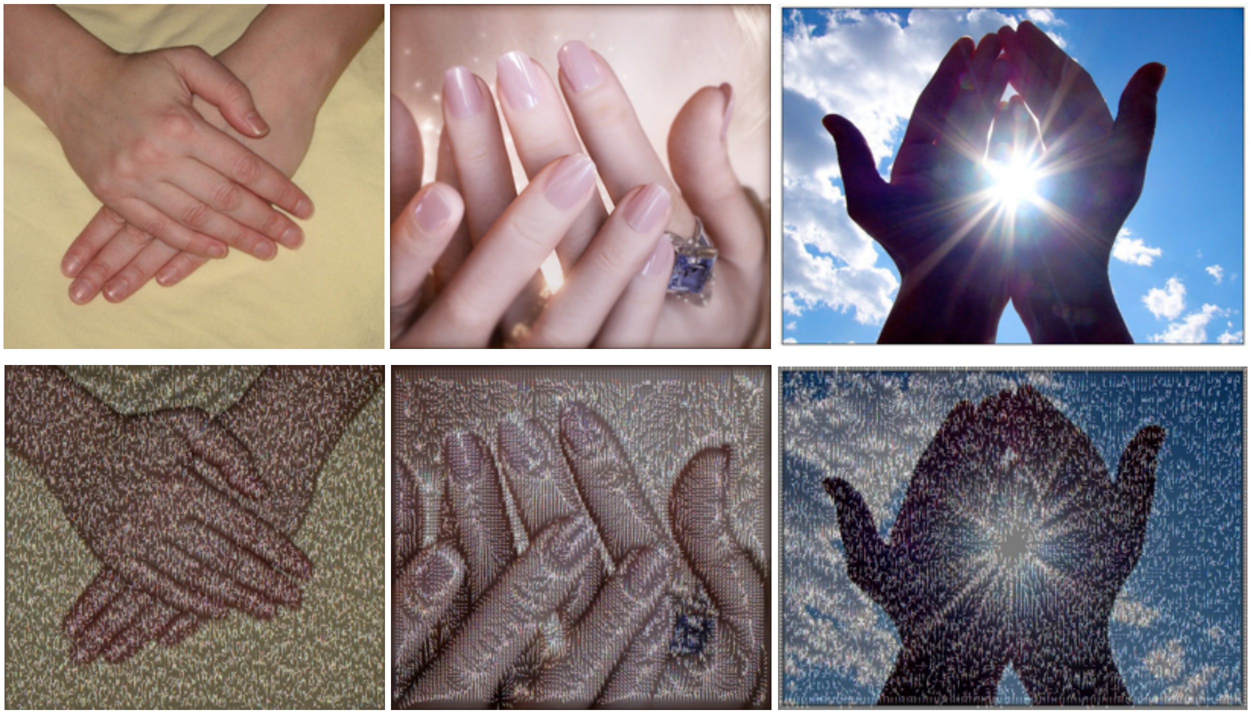


Figure 24. Luminance gradient examples. Left shows a pair of real hands.[25] Middle shows computer graphics.[26] Right shows real hands with a high-intensity and high contrast light source.[27]

While the arrows help illustrate the algorithm, they are undesirable when evaluating every 3x3 pixel square. In particular, arrows on every pixel will bump into adjacent arrows. Instead, the vector can be colorized and mapped

[25] N. Krawetz private collection.

[26] Department of Defense Cyber Crime Center forensic challenge 2007. Image "Her Hands.jpg".

[27] http://i146.photobucket.com/albums/r253/pjbaker_2006/hands.jpg retrieved 17-Jan-2008.

into a single pixel's RGB value. For example, the vector can be decomposed into horizontal and vertical components, and then mapped to red and green. A vector pointing due right has no green (value 0), and left is all green (value 255). Similarly, down has no red and up is all red. The blue channel can be mapped to the vector length – the minimum length is 0 and the maximum is 1 unit, mapped to 255.[28]

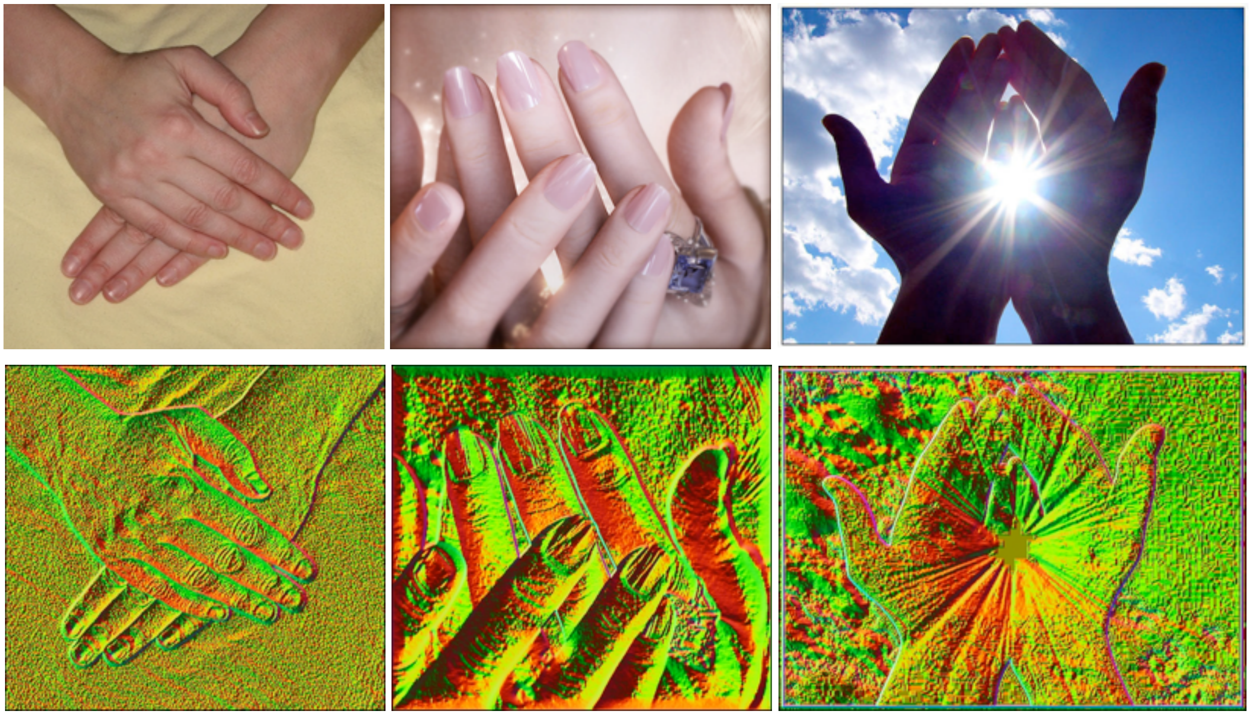


Figure 25. Luminance gradient examples showing colorized vector direction.

Colorized LG vectors better show a distinction between real and digital images (Figure 25). In general, real pictures show noise along uniform colors as a bumpy pattern. Smooth surfaces with even gradient transitions (or no transition) suggest digital manipulation or computer graphics. This algorithm also works as an edge detection algorithm. For example, the lines along the finger edges are noisy – the edges appear jaggy, irregular, and do not have uniform coloring. In contract, lines between computer-generated objects frequently appear sharp with no color permutations.

3.4.4.1 Example: "It's A Trap!"[29]

In January 2006, President George W. Bush visited the NSA headquarters at Fort Meade. A photo of his visit appeared in Newsweek and the Washington Post. Shortly after the photograph became available, a photoshopped version was circulated[30] that inserted Admiral Ackbar from Star Wars (Figure 26). The modified image was exceptionally well made. ELA and PCA analysis identify no abnormalities in the image.

The theory behind luminance gradient is that the vectors point toward the light source. When the vector changes direction, from left-to-right or up-to-down, it creates a high contrast transition in the colorized vector. While this cannot be used to identify the exact location of the light source, it can be used to identify if objects use alternate light sources. For example, each of the people in NSA photo has a dark-to-light colorized LG transition at the middle of their head. This identifies a light source around the middle of the room. Ackbar does not have a clear light transition,

[28] This is an example of one mapping method. There are many variations.

[29] Spoken by Admiral Ackbar in *Star Wars Episode VI: Return of the Jedi.*

[30] http://lists.sans.org/pipermail/list/2006-February/023785.html

but this could be due to his spotty skin coloring. In contrast, the left edge of each person has the same LG coloring, indicating a similar (or same) light source, while Ackbar has a different (darker) left edge. LG indicates that Ackbar has different lighting from the rest of the room. Therefore, LG identifies that Ackbar was likely added to the picture.

Although it is possible for complex set lighting to account for the light difference on Ackbar, the size of the light (his body height) and adjacency to the other people make this very unlikely. Specifically, complex set lighting requires a focus on the light, while the Ackbar image does not show any focused light source.

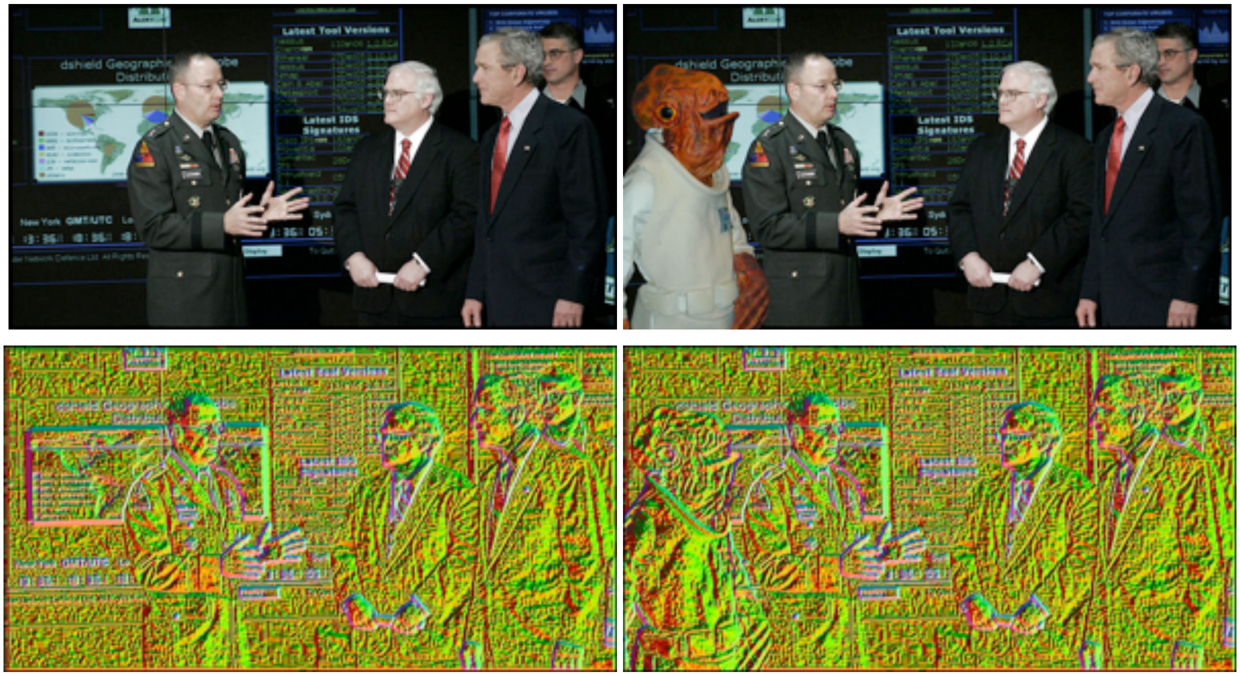


Figure 26. Top row: Washington Post's photograph[31] and photoshopped image[32]. Bottom row: luminance gradient of same images. The dark left side of the added alien indicates an alternate light source.

3.4.4.2 Example: Rachael Ray

In October 2003, celebrity chef Rachael Ray posed for the men's magazine *FHM*.[33] Years later, the photos began to circulate online. One of the pictures was identified as having visual oddities (Figure 27).

Basic observation identifies two significant abnormalities. First, the wall she is leaning against is bent; it bends in at her shoulder and out above her head. Straight objects, such as wall edges, should appear straight and not bent. The second observable abnormality comes from the wallpaper. As seen under the table, the wallpaper is a grid pattern. It contains straight horizontal and vertical lines. However, the wallpaper seen between her arm and her hip shows waviness to the lines, and a clear break in the middle. These two abnormalities identify likely digital manipulation, but do not identify the cause.

Error level analysis shows that the image has been resaved multiple times; most of the image does not change during a resave. However, there are a few modified areas. Specifically, the last thing added to the picture was the photographer's URL (bottom of the image). The FHM logo was either added at the same time, or one save prior.

[31] http://images.insecure.org/nmap/images/wash-post-nsa.jpg

[32] http://img378.imageshack.us/img378/5967/itsatrap7dz.jpg

[33] Wikipedia contributors, "Rachael Ray," *Wikipedia, The Free Encyclopedia*, http://en.wikipedia.org/w/index.php?title=Rachael_Ray&oldid=185268708 (accessed January 19, 2008).

The edging on her clothing appears to have been edited prior to the FHM logo, and some of the apples on the wall indicate manipulation.

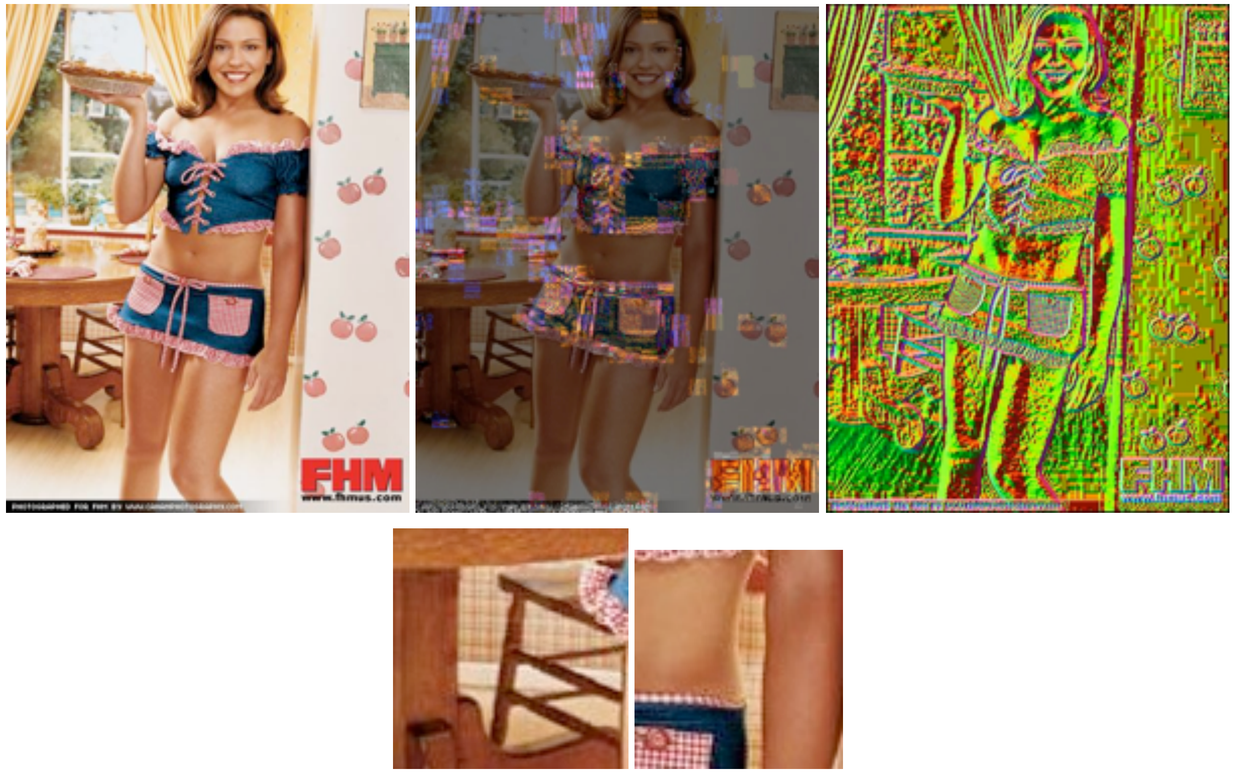


Figure 27. (Left) Rachael Ray from FHM[34], (middle) error level analysis, and (right) luminance gradient. The bottom row shows a close-up of the grid-pattern on the wallpaper.

However, the full extent of the edits is identifiable using luminance gradient:

- Her face, arms, and legs show a sharp LG transition. However, the transitions are not in the same place. Her head and straight-arm shows the light in the middle, but her bent arm (holding the pie) shows the light to the side. Her legs show the light on the opposite side compared to her pie arm. This implies very complex lighting, a "Frankenstein" image (person made from parts), or digital manipulation.
- In contrast to her face and limbs, her upper chest and belly do not show sharp color transitions. The LG transition is very smooth and linear, identifying digital manipulation.
- LG shows a bumpy noise pattern on the background window and her legs, but not on her arms, face, or upper chest. Her belly shows noise in the middle but not along the sides. The implication is that someone thinned her waistline (removed her love handles), and smoothed out parts of her body. However, the background and her legs appear to be original.
- The wall with the apples is missing noise, except immediately adjacent to the apples. It appears that someone took a tool, such as Photoshop's "magic wand" selection tool, selected the wall, and made it brighter – smoothing out the noise in the process. The reason ELA identifies modification around the apples is likely because everything around the apples was modified.

The image of Rachael Ray has been modified. The observed, ELA, and LG results combine to identify the full extent and method. It is probable that a tool such as Photoshop's "Liquify" was used to make Rachael Ray appear

[34] http://i78.photobucket.com/albums/j117/celebritymound/73974_rachel5_501lo.jpg

thinner. This tool altered the lighting on her waist, arms, and head. It also bent the wall and distorted the grid pattern on the wallpaper. The white part of the apple-wall was also brightened up, possibly by using the "magic wand" selection tool.

Following an initial analysis of this image, the photographer addressed the general claims of image manipulation[35]:

> 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

It is important to recognize that other images of Rachael Ray by the same photographer do not show the same degree of digital manipulation.

4 Analysis Limitations

The methods discussed in this paper provide a set of very powerful tools for evaluating images. However, they cannot be applied to all pictures and the results may not be conclusive. For example, very small images can invalidate some approaches. Wavelets require images that are at least a few hundred pixels in each direction (bigger is better), and image scaling can fool ELA and PCA.

Low quality images, either due to a very low JPEG quality setting or from color reduction (e.g., GIF) can impede most of the analysis techniques. In addition, media transitions may influence images. Scanning a photo, converting a printed magazine picture to JPEG, or capturing a TV signal can introduce artifacts from the conversion process.

More complicated analysis systems are dependent on image contents. For example, a picture with a sharp contrast or well-defined pattern could confound ELA and wavelet results. Wavelets may also lead to harmonic convergences, where a section of a picture may come into focus at a different rate from the rest of the image (see Section 5).

Images with a significant amount of recoloring (brightening, pallet skew, etc.) can also lead to false identification. Recoloring can appear indistinguishable from a modified image (ELA), imply multiple layers (wavelets), or look similar to complex lighting (LG).

On occasion, a skilled artist can create photo-realistic images that pass all of these evaluation methods. However, artists with this level of talent are extremely rare. Although photo-realistic tools are available, the image quality is highly dependent on the artist. In the more general case, an enhanced or CG image may pass one or more of these evaluation methods, but is unlikely to pass all of them.

It is important to recognize that the methods described in this paper only highlight aspects of the image. The evaluation is not automated and is heavily based on the observer. If the analyst does not recognize a particular artifact, then the evaluation may be inaccurate.

4.1 Defining Terms

Digital image analysis and forensics is a relatively new field. As such, definitions are not standardized. For example, what constitutes a computer graphics image? Does cropping or resizing an image count as a digital enhancement? Different organizations define these terms differently, and conflicting definitions may lead to incorrect categorization. A few examples:

- According the Defense's Cyber Crime Center (DC3), as long as the primary subject is real, the image is considered real. This is a useful definition for cases involving child pornography. In their forensic challenges[36], the DC3 is only concerned with "real" or "CG" and not levels based on digital enhancements.
- PhotoPortfolios.net considers an image to be "real" if the image contains nothing more than cropping or "minor tonal/contrast adjustments, cloning removal of dust spots and sharpening, [and] no post-shutter digital enhancement".[37] While some photography sites permit specular reflection removal and glare reduction, others sites consider even minor adjustments, such as red-eye removal, to be taboo.

[35] http://www.hackerfactor.com/blog/index.php?/archives/102-Rachael-Ray-enstein.html#comments

[36] http://dc3.mil/challenge/

[37] http://www.photoportfolios.net/menus/imageterms.html

- Currently, there are no legal definitions for "real" or "computer graphics". Instead, courts leave the definition to subject experts.[38]

Hacker Factor Solutions uses a four-level hierarchy:

- **Real**. An image is considered "real" if it contains no digital enhancements. Minimal scaling and cropping is permitted if it does not distort the image.
- **Digitally enhanced**. Any image that started as real but was digitally modified.
- **Compute graphics**. A CG image is any picture that did not start with a real image.
- **CG enhanced**. The use of post-rendering enhancements to a CG image.

This four-level definition is comparable to other definitions and can be applied to entire pictures or portions of pictures.

4.2 Analysis Accuracy

In 2007, the Department of Defense Cyber Crime Center provided 51 images for a blind evaluation. Of the 51 images, 25 were known to be real, 20 were CG, and 6 were unspecified. The analysis correctly classified 39 of the 45 known images, for an overall accuracy rate of 86.67%. However, the accuracy rate is a little misleading. In particular, there were no false-positives, where a real picture was classified as CG. Four of the six false-negatives (where a CG image was classified as real) were by award winners from the CG Society, indicating skill far above and beyond the average user. In addition, two of the false-negatives were by the same artist who spent more than two years working on the same model.[39]

		Analysis Determination		
		Real	**CG**	**Percent**
Observed Category	**Real**	19	0	100.00%
	CG	6	20	76.92%
	Percent	76.00%	100.00%	86.67%

All six of the unknown images were classified as CG, digitally enhanced, or CG enhanced. None were real.

Finally, six images, including four of the six false-negatives, were identified as "difficult" images to analyze. Each was identified as having possible CG elements, but the elements could be attributed to lighting, image quality, or other aspects. These classifications were considered "low confidence". All six of these images were actually CG, but only two were classified as CG.

5 Example: Soldier Picture

Modified images frequently appear in online forums, television, movies, and advertisements without much concern. However, there is a stigma when they appear in mass media and news outlets. Since news sources are supposed to report facts, enhanced or modified images could easily misrepresent real situations or mislead readers.

In 2003, photographer Brian Walski submitted a photo of a British soldier in Basra. This picture was identified as a fake after editors noticed duplication in the crowd of people – leading to Walski's dismissal from the *Los Angeles Times*.[40] The picture is widely believed to be a combination from two other Walksi photos (Figure 28).

[38] Example: http://caselaw.lp.findlaw.com/scripts/getcase.pl?navby=search&case=/data2/circs/1st/031741.html

[39] http://www.infinite-realities.net/, http://infinite.cgsociety.org/about/, and http://infinite.cgsociety.org/gallery/458968/

[40] http://blog.wired.com/wiredphotos54/2007/05/double_vision_i.html

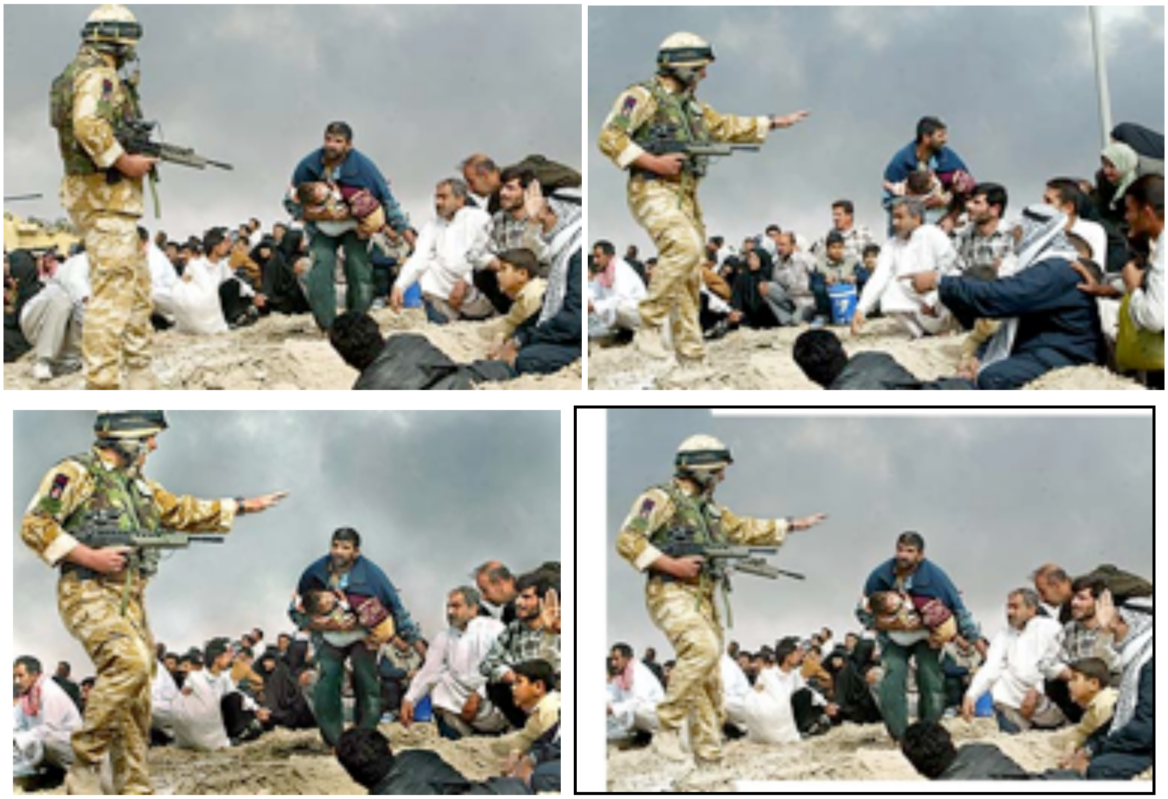


Figure 28. The first and second originals images (top) were combined to form the forgery (bottom-left). Bottom-right shows an attempt to recreate the forgery by overlaying the originals using Gimp.

The two original pictures can be combined to recreate the forgery and identify how it was likely created. A basic observation of the images suggests that the majority of the forgery comes from the first original – everything right of the soldier was kept. The second original picture was cut, leaving only the solder and people to his left. The second original was scaled larger and combined with the first image. Finally, the combination was significantly recolored and the sky was modified.

However, original sources are usually unknown. Using the techniques covered in this paper, the faked image can be analyzed without the assistance of the original images (Figure 29).

- **Observation**. In the combined picture, the squatting man in white (lower left, behind soldier's leg) is duplicated in front of the soldier's knee. However, the copies are not identical; one or both appear to have been scaled.
- **Meta Information**. The JPEG does not contain camera information. However, it does include Adobe-specific meta data, suggesting that it was edited using an Adobe tool.
- **Quantization Table Analysis**. The quantization matrix is unknown, but the quality is approximately 72%.
- **ELA**. The error level analysis shows that the sky is at one error level, but all of the people are at a different error level. This is actually due to the image recoloring, but ELA only identifies "a" change.
- **PCA**. PC1 shows that the sky has been resaved many times. The soldier and people to his left have been resaved a few times, as denoted by a fuzzy artifact halo. Other people in the crowd were not resaved as many times. PC1 summarizes the image into three regions: sky, soldier and people to his left, and people to his right.
- **Wavelets**. At 8% of the wavelets, the soldier appears crisp, but with many areas of sharp color transitions. This suggests that his image was resized to fit the picture. The man in the plaid shirt (right edge) has a

crisp face; it resolved at around 2% of the available wavelets. However, the man in white behind him and the child in front of him both have blurry faces, suggesting three different layers. The full wavelet analysis suggests that the people may form up to nine different layers.

By evaluating the ELA, PCA, observed duplication, and different percentages of rendered wavelets, it is clear that the image has been manipulated. However, the details of the manipulation are inconclusive. The significant recoloring, small size, and low resolution increase false-positive results in the analysis. If these issues are ignored, then the picture appears to include eleven or more distinct layers that were combined to form the image. However, this image has been significantly modified. The analysis shows that it may contain as few as three layers: soldier and people to his left, ground and people to the soldier's right, and the sky.

The largest inconsistency in the analysis comes from the PCA and wavelet analysis. (For other pictures, other analysis methods may yield inconsistent results.) In this case, wavelets show that the man in the plaid shirt resolves quickly. This is due to a frequency harmonic and not due to manipulation. In particular, neither of the original images shows an irregularity with the man in plaid, but the recreation attempt (Figure 29) does show the same wavelet characteristics.

Cynthia Baron offered the following suggestion for identifying harmonic convergence regions: Wavelet analysis uses symmetrical image sizes (e.g., 512x512 or 1024x1024) and processes one direction before the other (e.g., horizontal before vertical). Rotating the image before performing the wavelet analysis may shift or remove the location of any harmonic convergences. In this case, rotating the forged image removed the crispness from the plaid man's shirt, and reduced the crispness in his face.

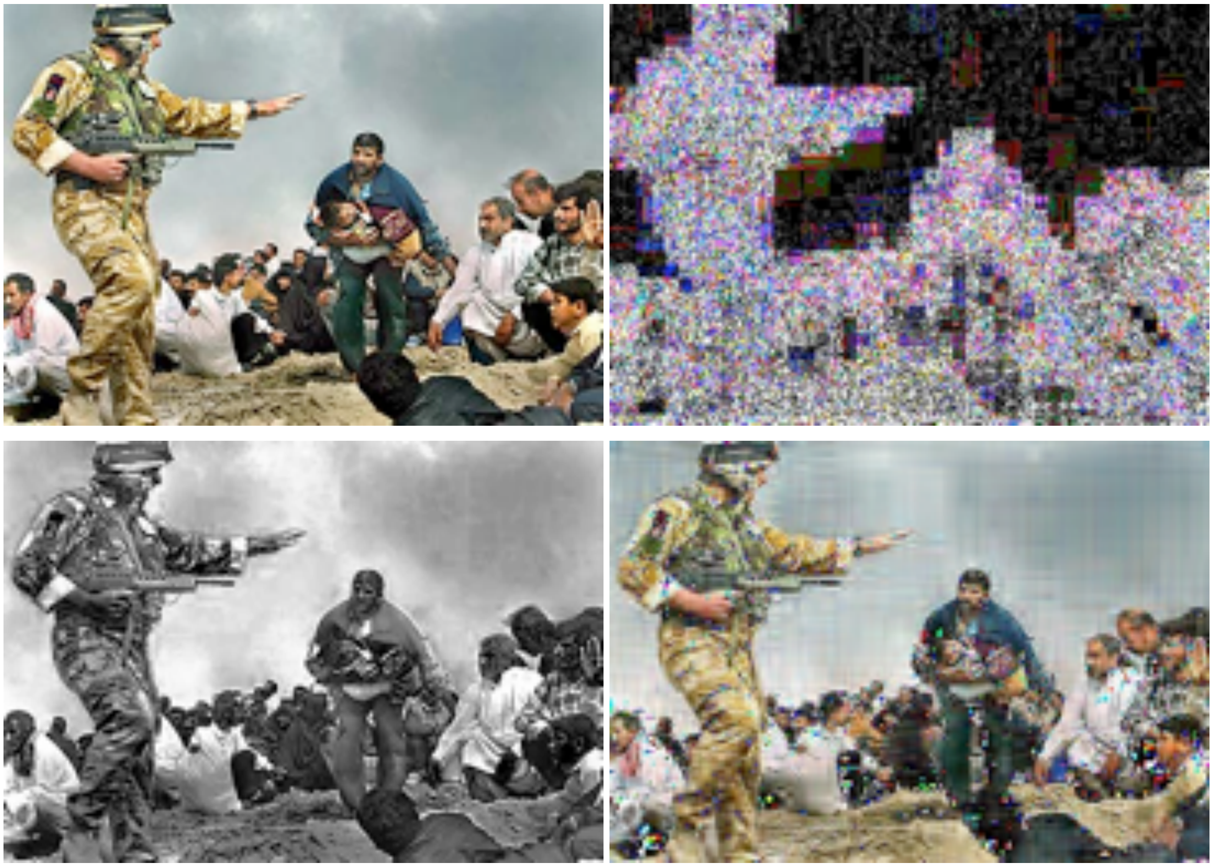


Figure 29. Modified picture from Brian Walski (top-left), 95% ELA (top-right), contrast-enhanced PCA (bottom-left), and rendered using 8% of available wavelets.

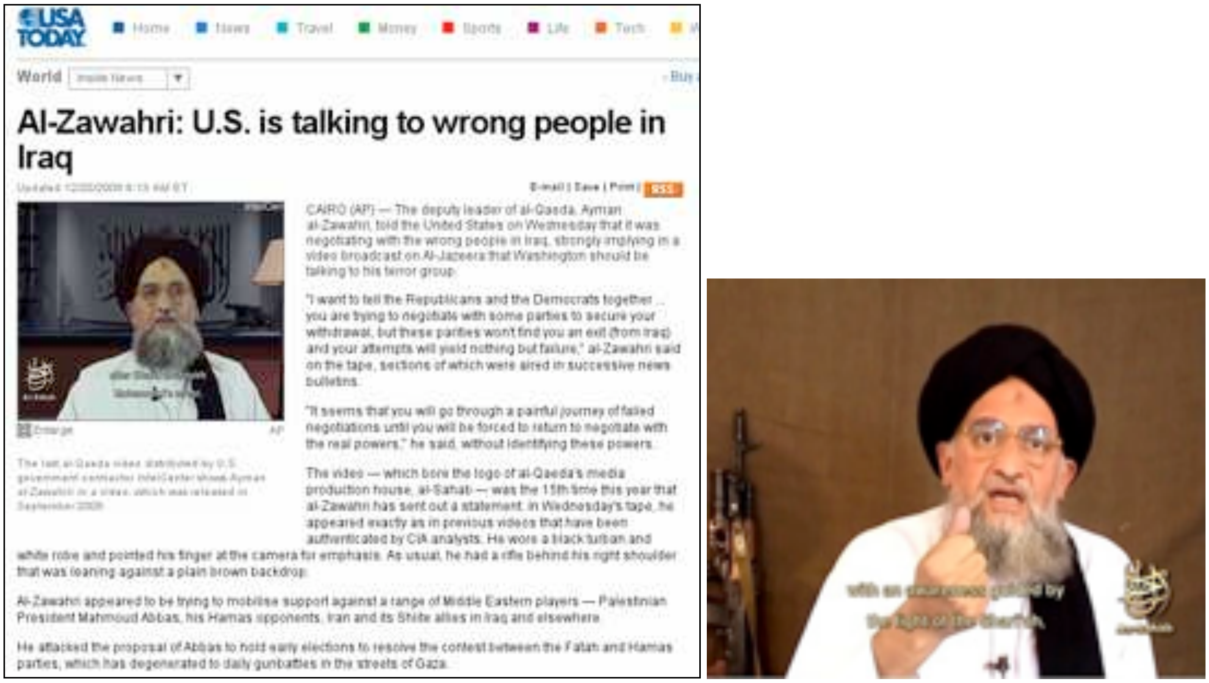
6 Case Study: As-Sahab and Al Qaeda

As-Sahab is the media arm of Al Qaeda. They periodically release videos designed to remind people of Al Qaeda's existence, issue threats, recruit members, and potentially act as a covert channel for triggering terrorist cells. Many of the videos include evidence of manipulation and tampering. These include videos featuring the Al Qaeda leaders Ayman al Zawahiri, Azzam al-Amriki, and Osama bin Laden.

6.1 Example: Dr. Ayman al Zawahiri

On 20-Dec-2006, Dr. Ayman al Zawahiri (#2 guy in Al Qaeda) released a video. *USA Today* covered the video release with a headline story (Figure 30).[41] *USA Today*'s description of the video says, "He wore a black turban and white robe ... he had a rifle behind his right shoulder that was leaning against a plain brown backdrop." While this is a valid description of the As-Sahab video[42], the picture used by *USA Today* did not show that image. Instead, *USA Today* used a picture from another video, dated 28-Sept-2006. This is an example of a mislabeled image.

The picture that *USA Today* chose to use with the story includes many other interesting features. First, it came from the IntelCenter (www.intelcenter.com) – an organization that tracks terrorist activities. The IntelCenter placed their logo in the top-right corner of the video. However, the company name is clearly cropped – likely by *USA Today*. A comparison with the same frame from the actual video shows many other observable differences. In particular, the IntelCenter adjusted the color and sharpness of the picture (Figure 31).



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

Updated 12/20/2006 8:13 AM ET

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) — The deputy leader of al-Qaeda, Ayman al-Zawahri, told the United States on Wednesday that it was negotiating with the wrong people in Iraq, strongly implying in a video broadcast on Al-Jazeera that Washington should be talking to his terror group.

"I want to tell the Republicans and the Democrats together ... you are trying to negotiate with some parties to secure your withdrawal, but these parties won't find you an exit (from Iraq) and your attempts will yield nothing but failure," al-Zawahri said on the tape, sections of which were aired in successive news 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.

Al-Zawahri appeared to be trying to mobilise support against a range of Middle Eastern players — Palestinian President Mahmoud Abbas, his Hamas opponents, Iran and its Shiite allies in Iraq and elsewhere.

He attacked the proposal of Abbas to hold early elections to resolve the contest between the Fatah and Hamas parties, which has degenerated to daily gunbattles in the streets of Gaza.

Figure 30. ***USA Today*** **announcing the video release with wrong picture and a frame from the correct video.**

[41] http://www.usatoday.com/news/world/2006-12-20-al-qaeda-palestinians_x.htm

[42] http://www.archive.org/details/Conflict-Between-Islam-and-Unbelief

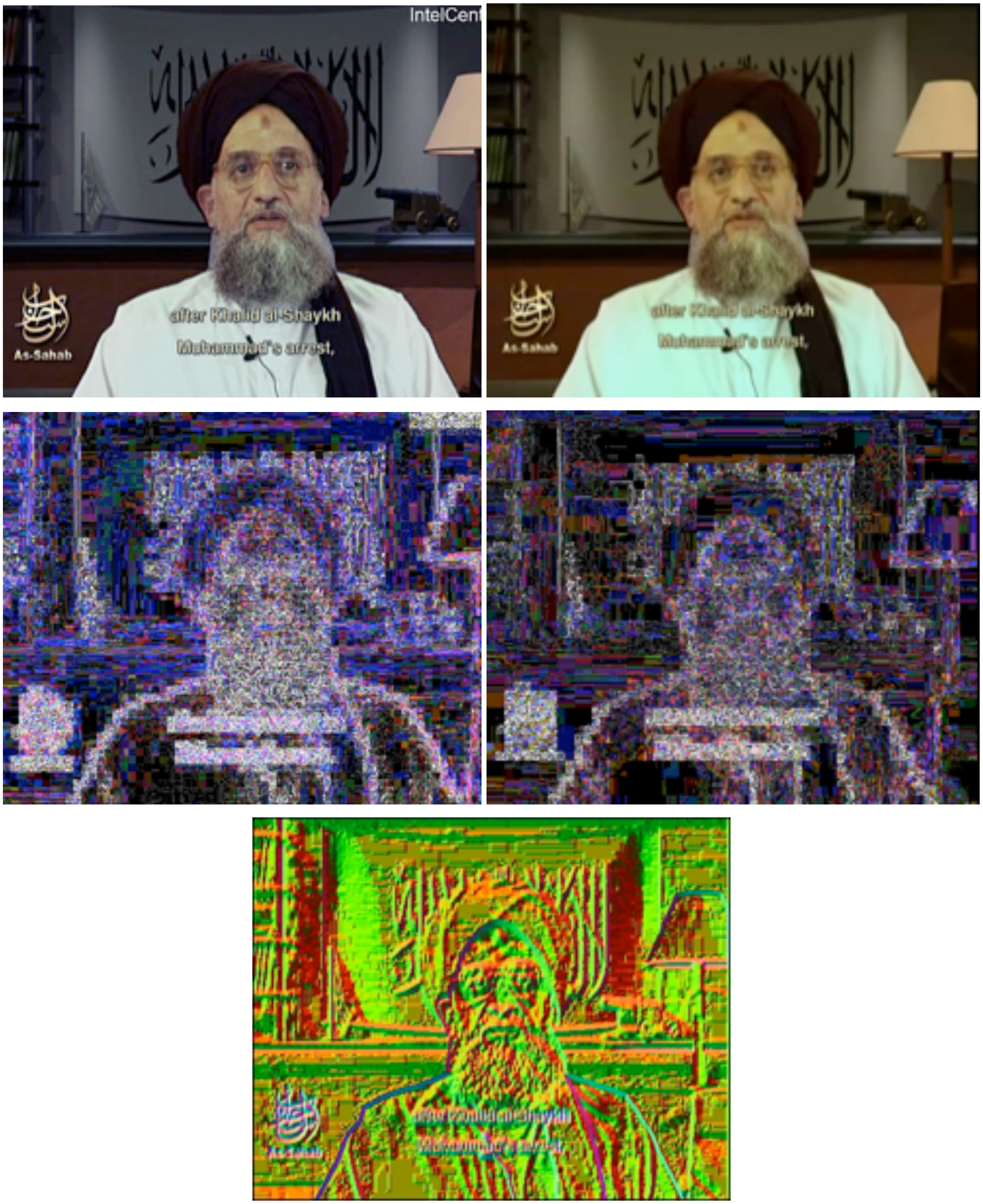


Figure 31. The image provided by IntelCenter and presented at *USA Today* (top-left) and the actual frame from the video (top-right). Middle row shows the 95% ELA for each image, and bottom row shows LG.

Observation combined with the 95% ELA shows the order of the modifications. The changes, starting with the most recent and working back:

- **Cropped**. The last change was the image being cropped.

- **IntelCenter**. Prior to the cropping, the IntelCenter added their logo.
- **Recolored**. The image was recolored and sharpened, modifying the error rate along the image.
- **As-Sahab**. Prior to the recoloring and IntelCenter logo, the As-Sahab logo and subtitle text were added.
- **Al Zawahiri**. Zawahiri was added to the picture. ELA clearly shows a crisp error level change between him and the background. This type of error-level halo is common for chroma-key (green screen) images. In particular, there is a distinctive error level feature generated by chroma-key replacements: since most chroma-key replacement algorithms are based on hue and not saturation or brightness, the different color channels along the seam are at different error levels.
- **Banner**. Just as text can be added to a blank sign (Section 2.2), someone appears to have added the text to the banner behind Zawahiri.

The PC1 and wavelet analysis for this image also supports these findings. The background office appears to be one layer that was saved multiple times. The text of the banner was likely added around the same time Zawahiri was added, and the As-Sahab logo and subtitles were added last.

6.1.1 Recreating the Desk

Following the initial release of this document in August 2007, multiple people suggested that the background was computer generated, possibly using 3D Studio Max[43]. This observation is supported by the LG analysis (Figure 31). Specifically, LG shows very little noise in the background, straight edges along the desk and bookshelf, and very sharp color transitions for the canned lighting over the desk.

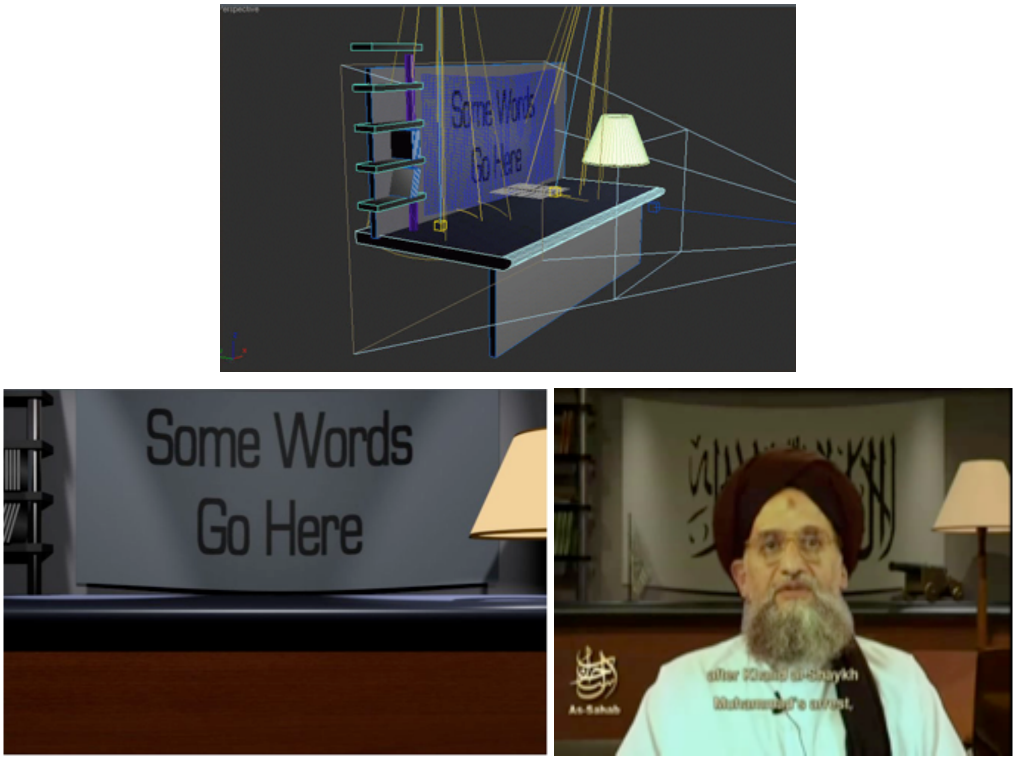


Figure 32. Wireframe model created by Spencer Lindsay (top). The rendered room (bottom-left) and real room (bottom-right).

[43] http://usa.autodesk.com/adsk/servlet/index?id=5659302&siteID=123112

Computer graphics specialist Spencer Lindsay speculated on the ease of using CG for an As-Sahab video.[44] Lindsay was challenged by Hacker Factor Solutions to recreate the Zawahiri room. After a few hours, Lindsay managed to recreate a room that appears similar to the Zawahiri video (Figure 32). The room created by Lindsay is not an exact match, but does strongly suggest that the background behind Zawahiri is a computer-generated room and was created using basic elements found in 3D Studio Max.

One item that is missing from the recreation is the canon on the desk. This appears to be a model of an 18th century British naval canon. While the exact canon was not found, wireframe models of similar canons are available for purchase for 3D Studio Max.[45]

6.2 Other Zawahiri Videos

The video from 28-Sept-2006 was not the only manipulated video. In fact, many of the videos featuring al Zawahiri test positive for chroma-key masking. Consider the video released on 27-July-2006 (Figure 33).[46] This video appears to show al Zawahiri sitting in a video studio.

When this video came out, many Americans became enraged at the US government. The main strife was generated by a single argument: if al Zawahiri is sitting in a studio making videos, the why can't we catch him? The answer is simple: he is *not* in a studio.

- **Observation**. The studio background shows a specular reflection above the scaffolding. However, there is no shadow below it.
- **ELA**. The 95% ELA shows a chroma-key halo around Zawahiri. The As-Sahab logo was added last. The background shows a repetitive pattern, indicating a solid color background. Since digital cameras add noise to images, the solid color is not from a digital camera; the background was drawn.
- **PCA PC1**. The line constructed from PC1 shows that the background was repeatedly resaved, the three pictures behind Zawahiri have three different resave levels, Zawahiri has a different level, and the As-Sahab logo shows no JPEG artifacts. This indicates that the picture was created using at least six layers.
- **PCA PC3**. The distance from the third principal component's line, as well as the first principal component's plane, emphasizes the lighting in the room. A large lighting ring can be observed on Zawahiri and the background. Since the lighted area behind Zawahiri is only slightly larger than the light on him, the background screen is no more than one to two feet behind him. The lighting contradicts the picture's rendering of a large studio.
- **Wavelets**. At 5%, the wavelets show six distinct layers. Mohammad Atef (left background) is very blurry, the World Trade Center is blurry, Mohammad Atta (right background) is nearly crisp, Zawahiri is crisp, and the As-Sahab logo is very crisp and showing sharp color transitions.

Since the background appears to have been drawn and includes other pictures, the question becomes: where did the other images come from? The image of Mohammad Atta (background right) appears to be from a photo found in the 9/11 Commission's final report.[47] This image was reduced in size and skewed to fit the background. The artist also removed the government exhibit tag from the lower corner. The image of Mohammad Atef (background-left) may be from a wedding video between his daughter and Osama bin Laden's son.[48]

[44] http://www.lindsaydigital.com/blog/wordpress/?p=90

[45] http://www.turbosquid.com/ FullPreview/Index.cfm /ID/253444 and http://www.turbosquid.com/ FullPreview/Index.cfm /ID/255325 are two examples of similar model canons.

[46] http://www.archive.org/details/Crusaders-and-Zoinism-war-on-Gaza-and-Lebnon

[47] http://www.rcfp.org/moussaoui/jpg/size600/GX00004.2-1.jpg

[48] In most videos, Mohammad Atef wears a collared shirt. Frames from the wedding video show him wearing a shirt similar to the one in pictured here.

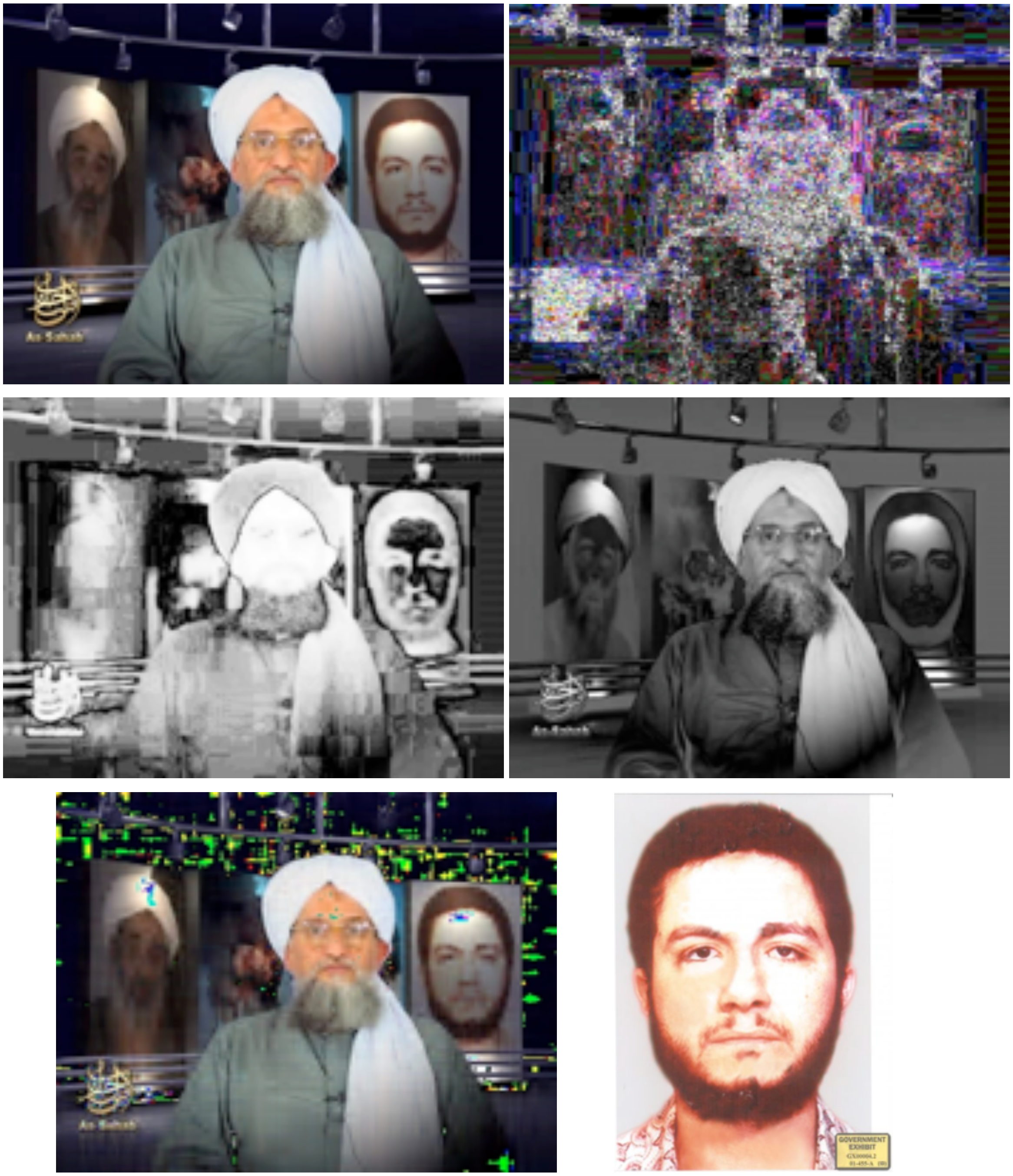


Figure 33. (From top to bottom) Still frame from the 28-Sept-2006 video, 95% ELA, PC1, PC3, 5% wavelets, and Mohammad Atta from the USA Government's 9/11 report.

6.3 Zawahiri: Back in Black

Claiming that there is a chroma-key background is not the same as actually seeing it. On 22-Jan-2007, SITE[49] – an organization that tracks terrorist activities – announced that they had intercepted an Al Qaeda video before it had been publicly released; SITE released it. Three days later, As-Sahab also released the video.[50]

The video (Figure 34) shows al Zawahiri in front of a black curtain. The curtain is not properly centered in the camera's view – there is a wedge of real background in the top-right corner. Changing the brightness of the image shows that the curtain is draped from a bar. While this could be a plain black fabric, the coloring appears to have a uniform hue. The fabric could be a type of chroma-key background called a "chroma-key sheet". Normally chroma-key backgrounds are mounted to reduce folds that could cause uneven lighting. However, a chroma-key sheet is suspended from a metal frame[51] (Figure 35) and may show suspension folds.

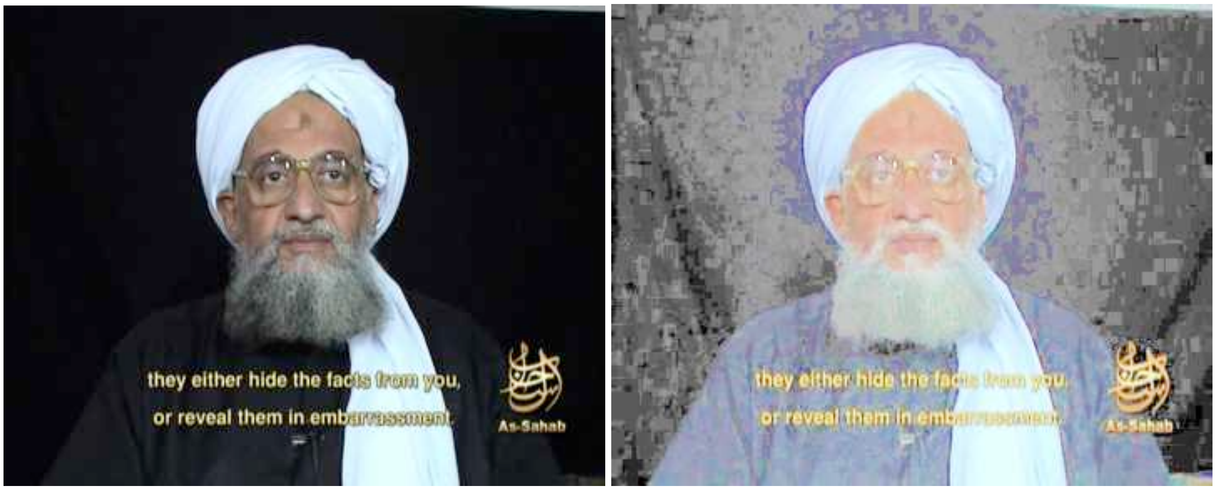

they either hide the facts from you,
or reveal them in embarrassment.
As-Sahab
they either hide the facts from you,
or reveal them in embarrassment.
As-Sahab

Figure 34. Frame from the 25-Jan-2005 As-Sahab video (left) and color enhanced (right).

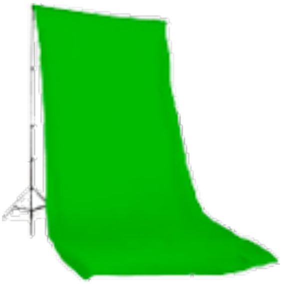


Figure 35. A chroma-key sheet.

6.4 Azzam al-Amriki Videos

Videos featuring al Zawahiri are not the only doctored As-Sahab videos. On 2-Sept-2006, Azzam al-Amriki (Azzam the American, aka Adam Gadahn) was featured in a video (Figure 36). The video appears to show him in a white room with a desk, computer, and some books. However, the 95% ELA suggests that the books do not exist in the room. While the computer, walls, desk, and Azzam are at one error level, the books, subtitles, and As-Sahab logo

[49] Previously called the SITE Institute (www.siteinstitute.org). As of January 2008, they have been renamed as the SITE Intelligence Group (www.siteintelgroup.org).

[50] http://www.archive.org/details/Correct-Equation

[51] http://www.videoguys.com/Emails/seriousmagic_blast.html

appear at a different error level. The PC1 analysis shows that the color range of the books is very different from the rest of the image, suggesting that they come from alternate sources.[52]

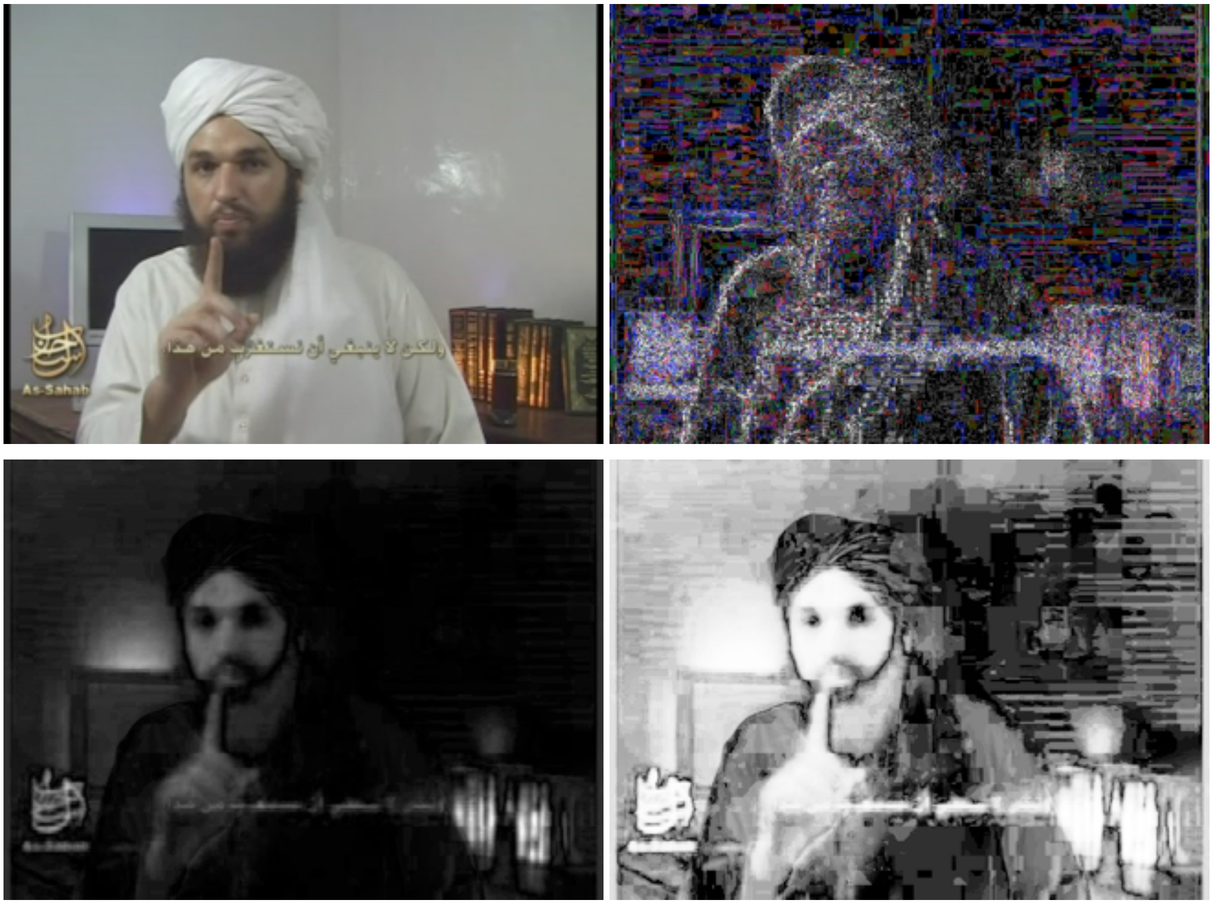


Figure 36. Azzam al-Amriki (Adam Gadahn) as seem in the 2-Sept-2006 video (top-left), 95% ELA (top-right), PC1 (bottom-left), and contrast-enhanced PC1 (bottom-right).

The PC1 image with an adjusted contrast shows a horizontal-line pattern. This is a media artifact; the original video of Azzam used an interlaced video source that is identified by the first principal component. However, while the horizontal lines are very visible on the background and mildly visible on Azzam, they are not visible on the books or As-Sahab logo. This suggests that these images did not come from an interlaced video source.

A scatter plot of the colors used in the al-Amriki video shows three distinct color regions (Figure 37). The main dense area contains the colors found across most of the picture: Azzam, walls, desk, and computer. The small cusp in the center of the plot consists of colors found in the As-Sahab logo and subtitle. These colors do not appear anywhere else in the image and form an independent cluster. Outside of the main image spectrum are the colors found in the books. The color spectrum for the books is distinctly outside of the main coloring for the image. This implies that the books are unlikely from the same footage as al-Amriki.

[52] Individual frames from the video, extracted by different people show the same attributes. http://www.memritv.org/data/thumbnails/clip_1257.jpg and http://rightvoices.com/wp-content/uploads/2006/09/adam_yehiye_gadahn.jpg

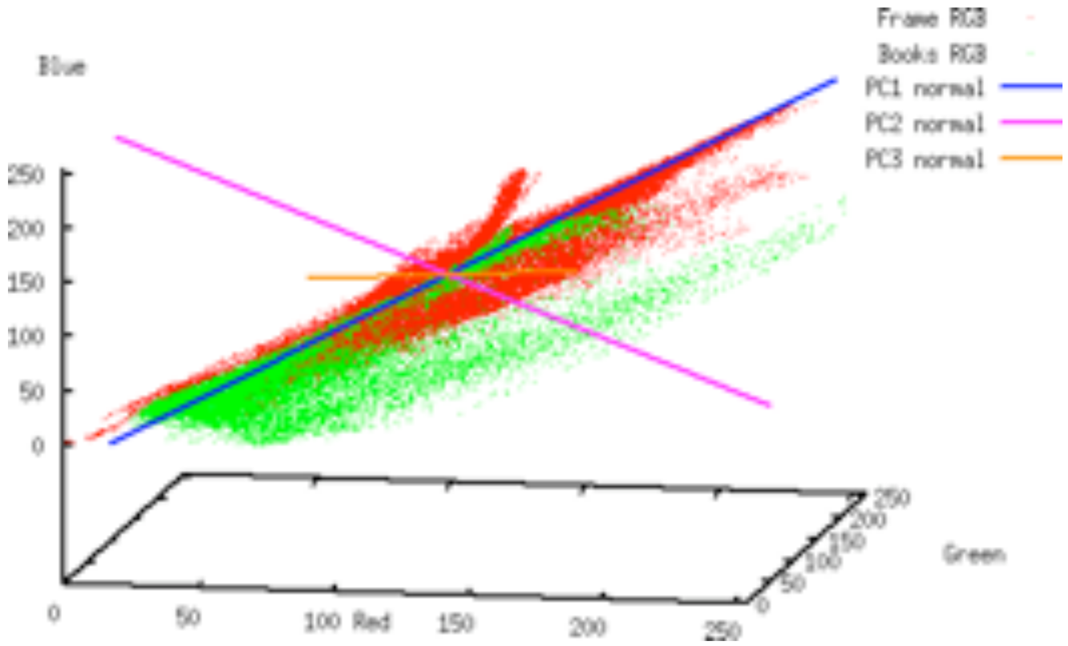


Figure 37. Scatter plot of the Azzam al-Amriki video frame. The book colors are in green.

6.5 Bin Laden's Beard

Prior to 2005, Osama bin Laden regularly appeared in videos and audio recordings. However, after 29-Oct-2004 he vanished. He did not reappear until 7-Sept-2007 – nearly 3 years later. The two videos (29-Oct-2004 and 7-Sept-2007 – Figure 38) each have oddities and raised debate over whether Bin Laden is still alive.

- **29-Oct-2004 video**. Commonly called the "**Graybeard**" video, this recording shows Osama bin Laden wearing a white shirt, yellow sweater (or robe), and a white hat. There is a plain brown background. He stands behind a podium and moves papers between piles as he reads. This video was initially released on Al Jazeera (TV). Low quality copies of the video are available from SITE[53] and the Internet Archive[54], and higher quality screen shots from the video are available from various sources.

  While most of the controversy around this video was based on the spoken message, there was some debate about his beard. In particular, the beard is very gray; previous footage of bin Laden showed a dark beard and gray streaks.

- **7-Sept-2007 video**. Commonly called "**Blackbeard**"[55], this video initiated more controversy. Bin Laden appears to wear the same clothing as the 2004 video. It appears to show the same set – he is in front of a podium, moves papers as he reads, and has a plain backdrop. However, there are some significant differences. The biggest difference is his beard: it is shorter than the 2004 video and it is dark. The contrast in beards made many people believe that the black beard was fake – either dyed or computer graphics.[56]

[53] http://siteintelgroup.org/multimedia/video/video_1099085219.wmv

[54] http://ia301231.us.archive.org/1/items/zimas00006/rasala.rm

[55] Technically, the beard in the video is a dark color, but not necessarily black. It may a dark red or brown color.

[56] http://timesofindia.indiatimes.com/Osama_Bin_Laden_is_dyed_and_alive/articleshow/2346119.cms

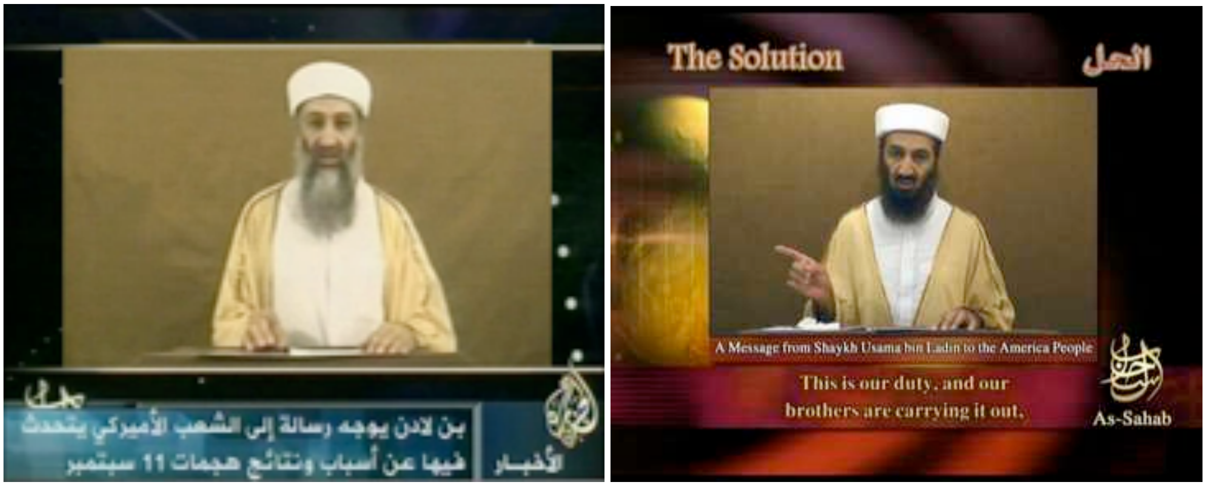

بن لادن يوجه رسالة إلى الشعب الأميركي يتحدث
فيها عن أسباب ونتائج هجمات 11 سبتمبر
الأخبار
The Solution
الحل
A Message from Shaykh Usama bin Ladin to the America People
This is our duty, and our brothers are carrying it out.
As-Sahab

Figure 38. Frames from the 2004 "Graybeard"[57] and 2007 "Blackbeard"[58] videos.

6.5.1 Blackbeard Analysis

The Blackbeard video includes multiple abnormalities. For example, the 26-minute video only contains 3.5 minutes of real animated footage; the remainder of the video shows a still-frame of bin Laden. Moreover, the animated segments are not adjacent: the video begins with two minutes (1:56) of animated video, followed by ten minutes of a still-frame (from 1:56 to 12:29). This is followed by a minute and a half of heavily spliced video footage (12:29 to 14:01), and ends with 12 minutes of the same still-frame (14:01 to 26:30).

The still-frame segments are associated with spliced audio. In particular, there is an audible click at 2:05. At this point, the background room echo changes, indicating that it is not part of the same video sequence. All references to current events (which date the video to 2007) are made during the still-frame sections.

The two animated sequences appear to have different aspect ratios (Figure 39). The second sequence appears wider but not taller. In particular, his eyes are farther apart, hat is wider, desk is wider, and even the patterns on the background are wider apart. In contrast, the vertical distance from the desk to his face and hat did not change. Because he is wider and not taller, it is likely that the aspect ratio changed and not the camera's focal length.

It is very rare to adjust the aspect ratio during a recording. The difference in aspect ratios within the 2007 video suggests that either the video consists of two different recording that were spliced together, or two segments that were post-processed differently.

[57] http://www.aljazeera.net/mritems/Images/2004/10/30/1_507945_1_34.jpg

[58] http://ia341235.us.archive.org/0/items/The-Solution/The-Solution.doc frame 19,800 (at 25fps).

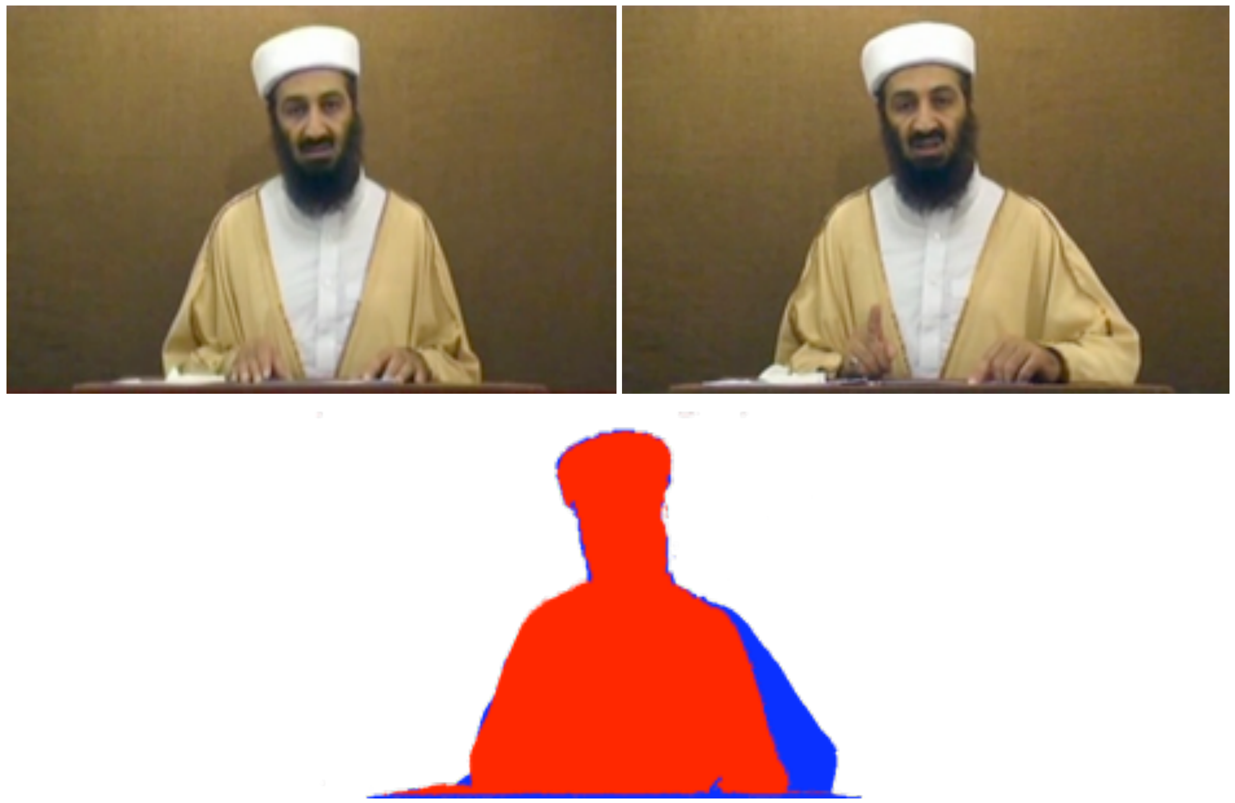


Figure 39. Blackbeard video frames 4,000 (left, 2:40) and 19,410 (right, 12:56) show that the figure is wider but not noticeably taller in the two animated sections.

6.5.2 Beard Digital Image Analysis

A basic digital image analysis does not identify any manipulation within the images themselves. ELA, PCA, and wavelets identified no manipulation. Luminance gradient also did not identify any abnormalities between the 2004 and 2007 videos. Both videos have realistic noise across the image.

LG analysis (Figure 40) does identify one interesting aspect. The lighting in the 2004 and 2007 videos appears similar, if not identical. In both videos, LG shows the left eye as light and right eye as dark, and left half of his beard as light with right-half dark. The transition from dark to light appears to be in the same location. In fact, the top edge along his hat shows a gradient coloring that appears virtually identical in both videos. The Blackbeard video appears to use the same lighting, three years after the Graybeard video.

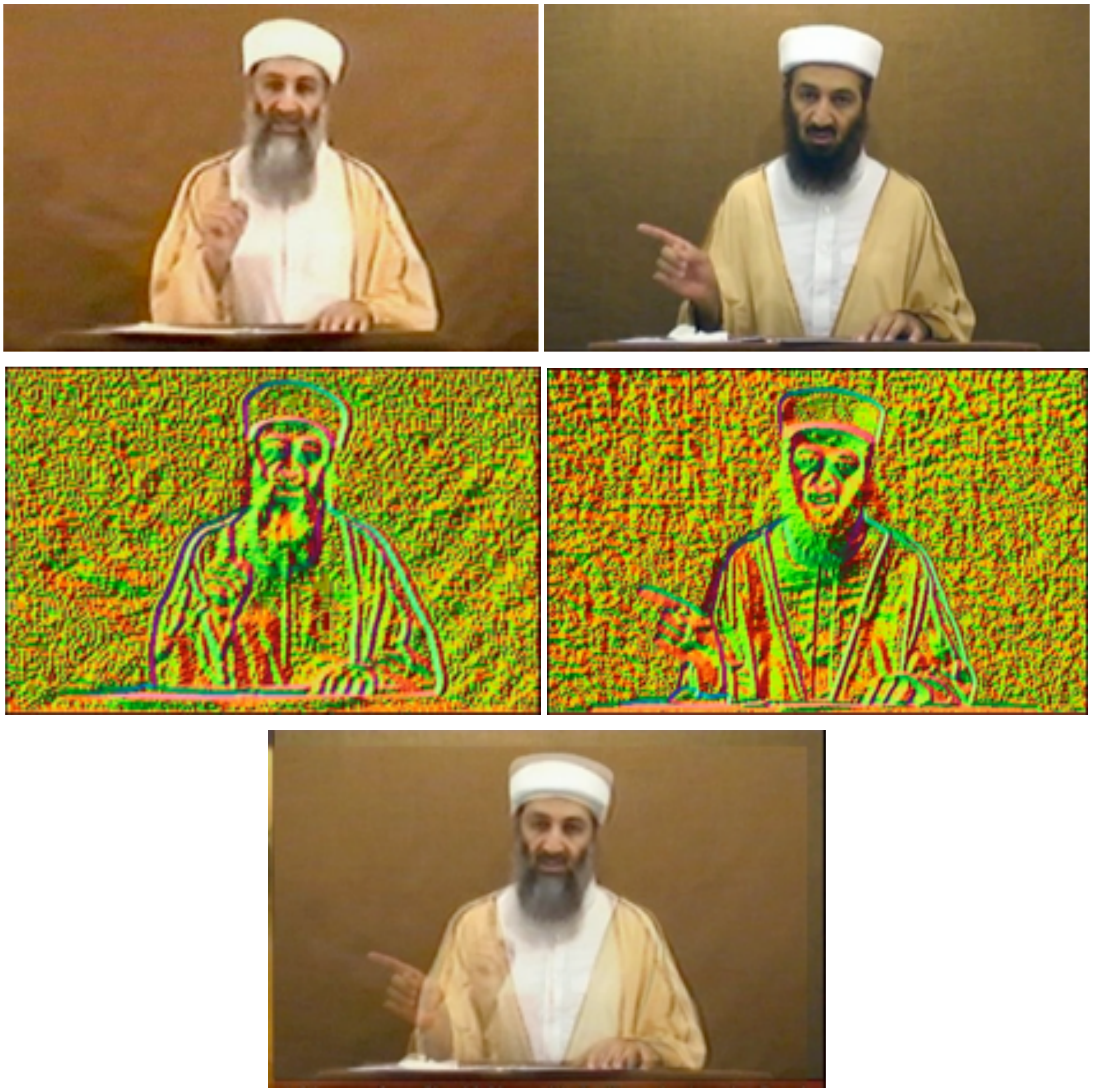


Figure 40. Frames from the Graybeard video[59] (top left) and 19,800 (top right, 13:12) of the Blackbeard video. Both frames show bin Laden with his head at the same angle. LG analysis (middle) indicates similar or identical lighting. An overlay of the two frames (bottom) indicates similar camera positions.

6.5.3 Beard Video Camera Angle

The Graybeard and Blackbeard videos have similar lighting and similar sets (background, table, clothing, etc.). Frame 19,800 (13:12) in the Blackbeard video shows bin Laden with his head cocked at the same angle as the frame from the Graybeard video. The overlay of these two videos, aligned only by shifting the frames to align on eye position, (Figure 40) shows many similarities:

- Significant features align: eyes, eyebrows, nose, face creases, mouth location, and dark beard border along his cheeks align perfectly. Similarly, his shirt collar perfectly lines up in the overlaid frames. His shoulders and desk are at the same height. This alignment occurs without adjusting the image's aspect ratio or scale.

[59] http://www.foxnews.com/images/143099/4_22_102904_binladen_450.jpg

- Although the 2007 Blackbeard video has a crumpled piece of paper on the stack, the stacks of paper appear to be at the same place on the podium.
- The podium appears to have a binder with the spine located off-center to the podium. Although the spine is not in the exact same place in both videos, it is nearly in the same position in both videos.

There are differences between the two videos. For example:

- Blackbeard wears his hat a little higher and Graybeard has his robe open instead of closed.
- The beards are different sizes; the gray beard is wider and longer than the black beard.
- The background behind the 2004 Graybeard video shows large wrinkles in the fabric, while the 2007 Blackbeard video only shows smaller wrinkles in the lower left corner.

The differences imply that these are not the same frame. However, the similarities suggest that it is the same person with very similar camera positions. In addition, since the 2007 video includes two segments with different aspect ratios, the alignment shows that Graybeard uses the same aspect ratio as the second section of the Blackbeard video.

6.5.4 Beard Theories

The sets, lighting, and camera angle appear similar, if not identical, between the 2004 and 2007 videos. There are only two remaining differences: the coloring and the beard.

The 2004 Graybeard video appears brighter than the Blackbeard video. The background is more reddish and the whites on his hat and shirt appear washed out. However, this color difference could be due to post-processing. For example, modifying the pallet to an over-exposed color space enhances the Blackbeard video (Figure 41). The background becomes reddish and the whites become overexposed, similar to the 2004 video.

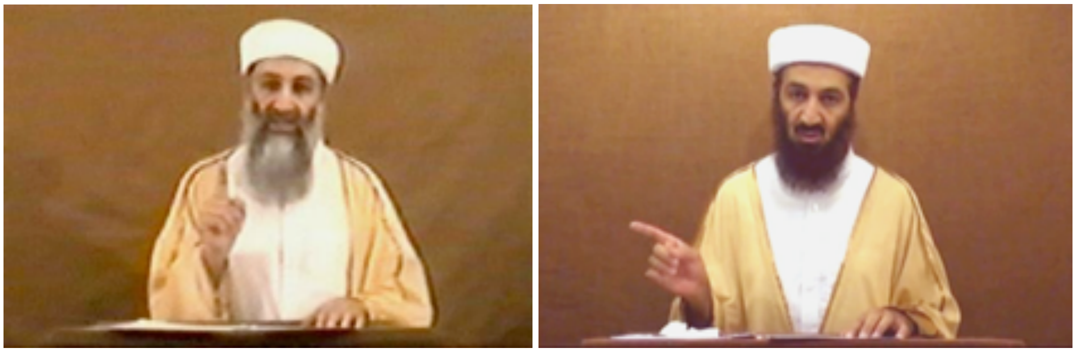


Figure 41. Adjusting the color of the 2007 video (right) creates similar frame coloring to the 2004 video (left).

Considering that the 2004 and 2007 videos appear to only differ by the color of the beards, there are two theories about the differences and similarities.

- **Theory #1: Both videos are from 2004**. This theory suggests that the similarities are due to both videos being recorded at nearly the same time – hours or possibly days apart, but not months or years. The recordings used the same clothing, studio, set, lighting, camera angle, and aspect ratio to record at least two videos, and possibly many more that have not been made public.

  Under this theory, the differences are primarily due to post processing (frame coloring). However, one of the beards must be fake. Considering that the gray beard is larger than the black beard, and both share the same hairline and dark facial patches, it is probable that the gray beard is a costume and/or dyed.

  In addition, this theory claims that the footage was recorded in 2004, even though the Blackbeard video was not released until 2007. Considering that the current events mentioned in the 2007 video do not happen during the animated sequences, and only happen after splices in the audio track (where the voice quality changes), this theory implies that Osama bin Laden has not been seen since 2004.

The different aspect ratios between the 2007 video segments could indicate multiple segments. It is likely that the two Blackbeard segments were recorded in sequence, but the order cannot be determined; either the 2-minute segment was recorded first, followed by the 1.5-minute segment, or vice versa.

Since the aspect ratio between the 2004 and second 2007 segments match, this implies that they were recorded around the same time or processed the same way. However, since the coloring differs, the difference is more likely due to recording that processing.

Together, this theory implies that the recording order was either Blackbeard segment #1, followed by segment #2, and then Graybeard, or the other way around (Graybeard followed by Blackbeard segment #2, then segment #1.). While the two Blackbeard segments were likely recorded shortly after each other, the wrinkles on the background imply that the 2004 and 2007 videos were recorded hours or possibly days apart. The recordings were made without breaking down the set, lighting, or camera position.

- **Theory #2: Recreation**. An alternate theory is that the videos come from a recreated set. The photographer and set designer would need to pay particular attention to the minor item alignment, such as desk position, distance to the subject, paper position on the podium, and clothing. This theory implies that the 2007 video is new and that special care went into making it look like the 2004 video.

  However, this theory has some glaring problems. For example, if they went out of their way to recreate the podium, lighting, and camera angle, then why did they not alter his beard or match the robe? Similarly, this theory does not explain why the aspect ratio changes during the 2007 video.

Occam's Razor says that the simplest solution is likely the correct solution. Considering all of the similarities and the difficulties associated with recreating the same set, it is probable that Theory #1 is correct – Osama bin Laden has not been seen alive since 2004.

7 Additional Research

The approaches covered in this paper represent a work-in-progress, both at Hacker Factor Solutions and in the field in general. As digital image technologies advance, so do methods for detecting manipulations and distinguishing reality from fiction. This is a fairly new field, and most researchers are working on different (but related) technologies. Example research project include:

- Hany Farid (Dartmouth College): Image manipulation detection.
- Jessica Fridrich (Binghamton University): Steganography detection and tamper detection.
- Shih-Fu Chang (Columbia University): Media forensics using signal processing and statistical pattern recognition.
- Nasir Memon (Polytechnic University): Image steganography and manipulation detection.
- Min Wu (University of Maryland): Digital media fingerprinting.

In addition, fields such as robotic vision are developing technologies that may be readily applicable to forensic image analysis.

8 Conclusion

As pictures are manipulated to sway opinions, image analysis and digital forensics grow in importance. This paper covers different methods for viewing and analyzing images. Although a single image may pass one or two tests, a modified image is unlikely to pass all of the tests. In addition, these methods are only the beginning of the available analysis approaches. Other methods do exist and are designed to catch other image manipulation techniques.

9 Acknowledgements

Neal Krawetz at Hacker Factor Solutions is grateful to the following people for their insight, suggestions, feedback, enlightening conversations, and technical opinions: Cynthia Baron (Northeastern University), Adam Bates (University of Colorado at Boulder), Paul Whyman, Bob Gobeille, Spencer Lindsay (lindsayDIGITAL), Pedro Bueno (Internet Storm Center), Hany Farid (Dartmouth College), Mark Rasch, J.D., and the Department of Defense Cyber Crime Center.