COMPARISON OF HUMAN AND MACHINE PERFORMANCE FOR COPY-MOVE IMAGE FORGERY DETECTION INVOLVING SIMILAR BUT GENUINE OBJECTS

Ye Zhu[1], *Ramanathan Subramanian*[2], *Tian-Tsong Ng*[3], *Stefan Winkler*[2], *Rama Ratnam*[2]

[1] College of Computer Science and Technology, Jilin University, Changchun, China
[2] Advanced Digital Sciences Center (ADSC), University of Illinois at Urbana-Champaign, Singapore
[3] Situational Awareness Analytics, Institute for Infocomm Research (I2R), Singapore

ABSTRACT

Copy-move forgery (CMF) is considered easier to detect than general forgery mechanisms, but detecting it in the presence of multiple similar but genuine scene objects (SGOs) is non-trivial. We study the efficacy of human visual perception for copy-move image forgery detection (CMFD) involving SGOs, and compare the same with machine performance. Via an eye tracking study performed with 16 users where pairs of images (one *real* and the other *tampered*) were displayed in either parallel or serial fashion, we make the following observations: (1) Forgery detection is quicker and more accurate when images are spatially aligned and presented serially, so that the tampering is conspicuous. (2) Eye fixations focus on corresponding regions of the real and tampered images, with fewer and more localized fixations noted during serial comparison. (3) A gap is noted between CMFD performance of humans and machines, with each being more sensitive to different tampering factors. Overall, results reveal the need for systematic visual comparisons to distinguish SGOs from forged objects, as well as the promise of a human-machine collaborative framework to this end.

1. INTRODUCTION

With the widespread popularity of image editing tools, a large number of photographs available on the Internet could be tampered. Image tampering detection mostly relied on human experts until digital forensics devised tamper detection tools [4]. Despite a decade of forensics research, how well computers perform with respect to humans for tamper detection remains unclear with few works examining human factors [2]. In most visual tasks, human performance is an optimistic benchmark for computational methods [3]. However, tamper detection is different as image forgeries are mainly designed to spoof human detection. Second, computers can analyze visual cues that are inconspicuous to human vision [9]. Therefore, comparing human and computer performance offers insights toward digital forensics through human-computer cooperation.

Two main goals of this work are to conduct human performance evaluation via user response and eye-movement behavior as in [6, 12], and a human-computer comparative study for copy-move forgery detection (CMFD). Copy-move forgery (CMF) is a common image tampering technique, where an image region is copied, manipulated graphically and pasted elsewhere in the same image. CMF is considered as a simpler instance of general forgery, as the tampering source resides in the same image; however, CMFD in the presence of multiple similar-but-genuine scene objects (SGOs) can be highly challenging (Fig. 1).

In order to study whether humans are adept at discriminating *original* from *forged* objects under such conditions, we performed a study with 16 users and 60 original-forged image pairs. Users' viewing behavior was monitored using an eye-tracker as they performed CMFD. Two comparison methods were investigated, where each image pair was shown (a) side-by-side or (b)spatially aligned in a temporally serial fashion. We then examined human and machine CMFD performance for attributes such as *scale*, *rotation*, *illumination*, *distortion* and a *combination* of these. Machine performance is evaluated via the state-of-the-art CMFD method [7], and its variant in a similar setting as human experiments. Experiments show that machine-based CMFD methods are adept at detecting rotation, scale and naive CMF, but are ineffective for distortion or illumination-based tampering which are efficiently detected by humans.

The main contributions of this paper can be summarized as follows: (1) We expressly compare human and automated CMFD performance. Humans and machines have complementary expertise at detecting various types of tampering, which highlights the need for better CMFD algorithms, and the promise of a collaborative CMFD framework. (2) We objectively examine the influence of viewing behavior and tampering scheme on CMFD performance. Systematic comparison of corresponding image regions is found to facilitate CMFD with SGOs, and human detection improves considerably when the original-forged images are spatially aligned and viewed serially so that tampering is conspicuous.

The rest of the paper is organized as follows: Section 2 describes the dataset and experiment, Section 3 discusses the results, and Section 4 concludes with key observations.

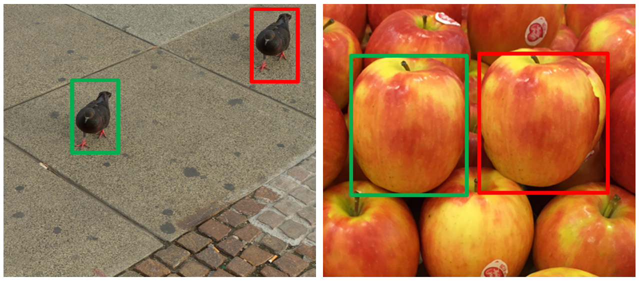

Fig. 1. Exemplar forged images from the CoMoFod [13] (left) and our dataset (right). The tampering is naive CMF, where a *natural* object (green) is duplicated to synthesize the *forged* region (red).

2. EXPERIMENT DESIGN

2.1. Dataset and Participants

54 image pairs, in each of which one image was *natural* and the other copy-move *tampered*, were used in our study. Tampering was achieved via: *rotation* (8 pairs), *scaling* (8 pairs), *illumination change* (9 pairs), *distortion* (8 pairs), *naive CMF* (9 pairs) and *combination* (9 pairs). Three panoramic image pairs were also included to explore human sensitivity in an elongated scene. To ensure viewers remained attentive, six pairs (10% of the total) comprising identical images were also shown during the experiment, but not used in the data analysis.

Exemplar image pairs used in our study are shown in Figs. 1,2. The images used in this study are unique and more challenging with respect to prior datasets, as they contain *at least one* natural object similar to the forged one (other datasets contain exactly one copy). Also, among CMFD datasets, only CoMoFod [13] considers tampering factors other than rotation, scale and naive CMF (it nevertheless does not consider illumination changes). 16 volunteers (6 female) aged 18–36 years (mean 26.6 ± 5.5) took part in the study.

2.2. Experimental Setup

The stimulus presentation protocol was developed using Matlab *Psychtoolbox*. Upon viewing each image pair, viewers were required to determine *'Which image is real?'* in a two-alternative forced choice (2AFC) design. Viewers had a maximum of 40 seconds to decide, failing which, the protocol proceeded to display the next image pair. Also, to examine whether comparison methodology impacts human CMFD, pairs were displayed in both *parallel* and *serial* formats – each format was viewed by a group of eight users. The real and tampered images were shown side-by-side in the parallel format (Fig. 2, cols 3-4), whereas they were spatially aligned and shown one-after-another in the serial format (Fig. 2, cols 5-6); here, users were able to move back-and-forth between the two images and indicate their selection using the keyboard.

Even though real-life tamper detection requires the human/machine to make a decision upon viewing a single image with no reference, we employed a comparison task in the user study owing to the following reasons: (1) CMFD has traditionally relied on key point [1] and block-based [10] matching strategies, which employ the number of point/block matches as a measure of CMF. When multiple SGOs exist in the original, this assumption breaks down necessitating the need for sophisticated CMFD strategies. (2) The larger aim of this work was to understand human search, (original and forged) object comparison and forgery detection mechanisms (in terms of eye movements and neural responses) so as to inspire similar automated approaches.

As viewers performed CMFD, their eye movements were recorded using an *Eyetribe* eye-tracker. Viewers sat at a distance of approximately 60 cm in front of a 14-inch screen with a resolution of 1366×768 pixels. The eye-tracker has a sampling rate of 60 Hz and is accurate to within 0.5^o visual angle upon calibration. All keyboard events were logged. The experiment lasted one hour, and the eye-tracker was recalibrated after every 15 image pairs to minimize drift. To eliminate systematic bias, display order of the image pairs and relative position of each image (left/right in parallel, and 1st/2nd in serial) was randomized across users.

3. DATA ANALYSIS

3.1. Parallel vs. Serial Comparison

To examine the influence of image comparison strategy for CMFD, we compiled the following statistics for each viewer in the serial and parallel presentation formats: detection accuracy (*i.e.*, % of times real image was detected), cumulative viewing frequency (CVF, *i.e.*, sum total of the number of times each image was viewed), and selection time (ST).

Mean and standard deviation for each of these measures is presented in Table 1. Detection accuracy in both serial and parallel presentation formats are well above chance, implying that humans are generally adept at CMFD via comparison. However, detection accuracy in the serial mode is considerably higher than in parallel and close to ceiling – this is because identifying differences is much easier when two images are spatially aligned and viewed one-by-one. Photo triaging studies [5, 8] have observed that human visual attention is sensitive on local scene changes while browsing through a sequence of spatially aligned images.

A two-sample t-test confirmed a highly significant difference in detection accuracy ($t(14) = 4.9024, p < 0.0005$) between the serial and parallel presentation formats. Ease of serial comparison also reflects via lower values of CVF and ST in the serial mode. Again, t-tests confirmed a significant difference in CVF ($t(14) = -3.0842, p < 0.01$) and a marginally significant difference in ST ($t(14) = -1.9476, p = 0.0718$) between the serial and parallel pre-

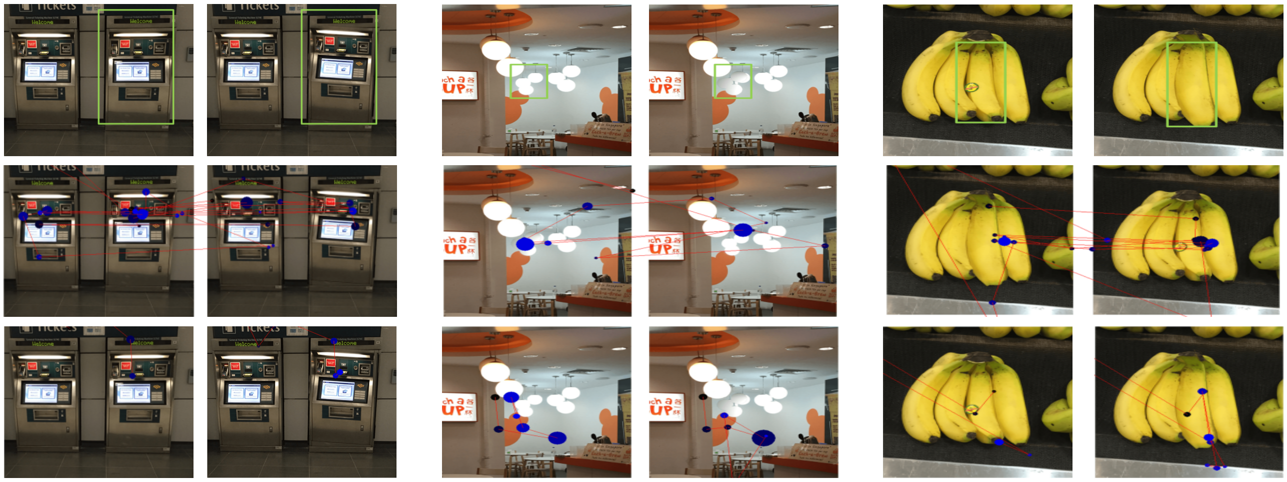

Fig. 2. Sample image pairs varying with respect to *rotation*, *illumination* and *distortion*, and corresponding eye movement patterns. Top row: Original and tampered images, with tampered region marked in green. Gaze patterns obtained with parallel display (middle row) and serial display paradigms (bottom row). Fixations are denoted by circles with dark/light blue shades denoting earlier/later fixations. Circle sizes denote fixation duration, while red lines denote saccades.

sentation formats. Overall, viewers are able to discriminate better, and perform comparisons easily on viewing spatially aligned counterparts in a serial fashion.

Table 1. Statistics for parallel and serial presentation formats.

Format	Acc [%]	CVF	ST [sec]	NF	FD [sec]
Parallel	76.9±10.1	10.3±5.1	12.2±4.7	8.7±3.8	1.0±0.2
Serial	94.9±2.6	4.6±1.1	8.1±3.7	5±1.2	0.6±0.1

3.2. Fixation Analysis

From the raw gaze data, we derived *fixations* by clustering points-of-regard within 0.5^o visual angle radius and gazed for more than 150 milliseconds. Exemplar image pairs corresponding to four different tampering schemes, and gaze patterns (acquired from one user) for the serial and parallel presentation formats are presented in Fig. 2. Earlier and later eye fixations recorded over the comparison time-frame are respectively denoted using darker/brighter circles, while red lines denoting *saccades* represent the shortest distance between successive fixations. Note that fixations occur in corresponding image regions for both presentation formats, but fewer and more localized fixations are noted for serial comparison.

To examine differences in fixation characteristics between the parallel and serial presentation formats, we computed the cumulative number of fixations (NF) on each image pair, and the mean fixation duration (FD) during image comparison. Average values of these measures computed for the two format-specific groups comprising eight users are presented in Table 1. Consistent with visual observations from Fig. 2, significantly higher number of fixations are noted in parallel than in serial comparison ($t(14) = 2.6289, p < 0.05$). Also, considerably longer fixations are noted in the parallel mode ($t(14) = -5.4283, p < 0.0001$).

To verify if visual attention is focused on the tampering during serial comparison, we marked rectangles around the original and forged objects (as in Fig. 2) and computed the cumulative percentage of fixations occurring *outside* of these rectangles for the real-forged image pair. 11% more fixations were observed outside of the tampered area in the parallel mode (72.6 vs 61.5), confirming that eye fixations are significantly more concentrated on the tampered region ($t(14) = 2.4082, p < 0.05$) for serial comparison. Finally, we did not note any significant difference in terms of fixation counts or durations between the serial and parallel presentation formats, suggesting that visual attention resources are equally divided between the real and tampered images for human forgery detection.

3.3. Sensitivity to Tampered Attributes

For benchmarking, we compare two automated CMFD methods, namely, SIFT-based [7] and its robust variant R-CMFD, with human performance considering the various tamper attributes. In R-CMFD, we adapt [7] such that it can distinguish between SGOs and copy-move tampered instances by constructing a pyramid scale-space and computing orientations at each level to achieve scaling and rotation invariance. LBP, DCT and SVD patch features are then extracted for precise texture description along with Harris corner point descriptors. Finally, random sampling consensus (RANSAC) is employed to refine matches obtained via extracted features.

Both [7] and R-CMFD compute a tampering score, s_t, based on the visual match count which is the number of validated matches over the total number of matches. The comparative experiment is conducted in parallel and serial presentation formats, in analogy to the user study. In the parallel setting, the real and forged images are independently evaluated by the automated methods (SIFT-Parallel and R-Parallel), and the image with a higher s_t is considered tampered. In the serial setting, the methods (SIFT-Serial and R-Serial) only evaluate visual matches in the tampered area, assumed to be known *a-priori*, to compute s_t.

CMFD accuracies for different tampered attributes in the parallel and serial settings are presented in Fig. 3 (only human performance is considered for the panoramic image pairs). Both the human and machine perform better in the serial setting, which highlights the importance of accurately localizing the tampered region. Note that machine performance in the parallel setting would be very similar to the real CMFD scenario, as both real and tampered images are evaluated independently.

SIFT-Parallel and R-Parallel outperform human detection for *rotation* and *naive CMF* while comparable performance is noted for *scaling* consistent with prior results [9, 11]. Even in the serial mode where human performance is considerably better, human and machine performance for these three attributes are very comparable. However, for the *illumination* and *distortion* factors, human CMFD performance far exceeds that of the machine in both presentation formats. This highlights the limitations of current CMFD methods, which are unable to detect certain forgery mechanisms even upon localization of the tampered region, calling for improved approaches.

Among the two automated approaches, R-CMFD outperforms the SIFT-based method for the *rotation*, *scaling* and *naive CMF* as it explicitly accounts for SGOs via precise texture description. Even in the serial mode, there is a 10% gain with R-Serial over SIFT-Serial for *scaling* and *naive CMF*, implying the need to account for SGOs around the tampered region. Given the complementary sensitivity of humans and machines to different tampering factors, it is worthwhile to consider a collaborative CMFD framework, where one entity determines potential forged regions, which are further processed by the other for precise detection.

4. DISCUSSION AND CONCLUSION

The presented study involving 16 users confirms the efficacy of human visual perception at discovering CMF in the presence of similar but genuine objects. A comparison-based tamper detection task was specifically designed in order to understand the visual search and tamper detection mechanisms that humans employ to discriminate the forged object from SGOs. Eye movement patterns reveal that systematic comparison of corresponding image regions facilitates tamper detection. CMFD is considerably easier in the serial viewing mode, where spatial alignment makes tampering conspicuous.

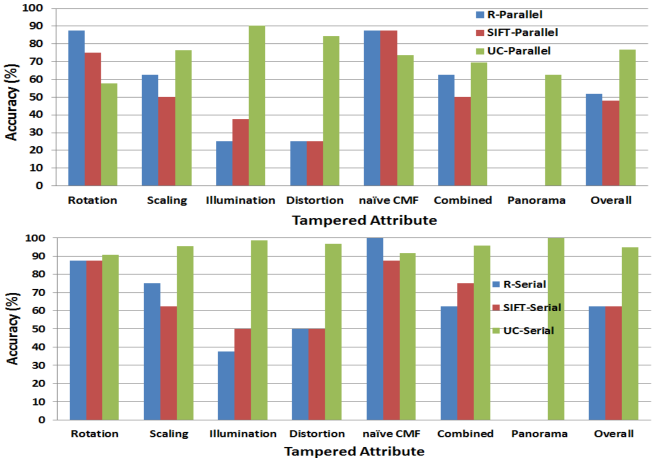

Fig. 3. Automated CMFD for different tamperings.

Machine performance also improves in the serial viewing mode, but is considerably lower than human accuracy for *distortion* and *illumination*, emphasizing the need to design better CMFD algorithms. Nevertheless, machines achieve comparable or better CMFD accuracies for *scale*, *rotation* and *naive CMF*-based tampering in both serial and parallel presentation formats. The need to account for SGOs in CMFD is demonstrated by the fact that R-CMFD, which specifically extracts textural descriptors to this end, performs as well/outperforms SIFT-based detection for the above factors.

Complementary sensitivity of humans and machines to tampering factors reveals promise of a collaborative CMFD framework – designing such a model will be the focus of our future work. We have also published a more comprehensive dataset with additional images and tampering attributes [14].

5. ACKNOWLEDGMENTS

This study was supported by the research grant for the Human-Centered Cyber-physical Systems Programme at the Advanced Digital Sciences Center (ADSC) from Singapore's Agency for Science, Technology and Research (A*STAR).

6. REFERENCES

[1] X. Bo, W. Junwen, L. Guangjie, D. Yuewei: "Image copy-move forgery detection based on SURF." in *Proc. Conference on Multimedia Information Networking and Security (MINES)*, pp. 889–892, 2010.

[2] P. S. Chandakkar, B. Li: "Investigating human factors in image forgery detection." in *Proc. ACM International Workshop on Human-centered Event Understanding from Multimedia (HuEvent)*, pp. 41–44, 2014.

[3] S. Fan, T.-T. Ng, J. S. Herberg, B. L. Koenig, C. Y.-C. Tan, R. Wang: "An automated estimator of image visual realism based on human cognition." in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[4] H. Farid: "Seeing is not believing." *IEEE Spectrum*, vol. 46, no. 8, pp. 44–51, 2009.

[5] S. O. Gilani, R. Subramanian, H. Hua, S. Winkler, S.-C. Yen: "Impact of image appeal on visual attention during photo triaging." in *Proc. IEEE International Conference on Image Processing (ICIP)*, pp. 231–235, 2013.

[6] S. O. Gilani, R. Subramanian, Y. Yan, D. Melcher, N. Sebe, S. Winkler: "PET: An eye-tracking dataset for animal-centric Pascal object classes." in *Proc. International Conference on Multimedia & Expo (ICME)*, 2015.

[7] H. Huang, W. Guo, Y. Zhang: "Detection of copy-move forgery in digital images using SIFT algorithm." in *Proc. Pacific-Asia Workshop on Computational Intelligence and Industrial Application (PACIIA)*, pp. 272–276, 2008.

[8] D. E. Jacobs, D. B. Goldman, E. Shechtman: "Cosaliency: Where people look when comparing images." in *Proc. ACM Symposium on User Interface Software and Technology (UIST)*, pp. 219–228, 2010.

[9] B. Li, T.-T. Ng, X. Li, S. Tan, J. Huang: "Revealing the trace of high-quality JPEG compression through quantization noise analysis." *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 3, pp. 558–573, 2015.

[10] L. Li, S. Li, H. Zhu, S.-C. Chu, J. F. Roddick, J.-S. Pan: "An efficient scheme for detecting copy-move forged images by local binary patterns." *Journal of Information Hiding and Multimedia Signal Processing*, vol. 4, no. 1, pp. 46–56, 2013.

[11] T.-T. Ng, M.-P. Tsui: "Camera response function signature for digital forensics-part I: Theory and data selection." in *Proc. IEEE International Workshop on Information Forensics and Security (WIFS)*, 2009.

[12] R. Subramanian, V. Yanulevskaya, N. Sebe: "Can computers learn from humans to see better? Inferring scene semantics from viewers' eye movements." in *Proc. ACM Multimedia*, pp. 33–42, 2011.

[13] D. Tralic, I. Zupancic, S. Grgic, M. Grgic: "CoMoFoD: New database for copy-move forgery detection." in *Proc. International Symposium ELMAR*, pp. 49–54, 2013.

[14] B. Wen, Y. Zhu, R. Subramanian, T.-T. Ng, X. Shen, S. Winkler: "COVERAGE – A novel database for copy-move forgery detection." in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2016.