Detecting Image Forgery

State of the Art

UTIA

March 10, 2009

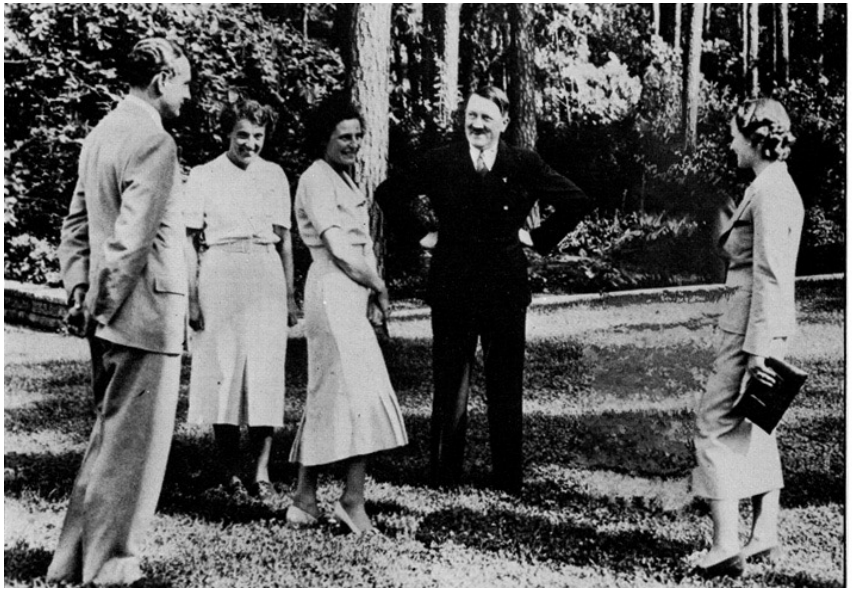
Motivation

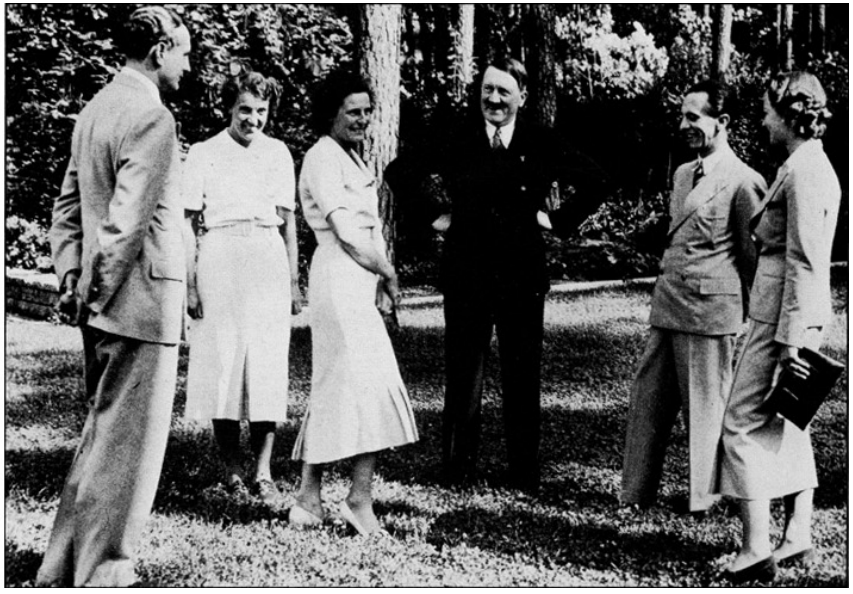
Stalin (1930)

- Seeing is believing ... or is not?

Mao Tse-tung and Po Ku (1936)

Adolf Hitler and Joseph Goebbels (1937)





Vladimir Clementis and Klement Gottwald (1948)

Motivation

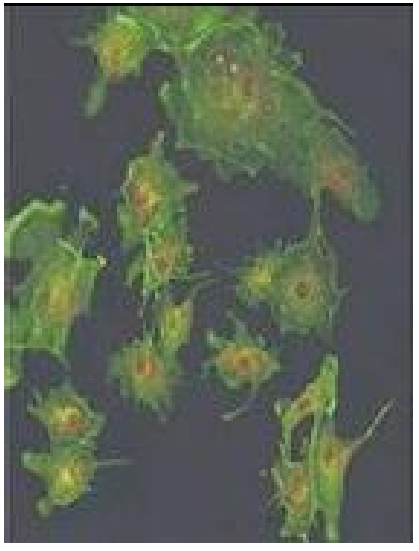
- Apparently, the photo manipulation has a long and rich history.
- In today's digital age, photos are very easy to manipulate and edit due to availability of many powerful editing software packages.
- For example, Adobe Photoshop has more than 6 mil regisrated users.
- It is possible to add or remove important features from an image without leaving any obvious traces of tampering.
- Resulted in a high number of image forgeries.

Motivation

- So, how much can we trust photos?

Motivation

- Today, we face image forgeries even in scientific literature.
- Journal of Cell Biology, estimates that around 25 percent of manuscripts accepted for publication contain at least one image that has been inappropriately manipulated (mostly changes do not affect the scientific meaning of the results).
- One of the recent famous cases of digital image forgeries in science area was the South Korean scientist Hwang Woo-suk's claim that he cloned stem cells.



Digital Forgery Detection Techniques

Digital Forgery Detection Techniques

- Active approach
  - Data hiding (e.g., watermarks)
  - Digital Signatures

Not very useful in forgery detection. Mostly must be inserted at time of recording.

- Passive (blind) approach
  - Regarded as a new direction
  - The area is growing rapidly
  - Does not need any prior information about the investigated image or its source
  - Mostly, try to analyze each forgery type separately (duplicated image regions, resampling, double JPEG compression, inconsistent noise patterns, etc.) and detect each type separately.

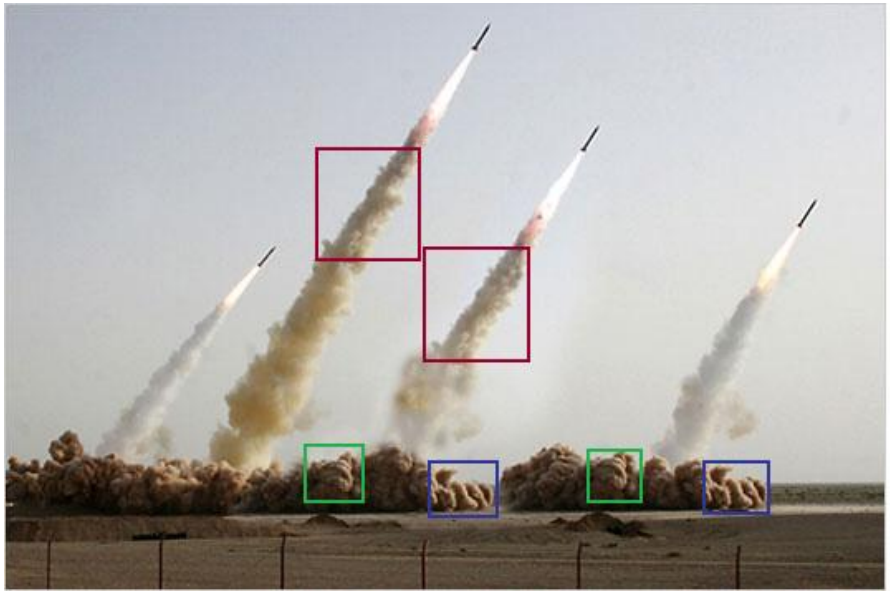
Cloning (Copy–Move Forgery)

Copy-Move Forgery

- A common type of digital image forgery.
- A part of the image is copied and pasted into another part of the same image typically with the intention to hide an object or a region of the image.
- Altered regions are from the same image, therefore they have compatible statistical properties with the rest of the image.
- To make the forgery harder to detect, one can use noise, blur, contrast changes, etc.
- JPEG makes the detection even more difficult.
- As the result duplicated regions mostly are not identical.

Example - Iranian missile test (a forgery easily detectable by our system)

Example



Detecting Copy-Move Forgery

- Tiling the image with overlapping blocks
  - Blocks are horizontally slid by one pixel rightwards starting with the upper left corner and ending with the bottom right corner.
  - Blocks ($R \times R$ pixels) are assumed to be smaller than the size of the duplicated regions.
  - The total number of overlapping blocks for an image of $M \times N$ pixels is $(M - R + 1) \times (N - R + 1)$. For instance, an image with the size of 640×480 with blocks of size 20×20 yields 286281 overlapping blocks.
- Blur moment invariants representation of the overlapping blocks.
  - Blur invariant $B(p, q)$ can be derived as a functional consisting of central moments:

$$B(p,q) = \mu_{pq} - \alpha \cdot \mu_{pq} - \frac{1}{\mu_{00}} \sum_{n=0}^{K} \sum_{i=m_1}^{m_2} \binom{p}{t-2i} \binom{q}{2i} \cdot B(p-t+2i, q-2i)\mu_{t-2i,2i}$$

Detecting Copy-Move Forgery

- We use 24 blur invariants up to the seventh order to create the feature vector of each block.
- In the case of an RGB image, the dimension of the feature vector is 72 (24 invariants per channel)
- Principal component transformation is used to reduce the space dimension.
- Blocks Similarity Analysis in a Kd-tree representation.
  - A duplicated region consists of many neighboring duplicated blocks. If we find two similar blocks (based on a distance measure in the Euclidean space) and if their neighborhoods are also similar to each other, there is a high probability that they are duplicated.

Results

(a)

(b)

(c)

Results

Discussion

- An invariant approach with respect to blur and contrast changes.
- Moment invariants are computed by a summation over the whole image, so they are not significantly affected by additive zero–mean noise.
- Problem with uniform areas in images, since we are looking for identical or similar areas.
- A high computational time.

Interpolation

Interpolation and Image Forgeries

- Typically, geometric transformations like scaling and rotation are commonly used tools for creating image forgeries.
  - Especially, when two or more images are spliced together.
- These procedures are typically based on a resampling and interpolation step.
- The interpolation step brings specific statistical changes into the image.
- Detecting these changes can be helpful in detecting and describing the geometric transformations which the image being investigated has undergone.

Detecting Inetrpolation

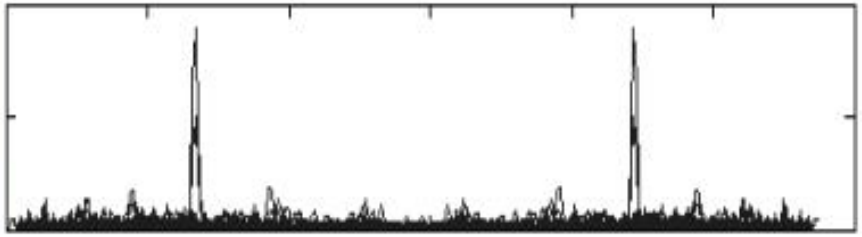
- We focus on low–order interpolation polynomials. They are used extensively because of their simplicity and implementation unassuming properties. Mainly on linear and cubic interpolation.
- These interpolation kernels are not ideal, are narrow–band and bring into the signal specific periodically varying artifacts.
- To detect the traces of interpolation in interpolated signals, we will focus on their frequency spectrum.
- Their frequency spectrum is correlated.

Interpolation

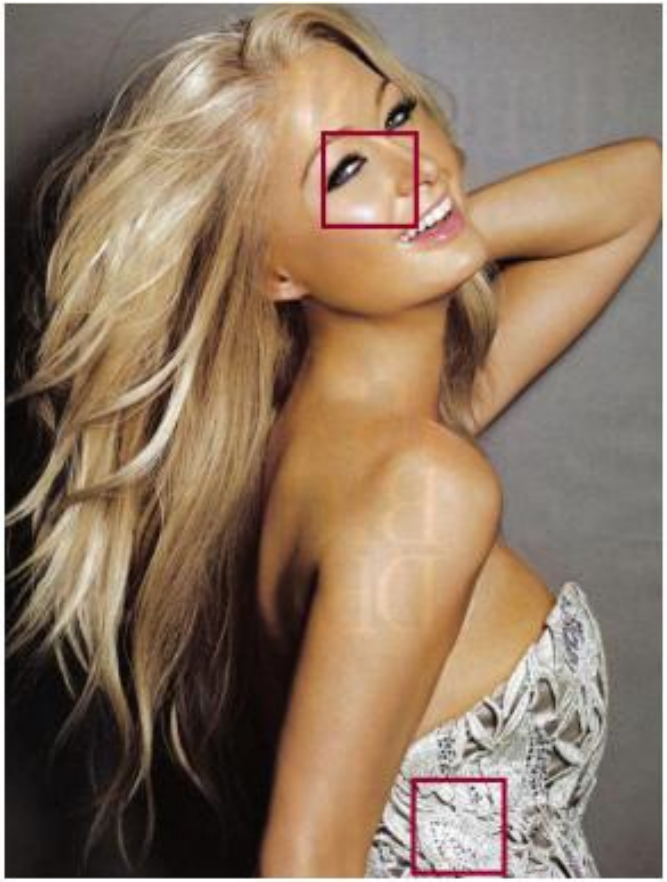
- Method:
  - ROI Selection.
    - A typical image consists of a discrete set of regions corresponding to objects needing verification.
  - Signal Derivatives Computation.
    - To emphasize the periodic properties present in a geometrically transformed image, a set of derivatives of different orders of ROI is computed.
  - Radon Transformation.
    - To detect the traces of rotation, the radon transformation is computed (at angles 0° to 179° in 1° increments).
  - Search for Periodicity.
    - Radon transformation vectors are analyzed for presence of periodicity using FFT. The output of the method is created by taking the maximum value at each frequency.

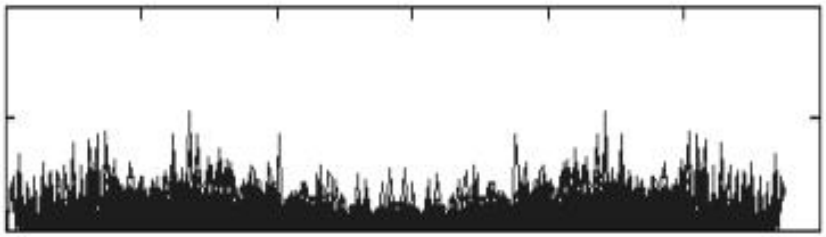
Example

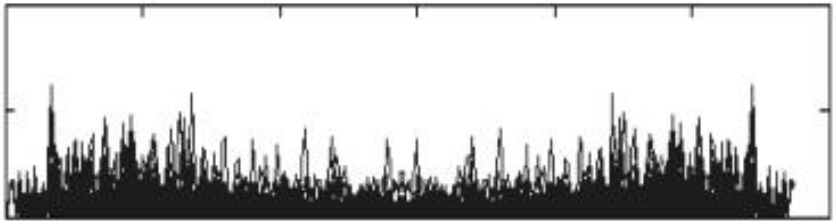
Example *rotation 15° followed by resizing 1.2, bicubic*



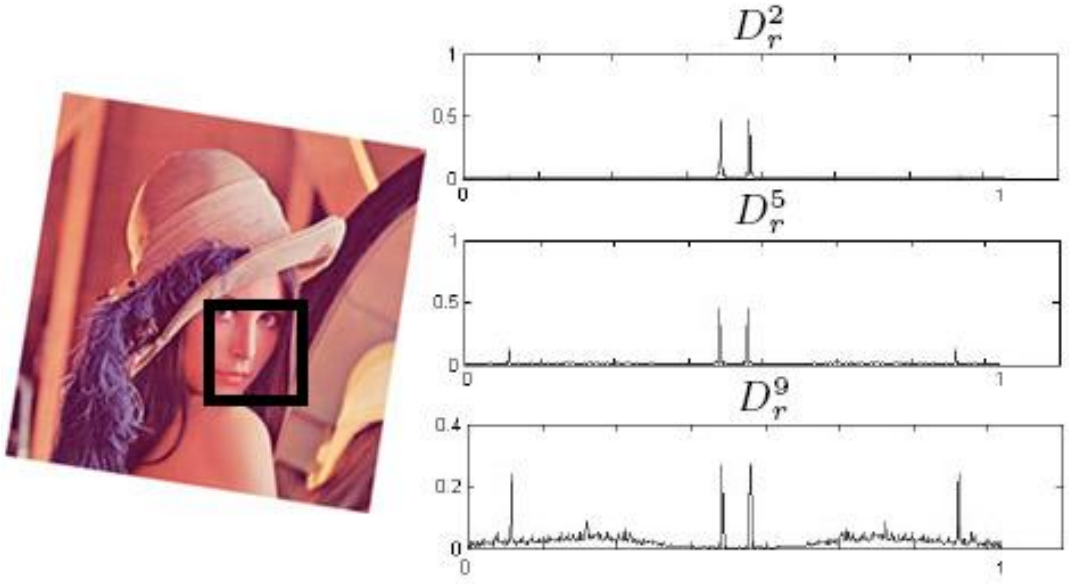
Example *non–interpolated*







Why a set of filters? *rotation 9° followed by resizing 1.9, bicubic*



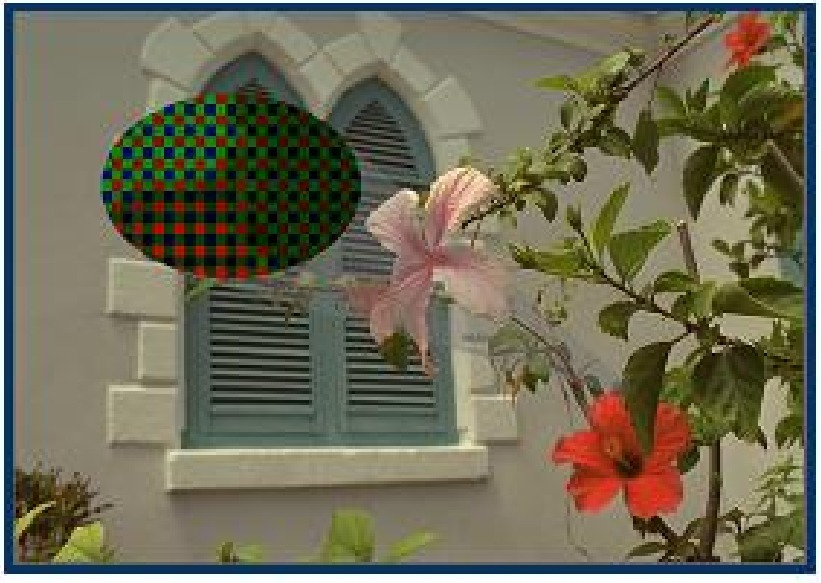
Discussion

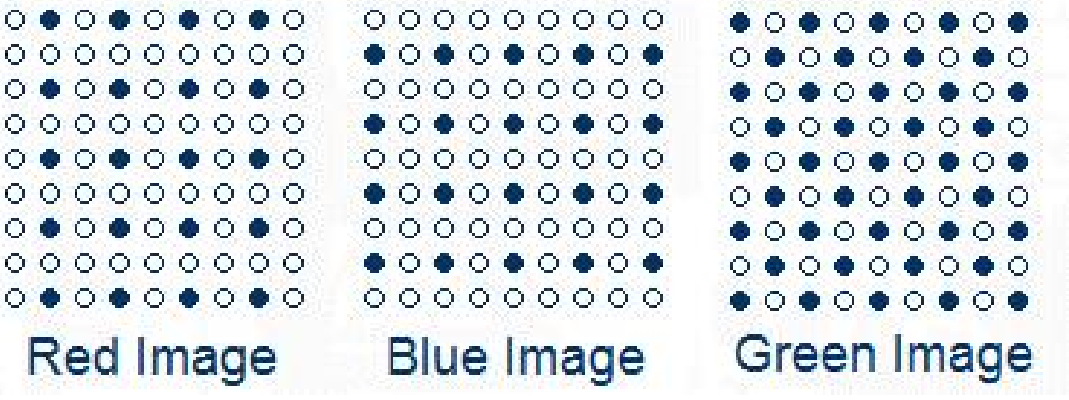
- Positions of detected peaks correspond to the scaling factors and rotation angles => helpful in estimating and describing the parameters of geometric transformation.
- Set of filters gives better results than using only one filter (specially when the image contains several transformations).
- The method works well for low–order interpolation polynomials: nearest neighbor, linear or cubic.
- The method is fast and simple.
- The detection performance decreases as the order of interpolation polynomial increases.
- Weak results for noise corrupted images and JPEG.

CFA Interpolation

- Color images in many cameras are obtained using a CFA interpolation process.
- Missing colors are computed using the neighboring pixels.
- There are many different CFA interpolation algorithms (bilinear, bicubic, median–based, gradient–based, adaptive, directional filtering, etc.)
- Some of them (e.g., bilinear) cause that a subset of pixels within a color channel are periodically correlated to neighboring pixels.

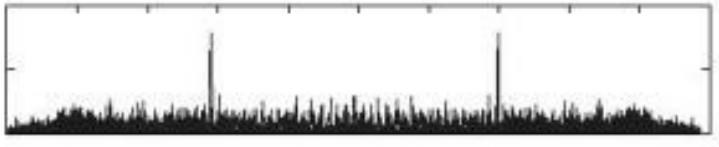
CFA Interpolation





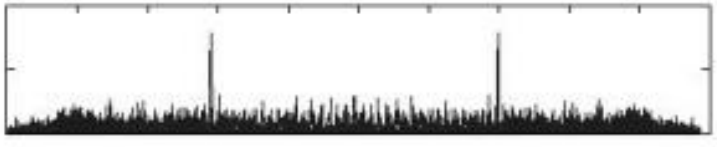
Effect of CFA

Bilinear CFA

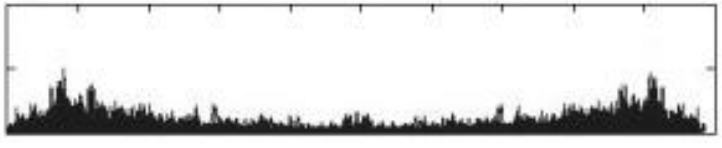


Effect of CFA

Bilinear CFA



Directional filtering CFA



Noise

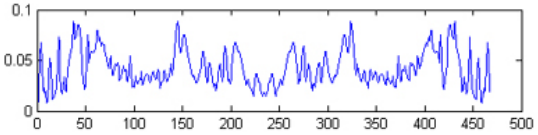
Image Noise Inconsistencies Analysis

- The interpolation–detection method is sensitive to noise.
- Noise is a commonly used tool to conceal the traces of tampering.
- When the described interpolation detector fails, it can be helpful investigate the noise consistency of the image.
- We assume white Gaussian noise with variance σ^2 which can spatially vary.
- Problem definition: Given an image containing a discrete set of regions with different noise levels, our task is to determine the presence of such regions and to localize them.

Method

- Tiling the HH_1 sub–band of the image's wavelet transform with non-overlapping blocks.
  - HH_1 denoted the diagonal details of the highest resolution.
- Blocks noise level estimation using the widely used median–based used estimator: $\hat{\sigma} = \frac{MAD_{HH1}}{0.6745}$
  - MAD denoted median absolute deviation. The median measurement is insensitive to isolated outliers of potentially high amplitudes.
- Blocks merging.
  - By employing a simple threshold–based region merging technique. The homogeneity condition is the estimated noise standard deviation of each block.

Noise Inconsistencies Detection



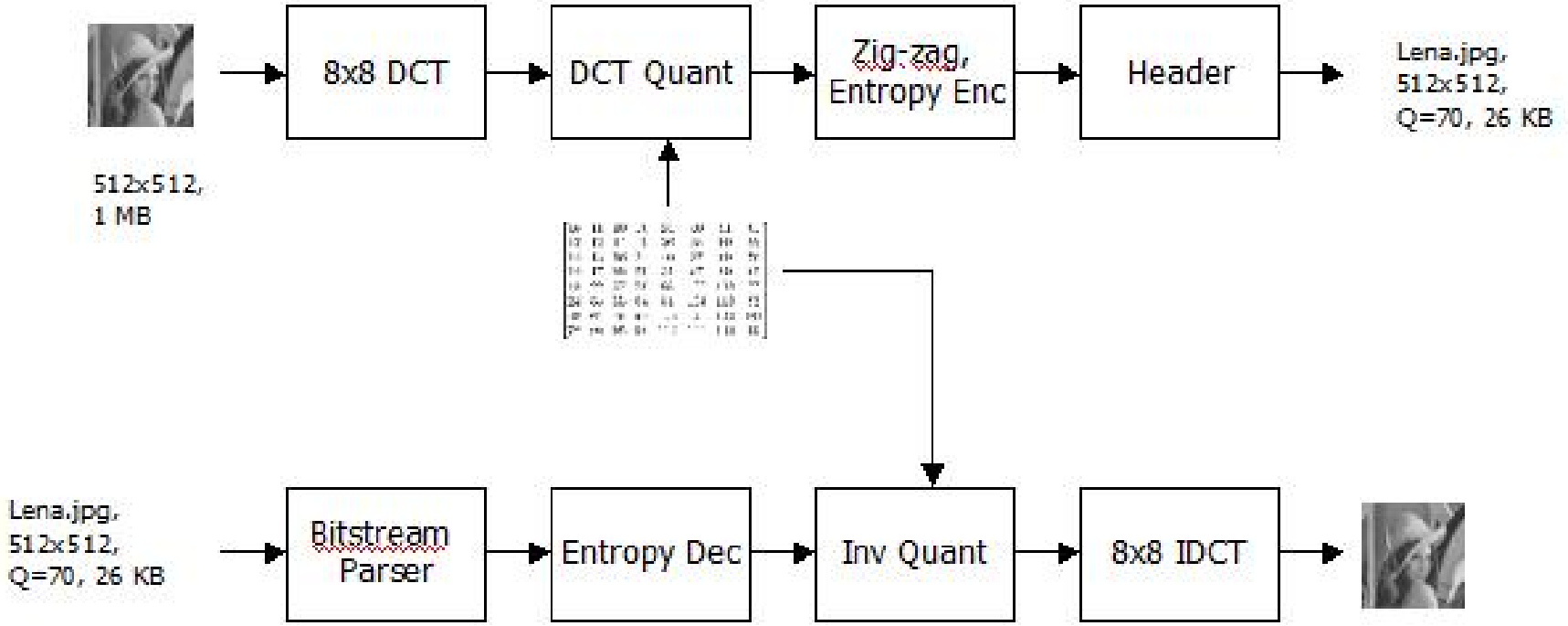
Discussion

- A simple way how to divide an investigated image into various segments with homogenous noise level.
- Not good as a stand alone detector. A good supplement to other forgery detection methods (interpolation detectors, near-duplicated image regions detectors, etc.).
- When the noise degradation is very small, the method fails. Not a big limitation. In the case of minor noise degradation, the resampling detector and other forgery detectors do not fail.

JPEG

JPEG Double Compression Detection

- JPEG is a popular image compression and file format standard.
- When an image is doctored, typically it is opened in an editing software and then re–saved.



JPEG Double Compression Detection

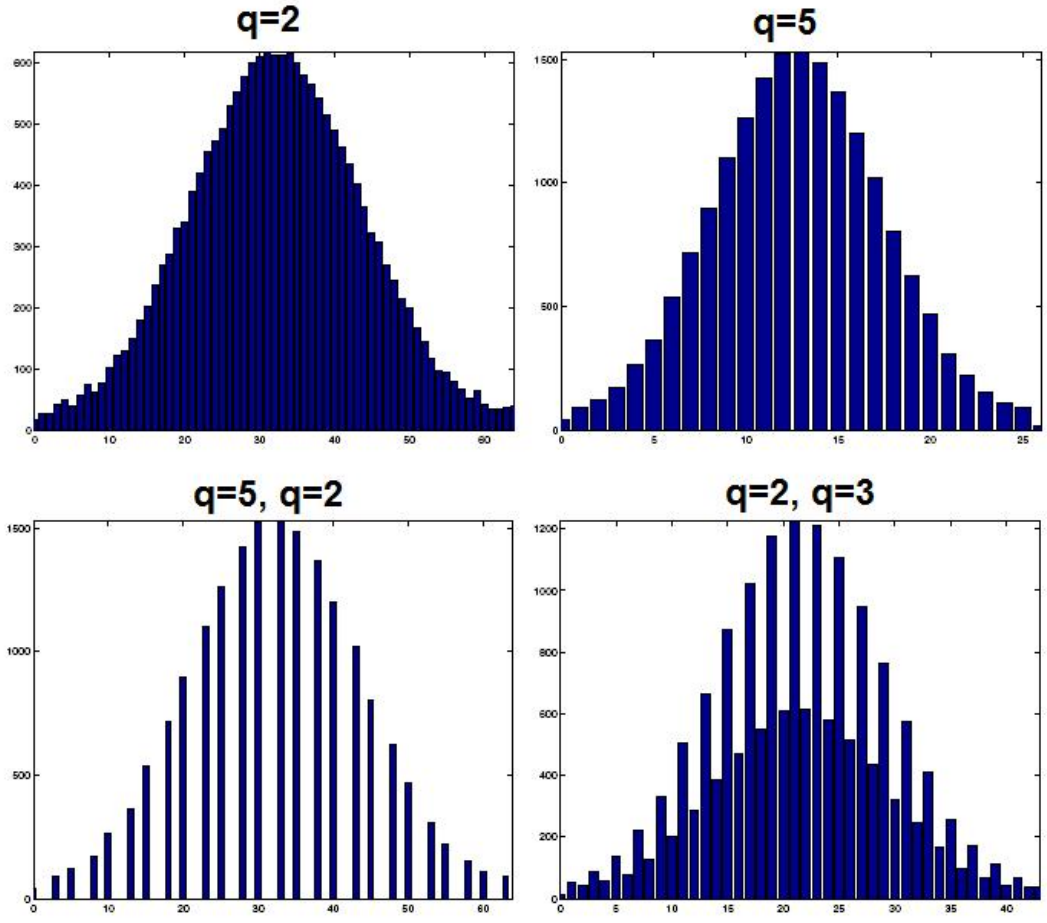
- Why DCT? DCT has a very desirable energy compaction property, specially in the lower frequency areas. That is, a DCT transformed signal has significant lower frequency components. This allows for easier quantization.

$$\begin{bmatrix} -76 & -73 & -67 & -62 & -58 & -67 & -64 & -55 \\ -65 & -69 & -73 & -38 & -19 & -43 & -59 & -56 \\ -66 & -69 & -60 & -15 & 16 & -24 & -62 & -55 \\ -65 & -70 & -57 & -6 & 26 & -22 & -58 & -59 \\ -61 & -67 & -60 & -24 & -2 & -40 & -60 & -58 \\ -49 & -63 & -68 & -58 & -51 & -60 & -70 & -53 \\ -43 & -57 & -64 & -69 & -73 & -67 & -63 & -45 \\ -41 & -49 & -59 & -60 & -63 & -52 & -50 & -34 \end{bmatrix} \longrightarrow \begin{bmatrix} -415 & -30 & -61 & 27 & 56 & -20 & -2 & 0 \\ 4 & -22 & -61 & 10 & 13 & -7 & -9 & 5 \\ -47 & 7 & 77 & -25 & -29 & 10 & 5 & -6 \\ -49 & 12 & 34 & -15 & -10 & 6 & 2 & 2 \\ 12 & -7 & -13 & -4 & -2 & 2 & -3 & 3 \\ -8 & 3 & 2 & -6 & -2 & 1 & 4 & 2 \\ -1 & 0 & 0 & -2 & -1 & -3 & 4 & -1 \\ 0 & 0 & -1 & -4 & -1 & 0 & 1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix} \longrightarrow \begin{bmatrix} -26 & -3 & -6 & 2 & 2 & -1 & 0 & 0 \\ 0 & -2 & -4 & 1 & 1 & 0 & 0 & 0 \\ -3 & 1 & 5 & -1 & -1 & 0 & 0 & 0 \\ -4 & 1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

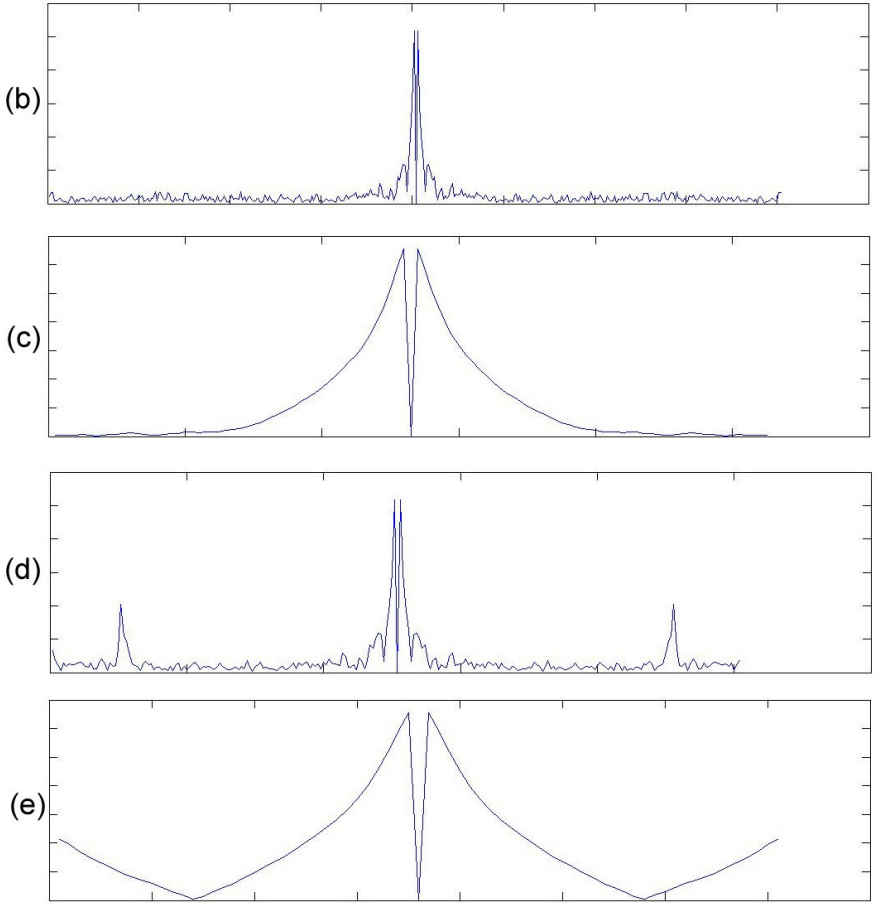
Double Quantization

- A JPEG re–saved image is a double quantized signal.



Double Quantization

(a)



Discussion

- A double quantized image is not necessarily a forgery.
- Not all double quantization procedures bring into the image detectable changes.
- Methods are sensitive to cropping, noise, etc.

Conclusion

The area is growing rapidly and results obtained promise a significant improvement in forgery detection in the never–ending competition between image forgery creators and image forgery detectors.

Thank You!