Comparative Study of Offline Scanned Documents and Digital Images for Alteration Using Analysis of Histograms

Pina Patel[1], Chhote Raja Patle[2], Anuwanshi Sharma[3], Anita Yadav[4]

How to cite this article:
Pina Patel, Chhote Raja Patle, Anuwanshi Sharma *et al*. Comparative Study of Offline Scanned Documents and Digital Images for Alteration Using Analysis of Histograms. Int Jr of Forensic Sci. 2024;7(1):63–67.

Abstract

This study aims to examine the comparative analysis of offline scanned documents and digital images to detect alterations by analyzing their histograms. In the digital age, document manipulation significantly threatens document authenticity. By leveraging histogram analysis, which provides a visual representation of pixel intensity distribution, this study seeks to identify differences between altered scanned documents and original digital images. The primary objective is to develop a reliable method for detecting document alterations by comparing histograms. Alterations made to documents can affect the pixel intensity distribution, leaving detectable traces within the histogram. Through meticulous analysis of these histogram variations, this study aims to establish a dependable baseline for identifying suspicious or fraudulent documents.

Keywords: Digital Forensics; Digital Document; Computer Manipulated Documents; Alteration; Histogram.

INTRODUCTION

A digital document is any piece of information that has been written down in such a way that a computer or other electronic device is required to display, interpret, store, and process it. Electronic mail communications, documents (text, photos, and spreadsheets), and audio-visual documents (jpeg, png, etc.) are all included in this.[1] Digital papers, however, are simple to replicate and falsify, unlike their analog counter parts. The information that a picture represents may be quickly created, altered, and modified in today's digital environment without leaving any evident signs of these activities.[2] This has led to significant uncertainty over the validity and dependability of digital materials. Even for a non-expert, due to the advancement of digital photography and powerful image modification tools, it is now fairly simple to create convincing forgeries of digital photographs.[3,4] The majority of the time, forgers modify numbers, letters, or words by copying and pasting them into digital software that is freely accessible on Google, changing the document's meaning in the process.[5,6,7] A document can be altered by changing its content, layout, or overall structure. It is possible to achieve this by adding, removing, or changing text, photos,

Author's Affiliation: [1]M.Sc Student, [2]Assistant Professor [4]Associate Professor, Department of Forensic Science, School of Sciences, Sanjeev Agrawal Global Educational University, Bhopal 462022, Madhya Pradesh, [3]Ph.D. Scholar, Division of Forensic Science, School of Basic and Applied Science, Galgotias University, Greater Noida 201312, Uttar Pradesh, India.

Correspondence: Anita Yadav, Associate Professor, Department of Forensic Science, School of Sciences, Sanjeev Agrawal Global Educational University, Bhopal 462022, Madhya Pradesh, India.

E-mail: anitakakas7@gmail.com

Received on: 08.08.2023

Accepted on: 29.10.2023

tables, charts, and other document components.[8-10] There are several reasons why a document could need to be changed. It would be immoral and may be unlawful to change facts in a document with the goal to deceive or mislead, for instance.

In the digital age, the proliferation of electronic devices and advanced image editing tools has given rise to concerns regarding the authenticity and integrity of visual information. Both offline scanned documents and digital images serve as critical components of modern communication, and their potential for manipulation has serious implications across various domains such as legal, forensic, historical, and artistic contexts.[11,12] As a result, there is a rising demand to create efficient and trustworthy techniques for identifying changes to these digital representations.

Histograms may be used to evaluate information, identify patterns, and acquire an understanding of how the data are distributed.[13,14] Using image histogram, the following study largely focuses on how documents are graphically represented as digital pictures, including offline scanned documents, and how to check for changes using the Image Histogram function of digital software named Irfan View. An effective tool for visualizing the tone distribution of an electronic image is the image histogram. It shows the frequency of the intensity value of each pixel, which in grayscale photographs generally ranges from 0 (black) to 255 (white).[15-18] The analysis of graphical representations of documents as digital photographs is one particular use of image histograms. This project's main objective is to convert offline, physical document scans into digital images that can be processed and examined using different software tools. This type of context calls for taking advantage of the "Image Histogram" function of the software Known as Irfan View. The histogram produced provides a graphical representation of the document's pixel intensity distribution.[19] Researchers may examine the readability and intelligibility of the scanned text overall, find out how often bright and dark tones are present in the document, and spot any potential picture quality problems by looking at the histogram. The image histogram can be used to identify changes or alterations in the scanned document. This is particularly useful for detecting forged or manipulated documents,[20] as any significant deviation from the expected histogram distribution may indicate tampering or modification.

METHODOLOGY

To study the comparative study of offline scanned documents and digital images for alteration using analysis of histograms, this comparative study was divided into two phases.

Phase-I, In Phase I, certain papers that are often utilized by forgers were chosen. The adjustments were created using GIMP software after the source documents were scanned at various DPIs.

Phase II, Image histograms of both original and modified documents will be used to compare histogram values in the online free software called Irfan View.

Collection of Sample

It involves working with a 10 scanned document (Mark sheet) that has been scanned at various DPI (dots per inch) settings, including 100 DPI, 200 DPI, 300 DPI, 400 DPI, and 500 DPI.

Manipulation of Sample

Using GIMP, the scanned document was opened, and a sample copy was created by selecting the desired DPI setting. Once the sample copy was generated, modifications could be made to the document. The modifications involved altering certain numerical values by increasing them, adjusting the spacing between numerical values, modifying grades, and even changing background colors. These changes were implemented using tools provided by GIMP, such as the cut and paste method.

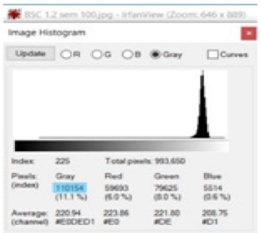
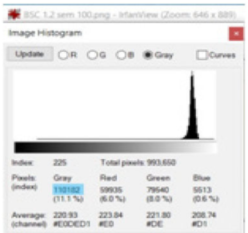
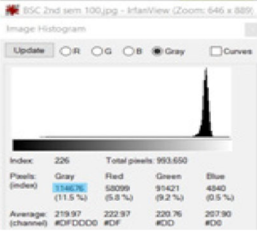
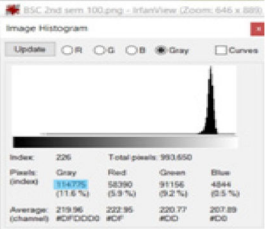
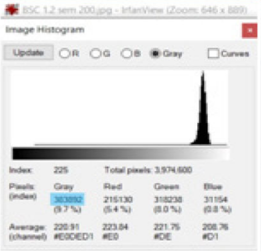
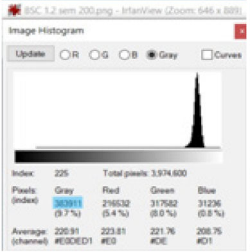
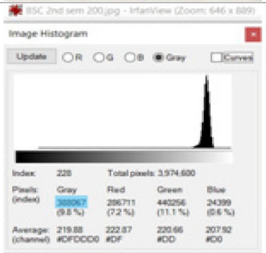
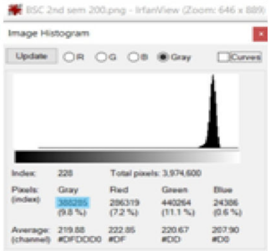
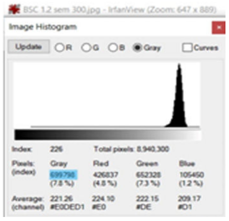
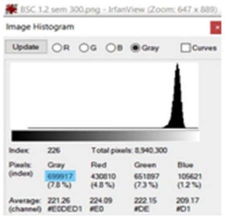
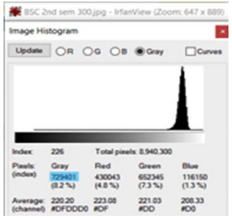
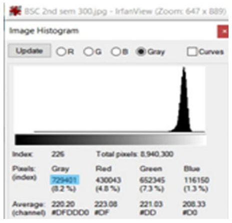
Analysis of Sample

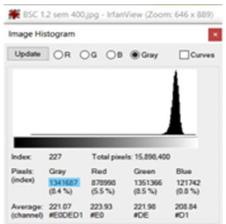
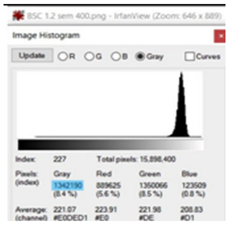
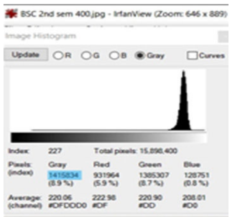
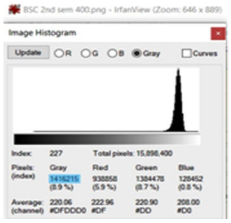
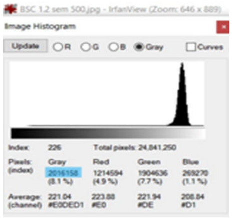
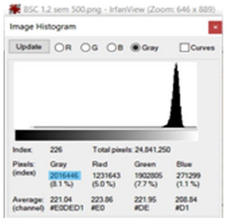
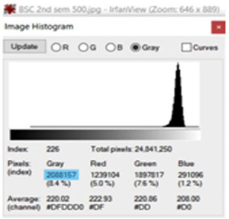
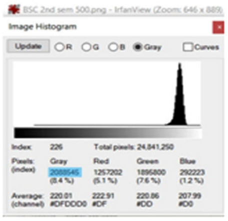
After that, the histogram value of the modified version of the document was checked using a free online program called Irfan View. By graphically displaying the distribution of pixel intensities within the image, the histogram demonstrated the tonal range and contrast of the picture. Once the histogram value of the altered document was obtained, the same process was repeated to check the histogram value of the original document. By comparing both the original and altered document's histogram values, it became possible to observe any changes that had occurred.

RESULT& DISCUSSIONS

The determination of histogram values for the allegedly changed picture and the supplied Standard image.

Table 1: Analysis of Standard and Altered Sample No. 1 and 2 at various DPI (***Source:*** Images provide by author)

	Sample 1		**Sample 2**	
DPI	**Standard**	**Altered**	**Standard**	**Altered**
100				
Result	 BSC 1.2 sem 100.jpg - IrfanView (Zoom: 646 x 889) Image Histogram Update ○ R ○ G ○ B ● Gray □ Curves Index: 225 Total pixels: 993,650 Pixels: (index) Gray 110154 (11.1 %) Red 59693 (6.0 %) Green 79625 (8.0 %) Blue 5514 (0.6 %) Average: (channel) 220.94 #E0DED1 223.86 #E0 221.80 #DE 208.75 #D1	 BSC 1.2 sem 100.png - IrfanView (Zoom: 646 x 889) Image Histogram Update ○ R ○ G ○ B ● Gray □ Curves Index: 225 Total pixels: 993,650 Pixels: (index) Gray 110182 (11.1 %) Red 59935 (6.0 %) Green 79540 (8.0 %) Blue 5513 (0.6 %) Average: (channel) 220.93 #E0DED1 223.84 #E0 221.80 #DE 208.74 #D1	 BSC 2nd sem 100.jpg - IrfanView (Zoom: 646 x 889) Image Histogram Update ○ R ○ G ○ B ● Gray □ Curves Index: 226 Total pixels: 993,650 Pixels: (index) Gray 114676 (11.5 %) Red 58099 (5.8 %) Green 91421 (9.2 %) Blue 4840 (0.5 %) Average: (channel) 219.97 #DFDDD0 222.97 #DF 220.76 #DD 207.90 #D0	 BSC 2nd sem 100.png - IrfanView (Zoom: 646 x 889) Image Histogram Update ○ R ○ G ○ B ● Gray □ Curves Index: 226 Total pixels: 993,650 Pixels: (index) Gray 114775 (11.6 %) Red 58390 (5.9 %) Green 91156 (9.2 %) Blue 4844 (0.5 %) Average: (channel) 219.96 #DFDDD0 222.95 #DF 220.77 #DD 207.89 #D0
200				
Result	 BSC 1.2 sem 200.jpg - IrfanView (Zoom: 646 x 889) Image Histogram Update ○ R ○ G ○ B ● Gray □ Curves Index: 225 Total pixels: 3,974,600 Pixels: (index) Gray 383892 (9.7 %) Red 215130 (5.4 %) Green 318238 (8.0 %) Blue 31154 (0.8 %) Average: (channel) 220.91 #E0DED1 223.84 #E0 221.75 #DE 208.76 #D1	 BSC 1.2 sem 200.png - IrfanView (Zoom: 646 x 889) Image Histogram Update ○ R ○ G ○ B ● Gray □ Curves Index: 225 Total pixels: 3,974,600 Pixels: (index) Gray 383911 (9.7 %) Red 216532 (5.4 %) Green 317582 (8.0 %) Blue 31236 (0.8 %) Average: (channel) 220.91 #E0DED1 223.81 #E0 221.76 #DE 208.75 #D1	 BSC 2nd sem 200.jpg - IrfanView (Zoom: 646 x 889) Image Histogram Update ○ R ○ G ○ B ● Gray □ Curves Index: 228 Total pixels: 3,974,600 Pixels: (index) Gray 388067 (9.8 %) Red 286711 (7.2 %) Green 440256 (11.1 %) Blue 24399 (0.6 %) Average: (channel) 219.88 #DFDDD0 222.87 #DF 220.66 #DD 207.92 #D0	 BSC 2nd sem 200.png - IrfanView (Zoom: 646 x 889) Image Histogram Update ○ R ○ G ○ B ● Gray □ Curves Index: 228 Total pixels: 3,974,600 Pixels: (index) Gray 388285 (9.8 %) Red 286319 (7.2 %) Green 440264 (11.1 %) Blue 24386 (0.6 %) Average: (channel) 219.88 #DFDDD0 222.85 #DF 220.67 #DD 207.90 #D0
300				
Result	 BSC 1.2 sem 300.jpg - IrfanView (Zoom: 647 x 889) Image Histogram Update ○ R ○ G ○ B ● Gray □ Curves Index: 226 Total pixels: 8,940,300 Pixels: (index) Gray 699798 (7.8 %) Red 426837 (4.8 %) Green 652328 (7.3 %) Blue 105450 (1.2 %) Average: (channel) 221.26 #E0DED1 224.10 #E0 222.15 #DE 209.17 #D1	 BSC 1.2 sem 300.png - IrfanView (Zoom: 647 x 889) Image Histogram Update ○ R ○ G ○ B ● Gray □ Curves Index: 226 Total pixels: 8,940,300 Pixels: (index) Gray 699917 (7.8 %) Red 430810 (4.8 %) Green 651897 (7.3 %) Blue 105621 (1.2 %) Average: (channel) 221.26 #E0DED1 224.09 #E0 222.15 #DE 209.17 #D1	 BSC 2nd sem 300.jpg - IrfanView (Zoom: 647 x 889) Image Histogram Update ○ R ○ G ○ B ● Gray □ Curves Index: 226 Total pixels: 8,940,300 Pixels: (index) Gray 729401 (8.2 %) Red 430043 (4.8 %) Green 652345 (7.3 %) Blue 116150 (1.3 %) Average: (channel) 220.20 #DFDDD0 223.08 #DF 221.03 #DD 208.33 #D0	 BSC 2nd sem 300.jpg - IrfanView (Zoom: 647 x 889) Image Histogram Update ○ R ○ G ○ B ● Gray □ Curves Index: 226 Total pixels: 8,940,300 Pixels: (index) Gray 729401 (8.2 %) Red 430043 (4.8 %) Green 652345 (7.3 %) Blue 116150 (1.3 %) Average: (channel) 220.20 #DFDDD0 223.08 #DF 221.03 #DD 208.33 #D0

400				
Result	 BSC 1.2 sem 400.jpg - IrfanView (Zoom: 646 x 889) Image Histogram Update ○R ○G ○B ●Gray □Curves Index: 227 Total pixels: 15,898,400 Pixels: (index) Gray 1341687 (8.4 %) Red 878998 (5.5 %) Green 1351366 (8.5 %) Blue 121742 (0.8 %) Average: (channel) 221.07 #E0DED1 223.93 #E0 221.98 #DE 208.84 #D1	 BSC 1.2 sem 400.png - IrfanView (Zoom: 646 x 889) Image Histogram Update ○R ○G ○B ●Gray □Curves Index: 227 Total pixels: 15,898,400 Pixels: (index) Gray 1342190 (8.4 %) Red 889625 (5.6 %) Green 1350066 (8.5 %) Blue 123509 (0.8 %) Average: (channel) 221.07 #E0DED1 223.91 #E0 221.98 #DE 208.83 #D1	 BSC 2nd sem 400.jpg - IrfanView (Zoom: 646 x 889) Image Histogram Update ○R ○G ○B ●Gray □Curves Index: 227 Total pixels: 15,898,400 Pixels: (index) Gray 1415834 (8.9 %) Red 931964 (5.9 %) Green 1385307 (8.7 %) Blue 128751 (0.8 %) Average: (channel) 220.06 #DFDDD0 222.98 #DF 220.90 #DD 208.01 #D0	 BSC 2nd sem 400.png - IrfanView (Zoom: 646 x 889) Image Histogram Update ○R ○G ○B ●Gray □Curves Index: 227 Total pixels: 15,898,400 Pixels: (index) Gray 1416215 (8.9 %) Red 938858 (5.9 %) Green 1384478 (8.7 %) Blue 128452 (0.8 %) Average: (channel) 220.06 #DFDDD0 222.96 #DF 220.90 #DD 208.00 #D0
500				
Result	 BSC 1.2 sem 500.jpg - IrfanView (Zoom: 646 x 889) Image Histogram Update ○R ○G ○B ●Gray □Curves Index: 226 Total pixels: 24,841,250 Pixels: (index) Gray 2016158 (8.1 %) Red 1214594 (4.9 %) Green 1904636 (7.7 %) Blue 269270 (1.1 %) Average: (channel) 221.04 #E0DED1 223.88 #E0 221.94 #DE 208.84 #D1	 BSC 1.2 sem 500.png - IrfanView (Zoom: 646 x 889) Image Histogram Update ○R ○G ○B ●Gray □Curves Index: 226 Total pixels: 24,841,250 Pixels: (index) Gray 2016446 (8.1 %) Red 1231643 (5.0 %) Green 1902805 (7.7 %) Blue 271299 (1.1 %) Average: (channel) 221.04 #E0DED1 223.86 #E0 221.95 #DE 208.84 #D1	 BSC 2nd sem 500.jpg - IrfanView (Zoom: 646 x 889) Image Histogram Update ○R ○G ○B ●Gray □Curves Index: 226 Total pixels: 24,841,250 Pixels: (index) Gray 2088157 (8.4 %) Red 1239104 (5.0 %) Green 1897817 (7.6 %) Blue 291096 (1.2 %) Average: (channel) 220.02 #DFDDD0 222.93 #DF 220.86 #DD 208.00 #D0	 BSC 2nd sem 500.png - IrfanView (Zoom: 646 x 889) Image Histogram Update ○R ○G ○B ●Gray □Curves Index: 226 Total pixels: 24,841,250 Pixels: (index) Gray 2088545 (8.4 %) Red 1257202 (5.1 %) Green 1896800 (7.6 %) Blue 292223 (1.2 %) Average: (channel) 220.01 #DFDDD0 222.91 #DF 220.86 #DD 207.99 #D0

The analysis focused on the histograms of the documents to identify any potential alterations or manipulations. Alterations in digital images were more diverse and often involved digital editing techniques. Common alterations included cropping, resizing, color adjustments, and image composite. Histogram analysis proved to be an effective method for detecting these alterations by revealing anomalies in the pixel value distribution. This could indicate that there was an increase in pixel intensity or a shift towards brighter or darker tones in the modified sample when compared to the reference sample. The analysis involved examining the reference (unaltered) sample and altered samples at various resolutions, including 100 DPI, 200 DPI, 300 DPI, 400 DPI, and 500 DPI. In each case, it was observed that the histogram values showed an increase, indicating that the alterations made to the samples had an impact on the distribution of pixel values throughout the document. This expanded analysis further reinforces the effectiveness of histogram analysis as a valuable tool for detecting alterations in both offline scanned documents and digital photographs. By comparing the histograms of the reference and altered samples at different resolutions, it becomes possible to identify changes in the distribution of pixel values and unveil potential manipulations.

Histogram analysis is particularly useful because it provides a visual representation of the frequency or count of pixel intensities within an image. This means that any alterations or modifications made to an image, whether intentional or unintentional, can be detected by examining the changes in the pixel value distribution. Any significant deviation from the reference sample's histogram can indicate potential tampering or unauthorized modifications.

The analysis highlights the importance of histogram analysis as a powerful technique for uncovering image manipulations. It provides forensic experts, researchers, and investigators with a reliable method to assess the authenticity and integrity of digital images and scanned documents. By leveraging histogram analysis, professionals can enhance their ability to detect alterations and ensure the credibility of visual evidence.

CONCLUSION

In this work, the comparison of offline scanned papers and digital pictures is used to look for changes using histogram analysis. Document authenticity faces a serious threat in the digital age from document tampering. The study tries to find differences between changed scanned documents and original digital pictures by using histogram analysis, which graphically depicts the distribution of pixel intensity. The major goal is to develop a trust worthy approach for identifying document changes by comparing histograms. Document changes may affect the histogram's pixel intensity distribution and leave recognizable traces. By carefully examining these histogram variances, the alterations made following these comparison studies may be quickly identified by contrasting histogram values.

REFERENCES

1. Math S, Tripathi RC. Digital forgeries: Problems and challenges. International Journal of Computer Applications. 2010 Aug;5(12):9-12.
2. Akhtar Z, Khan E. Revealing the traces of histogram equalisation in digital images. IET Image Processing. 2018 May;12(5):760-8.
3. Saini K, Kaur S. Forensic examination of computer-manipulated documents using image processing techniques. Egyptian Journal of Forensic Sciences. 2016 Sep 1;6(3):317-22.
4. Brin S, Davis J, Garcia-Molina H. Copy detection mechanisms for digital documents. In Proceedings of the 1995 ACM SIGMOD international conference on Management of data 1995 May 22 (pp. 398-409).
5. Bashir A, Fadlalla YA. Techniques of detecting forgery in identity documents. no. February. 2018.
6. Shabanian H, Mashhadi F. A new approach for detecting copy-move forgery in digital images. In2017 IEEE Western New York Image and Signal Processing Workshop (WNYISPW) 2017 Nov 17 (pp. 1-6). IEEE.
7. Al-Qershi OM, Khoo BE. Passive detection of copy-move forgery in digital images: State-of-the-art. Forensic science international. 2013 Sep 10;231(1-3):284-95.
8. Mahalakshmi SD, Vijayalakshmi K, Priyadharsini S. Digital image forgery detection and estimation by exploring basic image manipulations. Digital Investigation. 2012 Feb 1;8(3-4):215-25.
9. Cao Y, Gao T, Fan L, Yang Q. A robust detection algorithm for copy-move forgery in digital images. Forensic science international. 2012 Jan 10;214(1-3):33-43.
10. Kashyap A, Parmar RS, Agrawal M, Gupta H. An evaluation of digital image forgery detection approaches. arXiv preprint arXiv:1703.09968. 2017 Mar 29.
11. Sameria S, Saran V, Gupta AK. Analysis of offline scanned document and digital images for alteration through digital image processing.
12. Kumar A, Kansal A, Singh K. An improved anti-forensic technique for JPEG compression. Multimedia Tools and Applications. 2019 Sep 30;78(18):25427-53.
13. Santhi K, Banu RW. Adaptive contrast enhancement using modified histogram equalization. Optik-International Journal for Light and Electron Optics. 2015 Oct 1;126(19):1809-14.
14. Kanetkar S, Pathania A, Venugopal V, Sundaram S. Offline writer identification using local derivative pattern. In 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR) 2016 Oct 23 (pp. 355-360). IEEE.
15. Kaur A, Saran V, Gupta AK. Digital image processing for forensic analysis of fabricated documents. Digital Image Processing. 2014 Sep;1(2).
16. Nuzzo RL. Histograms: A useful data analysis visualization. PM & R. 2019 Mar;11(3):309-12.
17. Ho AT, Wang K, Cayre F. An effective histogram-based approach to JPEG-100 forensics. In 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA) 2016 Dec 12 (pp. 1-6). IEEE.
18. Nikolova M, Wen YW, Chan R. Exact histogram specification for digital images using a variational approach. Journal of Mathematical Imaging and Vision. 2013 Jul;46(3):309-25.
19. Anagha PH, Baskar A. An automatic histogram detection and information extraction from document images. International Journal of Speech Technology. 2021 March;24:77-85.
20. Stamm MC, Liu KR. Forensic detection of image manipulation using statistical intrinsic fingerprints. IEEE Transactions on Information Forensics and Security. 2010 Jun 17;5(3):492-506.