A Video Forensic Technique for Detecting Frame Integrity Using Human Visual System-inspired measure

Qianwen Wan, Karen Panetta, *Fellow, IEEE*
Department of Electrical and Computer Engineering
Tufts University
Medford, MA, USA
Qianwen.Wan@tufts.edu, karen@ece.tufts.edu

Sos Agaian, *Fellow, IEEE*
Department of Electrical and Computer Engineering
University of Texas at San Antonio
San Antonio, TX, USA
sagaian@utsa.edu

***Abstract*— Digital videos have been widely used for security purposes; hence, a significant research effort has been devoted to develop video forensic technology. Forensic video analysis aims to solve two main problems: (1) finding evidence present in a video, and (2) authenticating the original video source. In this paper, we propose an automatic jump-cut detection system to evaluate video altering and tampering using a novel, low-cost and accurate video forensic technique using Human Visual System-inspired measure, which can detect alterations that the human eye may not be able to perceive. Our method is intended to qualify the integrity of digital video content. The experimental results demonstrate that our measurement is able to perform a reliable identification and authorization for digital video forensic applications.**

Keywords — digit videos; video forensic technique; Human Visual System-inspired measure

I. INTRODUCTION

Digital videos have been widely used for forensic uses; however, there are many existing challenges. Two main challenges are: 1) poor quality videos make it difficult to capture and extract reliable evidence; 2) the commercial availability of image processing tools make it easy for forgers to manipulate the video contents that can be seemingly undetectable by the human eye. This can result in severe consequences when the digital content is used as legal evidence [1].

The goal of digital forensics is to detect whether information has been maliciously modified or erased from the original recorded scene, and to collect evidence[2, 3].

Researchers have focused on employing video forensic tools for authentication purposes; for example: (A) Kurosawa et al. has introduced a forensic tool that includes identification of acquisition devices in [3]. Studies suggest that the illegal reproduction of videos can be detected by: (i) detecting re-projected videos (for instance, active watermarking approach etc.[4] [5]); (ii) providing video retrieval techniques based on device fingerprinting [6]. (B) Uncompressed motion pictures are usually compressed into a lossy format because of the large bit rate of the video content[2]. This process has made modification of the video content possible. Hence, the advance of modern video compression techniques allows three aspects to be explored and studied: (i) Video coding parameter identification [7] [8, 9]; (ii) Video re-encoding [10, 11]; (iii) Network footprints identification[12]. (C) Videos are extensively used for surveillance worldwide, at the same time, the availability of video editing suites means that the video doctoring detection tools must consider many source scenarios: (i) Camera based editing detection [13] [14, 15]; (ii) Detection based on coding artifacts[16]; (iii) Detection based on inconsistencies in content[17]; (iv) Copy-move detection in videos[18].

In this paper, we focus on detecting inconsistencies in video content using a human visual system inspired image similarity measurement to find cuts or deletions that are not readily apparent. To our knowledge, though people have been using traditional image similarity measurements for shot transition detection or jump-cut detection in film or video post-production, however our literature search failed to find video forensic techniques using human visual system inspired image similarity measurement. Furthermore, the available software available either through openCV or commercially that do find jump cuts, cannot readily detect cuts that are non-obvious to the human eye.

An image similarity measure using enhanced human visual system characteristics introduced by Nercessian, Agaian, and Panetta in [19] will be utilized for jump-cut detection. Because HVS inspired image similarity measurement has shown strong performance compared to other existing similarity quality metrics for blurred images and noisy images. It has the ability to correlate better with subjective human evaluation for video processing. [19].

The remainder of this paper is structured as follows. Section II introduces related work on HVS inspired image similarity measurement. The proposed automatic jump-cut detection technique for detecting frame integrity is presented in Section III. In Section IV, experiments results are shown. Finally, Section V summarizes our contributions and lists the directions of future work.

978-1-5090-6356-7/17/$31.00 ©2017 IEEE

II. RELATED WORK

Image quality measures are crucial for image de-noising, compression, steganography, and other image processing applications, which require comparison to an ideal reference image to quantitatively assess algorithm performance [20].

The baseline objective image similarity measures are simple pixel-based distance metrics, such as the mean-squared error (MSE) and peak signal-to-noise ratio (PSNR) [21]. Equation (1) calculates the mean square error between two images x and y; and equation (2) defines the PSNR between two images in dB.

$$MSE(x,y) = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}[x(m,n) - y(m,n)]^2 \qquad (1)$$

$$PSNR(x,y) = 10 \cdot log_{10}\left(\frac{{MAX_x}^2}{MSE}\right) \qquad (2)$$

Here, MAX_x is the maximum possible pixel value of the image x. These measures are straightforward, easy to implement, and have a clear physical and mathematical interpretation. However, they do not correlate closely with human perception[22].

Because the human visual system is adapted to extract local structural information; the structural similarity index (SSIM) [20] has been developed, which focuses on local information, such as luminance, contrast, and structure. However, the SSIM index does not have good performance for blurry images. As a result, a variant of the SSIM index, the Gradient Structural Similarity index (GSSIM), was proposed [23]. The GSSIM index improves the performance of the traditional SSIM for badly blurred images by also considering the gradients of the input images. Accordingly, the 4-component SSIM (4-SSIM) and 4-component GSSIM (4-GSSIM) were proposed. These variants of the SSIM index combine local similarity measures with dynamic weights specified for changed edge, preserved edge, texture, and smooth regions[19], which can provide a better performance.

In this section, we explain a natural progression from the SSIM index to the 4-GSSIM index; in order to introduce human visual system (HVS) inspired 4-EGSSIM and its advantages by following the evolution of similarity measures.

A. SSIM index

The SSIM metric measures image similarity in terms of local luminance, contrast, and structure[20]. For two images, x and y, the luminance I, contrast C, and structure S for a window centered at pixel location (m,n) are given, respectively, by

$$I_{x,y}(i,j) = \frac{2\mu_x(i,j)\mu_y(i,j) + C_1}{\mu_x(i,j)^2 + \mu_y(i,j)^2 + C_1} \qquad (3)$$

$$c_{x,y}(i,j) = \frac{2\sigma_x(i,j)\sigma_y(i,j) + C_2}{\sigma_x(i,j)^2 + \sigma_y(i,j)^2 + C_2} \qquad (4)$$

$$s_{x,y}(i,j) = \frac{\sigma_{xy}(i,j) + C_3}{\sigma_x(i,j)\sigma_y(i,j) + C_3} \qquad (5)$$

Where μ_x and μ_y are the sample means of x and y; $\sigma_x$and σ_y are the sample standard deviations of x and y; σ_{xy} is the sample covariance of x and y; and C_1, C_2, and C_3 are small constants included to provide stability for small denominators. The sample statistics are calculated using an 11x11 window and a Gaussian window profile[19]. The SSIM map is given by,

$$SSIM_{x,y}(m,n) = [I_{x,y}(m,n)]^{\alpha}[c_{x,y}(m,n)]^{\beta}[s_{x,y}(m,n)]^{\gamma} \qquad (6)$$

Here, α, β, γ are parameters we choose, in common implementation, C1 = C2, C3 = C1/2, and $\alpha = \beta = \gamma = 1$. And the mean SSIM index is given by,

$$SSIM(x,y) = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} SSIM_{x,y}(m,n) \qquad (7)$$

B. GSSIM index

The GSSIM index was proposed by Chen et. al [24] upon noting that the performance of the SSIM index degraded substantially when assessing Gaussian blurred images. The contrast and structure term of SSIM are modified by

$$c_{x',y'}(i,j) = \frac{2\sigma_{x'}(i,j)\sigma_{y'}(i,j) + C_2}{\sigma_{x'}(i,j)^2 + \sigma_{y'}(i,j)^2 + C_2} \qquad (8)$$

$$s_{x',y'}(i,j) = \frac{\sigma_{x'y'}(i,j) + C_3}{\sigma_{x'}(i,j)\sigma_{y'}(i,j) + C_3} \qquad (9)$$

Where, $\sigma_{x'}$and $\sigma_{y'}$ are the sample standard deviations of x' and y', respectively, $\sigma_{x'y'}$ is the sample covariance of x' and y'; and x' and y' are corresponded to x and y calculated using Sobel edge detector. Then the GSSIM map is given by,

$$GSSIM_{x,y}(m,n) = \left[I_{x,y}(m,n)\right]^{\alpha}[c_{x',y'}(m,n)]^{\beta}\left[s_{x',y'}(m,n)\right]^{\gamma} \qquad (10)$$

And the mean SSIM index is given by,

$$GSSIM(x,y) = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} GSSIM_{x,y}(m,n) \qquad (11)$$

C. 4-SSIM and 4-GSSIM indexes

The 4-SSIM and 4-GSSIM attempt to further improve the SSIM and GSSIM in order to provide a higher consistency with human subjective judgment of blurry and noisy images[25]. The images are thresholded into four regions, preserved edge pixel region, change edge pixel region, smooth region and texture region. The 4 regions are determined as follows:

Preserved edge pixel region (R_1)

$$x'(m,n) > T_1 \ AND \ y'(m,n) > T_1 \qquad (12)$$

Change edge pixel region (R_2)

$$(x'(m,n) > T_1 \ AND \ y'(m,n) \le T_1) \ OR \ (y'(m,n) > T_1 \ AND \ x'(m,n) \le T_1) \qquad (13)$$

Smooth region (R_3)

$$(x'(m,n) > T_1 \ AND \ y'(m,n) \le T_1) \ OR \ (y'(m,n) > T_2 \ AND \ x'(m,n) \le T_2) \qquad (14)$$

Texture Region (R_4)

Otherwise (15)

Here, the thresholding parameters $T_1 = 0.12(x'_{max})$, and $T_2 = 0.06(x'_{max})$ are calculated, where x'_{max} is the maximum value of the gradient magnitude of the referenced image. Meanwhile, $x'(m,n)$ is the gradient that calculated on the referenced image, and $y'(m,n)$is the value of the gradient magnitude of compared image.

The 4-SSIM is determined by calculating the average in each region, given as

$$SSIM_{R_1}(x,y) = \frac{1}{|R_1|}\sum_{(m,n)\in R_1} SSIM_{x,y}(m,n) \quad (16)$$

$$SSIM_{R_2}(x,y) = \frac{1}{|R_2|}\sum_{(m,n)\in R_2} SSIM_{x,y}(m,n) \quad (17)$$

$$SSIM_{R_3}(x,y) = \frac{1}{|R_3|}\sum_{(m,n)\in R_3} SSIM_{x,y}(m,n) \quad (18)$$

$$SSIM_{R_4}(x,y) = \frac{1}{|R_1|}\sum_{(m,n)\in R_4} SSIM_{x,y}(m,n) \quad (19)$$

Also, the mean 4-SSIM is then determined by,

$$4-SSIM(x,y) = \sum_{i=1}^{4} w_i SSIM_{R_i}(x,y) \quad (20)$$

Where w_i are weights for each region R_i.

In[25], the weights were chosen such that w3 = w4 = 0.25. Also, w2 = 0.5 if R_1 is an empty set. We set w1 = 0.5, and w1 = w2 = 0.25 and if R_2 is an empty set. The 4-GSSIM is calculated in a similar fashion.

D. 4-EGSSIM index (The image similarity measure using enhanced human visual system characteristics)

In [19] , Nercessian, Agaian, and Panetta pointed out that, the more the salient features of the images are stressed, the better the performance of the quality metric is.

After analyzing the evolution of the SSIM index and its variants, they performed dynamic range compression on the image gradient using a simple pixel-by-pixel transformation.

In this case, a logarithmic transformation was added, Namely, the gradient images are enhanced by

$$x'_{Enh} = \log(x'+1) \quad (21)$$

$$y'_{Enh} = \log(y'+1) \quad (22)$$

The enhanced gradient magnitudes are used in calculating the local similarity terms, while the standard gradient magnitudes are used to determine the image regions. The schematic of 4-EGSSIM is in Fig.1.

III. PROCEDURE OF THE AUTOMATIC VIDEO FORENSIC TECHNIQUE FOR DETECTING FRAME INTEGRITY

A. Testing Database

Since there is no specific public database for testing videos with disconnected frames, this research created video incoherence using the SULFA database[26]. More specifically tested was the movie *AWAKE: The life of Yogananda* [27], and original videos filmed in the Dr. Panetta's Vision & Sensing System Lab.

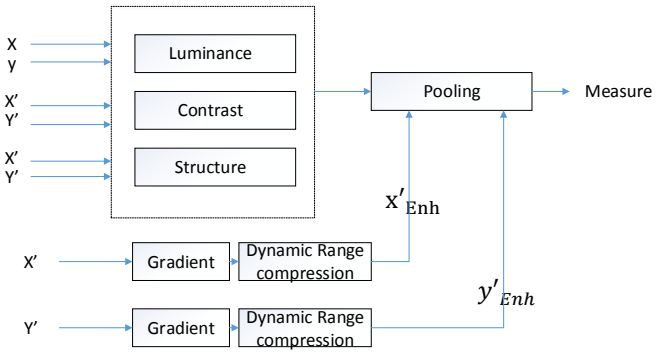
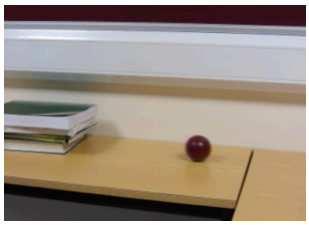


Fig. 1 Schematic of 4-EGSSIM image quality measure.

SULFA is a freely available database for through the University of Surrey website for the purpose of forensic research via http://sulfa.cs.surrey.ac.uk/.

A wide variety of video scenes have been collected. For example, Fig.2 (a) and (b) show two frames of video shot indoors, with a Canon SX220, of a rolling red ball; Fig.2(c) and (d) show two frames of video shot outdoors, with FUJIFILM 2800, of a busy street view. Both sets of figures are from the SULFA database. Fig.2 (e) and (f) show two frames of a motion detection video database, with a low-quality camera and non-processed noise; Fig.2 (g) and (h) show two frames with missing frame(s) between, also called a jump-cut, which is mistakenly edited between two shots. Fig.2 (i) and (j) show the two frames that we filmed using a mobile camera.

(a)



(b)

(c)

(d)

(e)

(f)

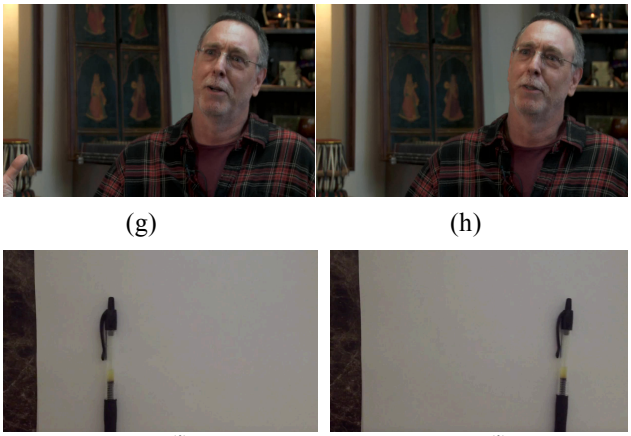


Fig.2. Example Database for testing the Video Forensic Technique for Detecting Frame Integrity Using Human Visual System-inspired measure; (a)-(d) are videos from SULFA database; (e) and (f) are from a motion detection video database, with a low-quality camera and non-processed noise; (g) and (h) Movie frames from the film documentary *AWAKE: The life of Yogananda* with subtle jump-cut; (i) and (j) are from the video database we filmed using a mobile camera.

B. *An automatic video forensic technique for detecting frame integrity*

Fig.3 below shows the flow diagram of the automatic frame integrity detection system for the 4-EGSSIM video forensic technique. First, the video is input, which is then separated into individual frames. The next step is to send neighboring frames into the 4-EGSSIM procedure to calculate an image similarity assessment value based on the human visual system. Finally, by finding the smallest 4-EGSSIM-index value, the system can detect the altered frames.

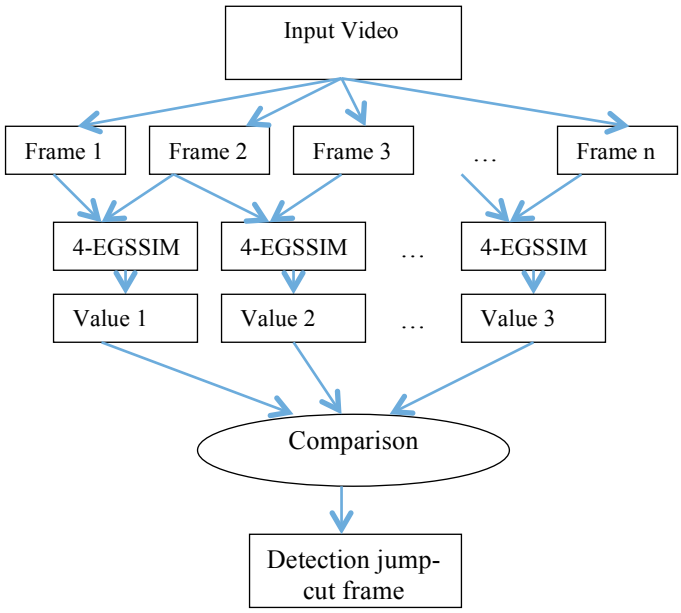


Fig. 3 Flow Chart of the video forensic technique using Human Visual System inspired image quality measure

IV. Experimental Result

In this section, we present the experimental results in order to show that our video forensic technique for detecting frame integrity using human visual system-inspired measure is accurate, and low-cost.

Below, some example results are shown from testing the database introduced in section III. Fig. 4 was the result of the pen-shifting movie, which is filmed using an iPhone. We manually removed only one frame before testing. The human eye can barely detect the disconnection, however, our video forensic method could easily detect the disconnection by finding the smallest value. Fig.5 is the result of a car movement video captured by a low-quality camera, which contained a lot of noise. We manually removed only five frames out of 529 frames before testing. Regardless of the large amount of noise, our method can realize a successful detection. These results show that our forensic technique can be used to detect the integrity of legal evidences.

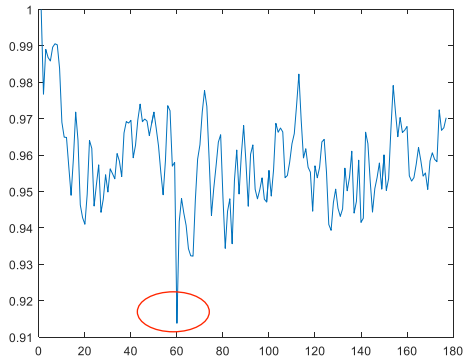


Fig.4. Detecting result of pen-shifting movie (manually remove one fame out of 180 frames.); the red circle of peak value indicated the jump cut

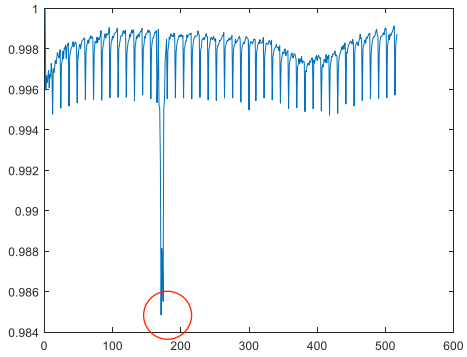


Fig. 5. Detecting result of parking lot car movement movie (manually remove 5 fames out of 529 frames.); the red circle peak value indicated the jump cut

978-1-5090-6356-7/17/$31.00 ©2017 IEEE

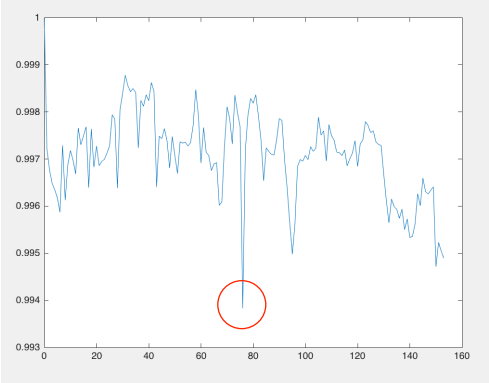


Fig. 6. Detecting result of Krishadas movie; the red circle of peak value indicated the jump cut between 76th frame and 77th frame

Fig.6 is a practical example of the movie *AWAKE: The life of Yogananda* [27] with a subtle disconnection created by editing two shots of a film together. In the movie industry, manually detecting and editing disconnections usually do the film post-production process. Therefore, as another application, our forensic tool can be used to reduce human labor.

Below is a table that presents the detection rate using our proposed video forensic tool using human visual system (HVS) inspired image quality measurement. By comparing the results, HAS inspired 4-EGSSIM can provide a better jump-cut detection for video forensic.

TABLE I. DETECTION RATE USING OUR HUMAN VISUAL SYSTEM INSPIRED IMAGE QUALITY MEASUREMENT AND COMPARISON WITH OTHER BASELINE IMAGE SIMILARITY MEASUREMENT

Testing Video	**SULFA video forensic database**			**Video database with Low-quality camera**			**Video filmed by iPhone**		**Movie**
Remove Frames	***1***	***5***	***10***	***1***	***5***	***10***	***1***	***5***	***0***
HVS Detection rate (%)	90	100	100	75	78	90	95	100	100
MSE Detection rate (%)	40	50	75	25	40	50	0	0	0
PSNR Detection rate (%)	40	50	75	25	35	50	0	0	0
SSIM Detection rate (%)	75	80	90	25	50	50	75	90	100

V. CONCLUSION

In this paper, a video forensic tool using a Human Visual System-inspired measure is presented to address the challenge of detecting video integrity, in a low-cost, accurate and efficient manner. The availability of inexpensive digital multimedia devices and the high quality of data processing tools have made the video forensic topic more and more popular and crucial. Although there are existing open source software, using traditional image similarity measurements for shot transition detection or jump-cut detection, our literature search failed to find any video forensic techniques using human visual system inspired image similarity measurement. We believe our quality measurement based video forensic system is valuable asset that can be utilized in applications performing identification and authorization tasks for video forensics. Furthermore, it will aid in improving existing video forensic tools aimed at detecting alteration and tampering.

In the future, improvements will be made to the autonomous prototype of the human visual system inspired image quality measure by adding brightness and colorfulness modules to extract more information contained by the videos. Moreover, we will integrate the proposed forensic tool in practical computer-aided vision systems.

REFERENCES

[1] H. Farid, "Exposing digital forgeries in scientific images," presented at the Proceedings of the 8th workshop on Multimedia and security, Geneva, Switzerland, 2006.

[2] P. Bestagini, K. M. Fontani, S. Milani, M. Barni, A. Piva, M. Tagliasacchi, *et al.*, "An overview on video forensics," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, 2012, pp. 1229-1233.

[3] K. Kurosawa, K. Kuroki, and N. Saitoh, "CCD fingerprint method-identification of a video camera from videotaped images," in *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, 1999, pp. 537-540 vol.3.

[4] J.-W. Lee, M.-J. Lee, T.-W. Oh, S.-J. Ryu, and H.-K. Lee, "Screenshot identification using combing artifact from interlaced video," presented at the Proceedings of the 12th ACM workshop on Multimedia and security, Roma, Italy, 2010.

[5] M.-J. Lee, K.-S. Kim, and H.-K. Lee, "Digital Cinema Watermarking for Estimating the Position of the Pirate," *Trans. Multi.*, vol. 12, pp. 605-621, 2010.

[6] "Proceedings of the 1st ACM international conference on Multimedia information retrieval," Vancouver, British Columbia, Canada, 2008, p. 476.

[7] Z. G. Fan and R. L. de Queiroz, "Identification of bitmap compression history: JPEG detection and quantizer estimation," *Ieee Transactions on Image Processing*, vol. 12, pp. 230-235, Feb 2003.

[8] M. Tagliasacchi and S. Tubaro, "Blind estimation of the QP parameter in H.264/AVC decoded video," in *Image Analysis for Multimedia Interactive Services (WIAMIS), 2010 11th International Workshop on*, 2010, pp. 1-4.

[9] G. Valenzise, M. Tagliasacchi, and S. Tubaro, "Estimating QP and motion vectors in H.264/AVC video from decoded pixels," presented at the Proceedings of the 2nd ACM workshop on Multimedia in forensics, security and intelligence, Firenze, Italy, 2010.

[10] D. Fu, Y. Q. Shi, and W. Su, "A generalized Benford's law for JPEG coefficients and its applications in image forensics," in *Electronic Imaging 2007*, 2007, pp. 65051L-65051L-11.

[11] S. Milani, M. Tagliasacchi, and S. Tubaro, "Discriminating multiple JPEG compressions using first digit features," *APSIPA Transactions on Signal and Information Processing*, vol. 3, p. e19, 2014.

[12] A. R. Reibman and D. Poole, "Characterizing packet-loss impairments in compressed video," in *ICIP (5)*, 2007, pp. 77-80.

[13] N. Mondaini, R. Caldelli, A. Piva, M. Barni, and V. Cappellini, "Detection of malevolent changes in digital video for forensic applications," in *Electronic Imaging 2007*, 2007, pp. 65050T-65050T-12.

[14] C.-C. Hsu, T.-Y. Hung, C.-W. Lin, and C.-T. Hsu, "Video forgery detection using correlation of noise residue," in *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*, 2008, pp. 170-174.

[15] M. Kobayashi, T. Okabe, and Y. Sato, "Detecting forgery from static-scene video based on inconsistency in noise level functions,"

Information Forensics and Security, IEEE Transactions on, vol. 5, pp. 883-892, 2010.

[16] W. Wang and H. Farid, "Exposing digital forgeries in interlaced and deinterlaced video," *Information Forensics and Security, IEEE Transactions on,* vol. 2, pp. 438-449, 2007.

[17] J. Zhang, Y. Su, and M. Zhang, "Exposing digital video forgery by ghost shadow artifact," in *Proceedings of the First ACM workshop on Multimedia in forensics*, 2009, pp. 49-54.

[18] W. Wang and H. Farid, "Exposing digital forgeries in video by detecting duplication," in *Proceedings of the 9th workshop on Multimedia & security*, 2007, pp. 35-42.

[19] S. Nercessian, S. S. Agaian, and K. A. Panetta, "An image similarity measure using enhanced human visual system characteristics," in *SPIE Defense, Security, and Sensing*, 2011, pp. 806310-806310-9.

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on,* vol. 13, pp. 600-612, 2004.

[21] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? A new look at signal fidelity measures," *Signal Processing Magazine, IEEE,* vol. 26, pp. 98-117, 2009.

[22] E. Wharton, K. Panetta, and S. Agaian, "Human visual system based similarity metrics," in *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on*, 2008, pp. 685-690.

[23] G.-H. Chen, C.-L. Yang, and S.-L. Xie, "Gradient-based structural similarity for image quality assessment," in *Image Processing, 2006 IEEE International Conference on*, 2006, pp. 2929-2932.

[24] G. h. Chen, C. l. Yang, and S. l. Xie, "Gradient-Based Structural Similarity for Image Quality Assessment," in *2006 International Conference on Image Processing*, 2006, pp. 2929-2932.

[25] C. Li and A. C. Bovik, "Content-partitioned structural similarity index for image quality assessment," *Signal Processing: Image Communication,* vol. 25, pp. 517-526, 2010.

[26] G. Qadir, S. Yahaya, and A. T. Ho, "Surrey university library for forensic analysis (SULFA) of video content," in *Image Processing (IPR 2012), IET Conference on*, 2012, pp. 1-6.

[27] . http://www.awaketheyoganandamovie.com.

978-1-5090-6356-7/17/$31.00 ©2017 IEEE