COVERAGE – A NOVEL DATABASE FOR COPY-MOVE FORGERY DETECTION

Bihan Wen[1,2]*, Ye Zhu*[3,4]*, Ramanathan Subramanian*[1,5]*, Tian-Tsong Ng*[4]*, Xuanjing Shen*[3]*, Stefan Winkler*[1]

[1] Advanced Digital Sciences Center, University of Illinois at Urbana-Champaign, Singapore.
[2] Electrical and Computer Engineering and CSL, University of Illinois at Urbana-Champaign, IL, USA.
[3] College of Computer Science and Technology, Jilin University, Changchun, China.
[4] Situational Awareness Analytics, Institute for Infocomm Research, Singapore.
[5] Center for Visual Information Technology, Int'l Institute of Information Technology, Hyderabad, India.

ABSTRACT

We present **COVERAGE** – a novel database containing copy-move forged images and their originals with similar but genuine objects. COVERAGE is designed to highlight and address tamper detection ambiguity of popular methods, caused by self-similarity within natural images. In COVERAGE, forged–original pairs are annotated with (i) the *duplicated* and *forged* region masks, and (ii) the tampering factor/similarity metric. For benchmarking, forgery quality is evaluated using (i) computer vision-based methods, and (ii) human detection performance. We also propose a novel sparsity-based metric for efficiently estimating forgery quality. Experimental results show that (a) popular forgery detection methods perform poorly over COVERAGE, and (b) the proposed sparsity based metric best correlates with human detection performance. We release the COVERAGE database to the research community.

Index Terms— Image Forensics, Copy-move forgery, Benchmark database, Sparsity-based metric, Forgery quality

1. INTRODUCTION

Copy-move image forgery is a highly popular tampering method, where a *forged image* is produced by copying an object from the *duplicated region* of the original, manipulating it via specific operations, and pasting it onto the *forged region* within the same image. Copy-move forgery detection (CMFD) is considered simpler than general image forgery detection, since the source of the forged object resides in the same image. In this paper, we will show that CMFD is much more challenging than generally thought, especially when the original image contains multiple similar objects. Most CMFD algorithms [1–4] are based on key-point/block-based region matching, and simply treat region similarity as a measure of copy-move forgery, ignoring ambiguity that may arise between a copy-move forged image vis-á-vis natural image with multiple similar-but-genuine objects (SGOs). To this end, we propose the COpy-moVe forgERy dAtabase with similar but Genuine objEcts (COVERAGE).

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

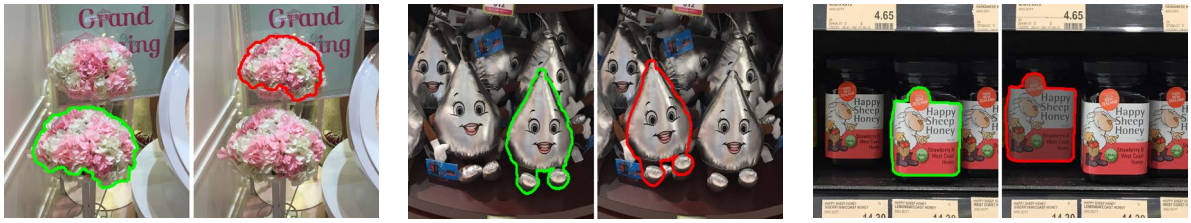
Attribute		CoMoFod	Manip	GRIP	COVERAGE
# image pairs		260	48	100	100
Average Image Size		512 ×512	2305 ×3020	1024 ×768	400 ×486
SGO		✗	✗	✗	✓
Mask	forged	*comb*	✓	*comb*	✓
	duplicated		✗		✓
Tamper Annot.		✗	✗	✗	✓
Tamper Types		*S*,*C*	*S*	*S*	*S*,*C*

Table 1. Overview of CMFD databases. *S* and *C* respectively denote *simple* and *complex*. *comb* denotes combined.

Among datasets available for forgery detection, the Columbia dataset [5] contains authentic and forged gray-scale images, but focuses on image splicing detection instead of CMFD. The MICC databases [6] include forged images without originals, which inconveniences analysis and evaluation. The CoMoFod [7], Manipulation (Manip) [8], and GRIP [4] datasets provide both the forged and original images. Among these, however, only CoMoFod considers complex manipulations (see Sec. 2 for types of complex tampering factors) for forged image synthesis. For evaluation, CoMoFod and GRIP combine the duplicated and forged region masks in a single image without demarcation. COVERAGE explicitly specifies the duplicated and forged region masks, and also considers complex tampering factors. Most importantly, no existing CMFD database accounts for SGOs in the original, which is a commonplace phenomenon in nature. Table 1 overviews CMFD datasets and motivates the need for COVERAGE.

COVERAGE [1] contains 100 original–forged image pairs where each original contains SGOs, making discrimination of forged from genuine objects highly challenging. Both simple and complex tampering factors are employed for forging (see Sec. 2). Also, annotations relating to (i) *duplicated* and *forged* region masks (Fig 1) and (ii) the tampering factor or level of similarity between the original and tampered image are available for all pairs. For benchmarking, we evaluate COVERAGE with several popular computer vision (CV)-based CMFD algorithms, as well as human performance based on visual perception (VP). We also propose a sparsity-based metric which correlates well with human performance.

[1] COVERAGE is available at https://github.com/wenbihan/coverage.



(a) Scaling to $\phi = 0.8$ (b) Rotation by $\theta = 10°$ (c) Illumination change

Fig. 1. Exemplar original–forged image pairs with tampering factor specified for simple transformations (a) and (b). The *duplicated* and *forged* regions are respectively highlighted in green and red.

2. DATABASE GENERATION

Original images in COVERAGE were acquired using an Iphone 6 front camera. We captured both indoor and outdoor scenes for the purpose of this study. These scenes are highly diverse, including stores, offices, public spaces, places of leisure, *etc*, which are highly representative of everyday scenes [9]. A *region of interest* containing at least two SGOs was cropped from each image, and stored in lossless TIFF format as the *original image* corresponding to each pair. A *forged image* was synthesized via graphical manipulation of the original using Photoshop CS4 and also stored as lossless TIFF: One of the original SGOs served as the *duplicated* region, and was transformed via one of the *simple* or *complex* tamperings specified below to *replace* another similar object, and synthesize the forged image.

Masks corresponding to the *duplicated* and *forged* regions are annotated for all image pairs as in Fig. 1. Six types of tampering were employed for forged image generation:

i. **Translation**– the duplicated object (DO) is directly translated and pasted onto the forged region without any manipulation.

ii. **Scaling**– DO is scaled by a factor of ϕ followed by translation (FBT).

iii. **Rotation**– DO is rotated clockwise by θ FBT.

iv. **Free-form**– DO is distorted via a free transform FBT.

v. **Illumination**– DO is modified via lighting effects FBT.

vi. **Combination**– DO is manipulated via more than one of the above factors FBT.

Factors (i) to (iii) denote ***Simple tampering***, since they are easily reproducible with a given parameter. Factors (iv) to (vi) denote ***Complex tampering***, and normally contain multiple transformations. For simple tampering, we annotate the original–forged image pair with the tampering level specified by a single parameter. For complex tampering, we employ similarity metrics to discriminate forged regions from originals as discussed in Section 3.

3. FORGERY QUALITY ESTIMATION METRICS

For the purpose of (a) describing complex transformations, (b) benchmarking CMFD performance on COVERAGE, and (c) demonstrating the challenge posed by SGOs, we attempt to estimate forgery quality in this section. As forgery quality estimation is an open question [10, 11], we explore various metrics to this end. These metrics are intended to serve as a guide for tamper detection difficulty, and not for CMFD performance evaluation *per se*.

3.1. CMFD Benchmark

To benchmark the forgery quality, we used both CV and VP-based detection performance. Popular CV-based CMFD methods include SIFT [1] and SURF [3] based on key-point features, and the recent dense-field method [4] which employs features of densely overlapping blocks. All three methods are capable of matching copy-move key point/block pairs automatically. We follow the original/forged image identification methodology in [1] to compute detection accuracy.

Though CV methods provide for fast evaluation, the copy-move image forgery is essentially aimed to cheat the human eye. In order to study VP-based tamper detection performance, we designed a user study where 30 viewers were required to determine the forged image from each pair upon visual inspection. Mean human detection accuracy in also specified for each COVERAGE pair.

3.2. Spatial Similarity Metrics

Tampering methods are designed to enhance the spatial similarity between the forged and original images, especially for those with SGOs. A poorly forged image usually deviates considerably from its original version, and can thus be easily detected. Amongst existing metrics, we consider Peak Signal-to-Noise Ratio (PSNR) [12] (in decibels) and structural similarity (SSIM) [13] computed on 3×3 overlapping windows of the forged and original images, as global similarity metrics. PSNR and SSIM are commonly used for measuring image reconstruction quality.

Since the PSNR and SSIM measures capture global image similarity, they tend to decrease when the image and *forged* region differs in size. A better similarity metric in such cases is the forged region PNSR (fPSNR), which focuses solely on the tampered region. Denoting the forged and original images as $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{P \times K}$, where P is the number of image pixels and K is the number of color channels ($K = 1$ for grayscale and 3 for RGB), fPSNR is defined as

$$\text{fPSNR} = 20 \log_{10} \left(\frac{255\sqrt{K|C_f|}}{\sqrt{\sum_{k=1}^{K} \sum_{i \in C_f} (X_k^i - Y_k^i)^2}} \right) \qquad (1)$$

Fig. 2. Original (top) and forged images (bottom), with the forged region boundary in red (left), and a zoom-in of the patches centered at the boundary (right).

where C_f denotes the set of pixels corresponding to the forged region. $\{X_k^i, Y_k^i\},\ i \in C_f$ denotes the i^{th} pixel pair in C_f derived from the k^{th} channel of $\mathbf{X}$ and $\mathbf{Y}$. Different from PSNR and SSIM, fPSNR is a tamper quality estimation metric invariant to forged region size.

3.3. Forgery Edge Abnormality Using Adaptive Sparsity

Human forgery detection typically employs the forged edge abnormality (FEA) as an important clue [14, 15]. To circumvent the need for involving users but emulate human detection performance, we propose a sparsity-based metric measuring FEA. Figure 2 visualizes FEA for the original and forged images. Natural images are usually sparsifiable [16, 17], while the imperfect object embedding in image tampering produces unnatural local artifacts at the boundary [18] as highlighted in Fig. 2. CMFD algorithms are normally insensitive to such artifacts, which can be easily captured by the human eye and facilitate VP-based forgery detection. Inspired by the union-of-transforms (UOT) model [19], we compute the FEA metric via adaptive UOT-domain sparsity.

To this end, we first calculate the adaptive sparse modeling error for forged images. Define $\mathbf{Y}$ and $\mathbf{X}$ to be the original and forged images, and B to be the set of pixels at the boundary of the forged region C_f and the background $\neg C_f$. We extract overlapping patches that are centered at the interior of C_f, denoted as $\{R_j\mathbf{Y}\}_{j\in C_f,\, j\notin B}$, where R_j extracts the j^{th} patch within the set of pixels. The sparsifying transforms W_f and W_o model C_f and $\neg C_f$ respectively. The optimal $\hat{W}_f$ and $\hat{W}_o$ are obtained by solving (P1),

$$\text{(P1)}\quad \hat{W}_f = \underset{W_f,\{\alpha_j\}}{\operatorname{argmin}} \sum_{j\in C_f,\, j\notin B} \|W_f R_j\mathbf{X} - \alpha_j\|_2^2$$

$$\hat{W}_o = \underset{W_o,\{\beta_i\}}{\operatorname{argmin}} \sum_{i\in \neg C_f,\, i\notin B} \|W_o R_i\mathbf{X} - \beta_i\|_2^2$$

$$s.t.\ \ \|\alpha_j\|_0, \|\beta_i\|_0 \le s,\ \ \forall\ i,\ j,\ W_f^H W_f = I,\ W_o^H W_o = I.$$

Here $\{\alpha_j\}$ and $\{\beta_i\}$ are the sparse codes of the extracted patches with sparsity level smaller than s, while $(.)^H$ denotes the Hermitian operator. The solutions $\hat{W}_f$ and $\hat{W}_o$ are both restricted to be unitary. Problem (P1) can be solved by alternating between *sparse coding* and *transform update* steps, both of which have closed-form solutions [19]. Denote the learned UOT as $U = \left\{\hat{W}_f,\ \hat{W}_o\right\}$. We evaluate the normalized UOT modeling error for the forged image, denoted as E_f, using the patches centered at the boundary B. Each patch selects either $\hat{W}_f$ or $\hat{W}_o$ to ensure smaller modeling error. The normalized forged image UOT modeling error E_f is calculated as

$$E_f = \frac{\sum_{j\in B} \min_{W\in U} \left\|WR_j\mathbf{X} - \mathbf{Proj}_{0,s}(WR_j\mathbf{X})\right\|_2^2}{\sum_{j\in B} \|R_j\mathbf{X}\|_2^2} \qquad (2)$$

where $\mathbf{Proj}_{0,s}(\cdot)$ denotes l_0 ball projection [19] with maximum sparsity level s.

To calculate the normalized UOT modeling error for original images, we follow the same algorithm but using boundary patches extracted from $\mathbf{Y}$ to obtain E_o in a similar way. The proposed FEA metric is defined as the difference of the normalized UOT modeling errors,

$$\text{FEA} = E_f - E_o \qquad (3)$$

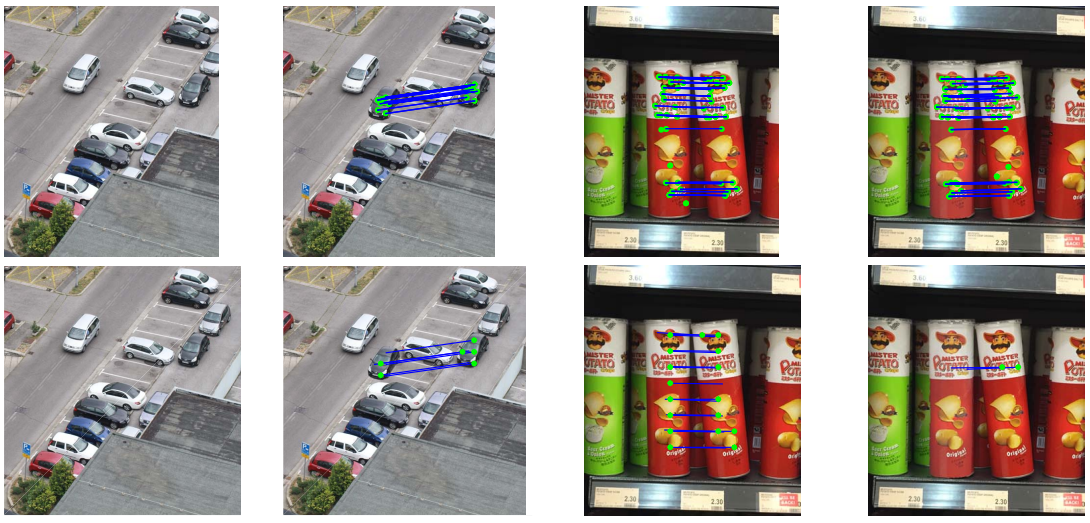
Forged images with boundary patches satisfying the adaptive UOT model usually correspond to a small FEA value, and are thus considered well tampered (*i.e.*, contain less unnatural artifacts). A well-forged image is also less likely to be detected by a human, and therefore corresponds to a lower VP-based detection accuracy. We observe this linear correlation between FEA and human performance in Section 4.

4. EXPERIMENTAL RESULTS AND ANALYSIS

Tamp. factor (#images)		Spatial Similarity			FEA (%)	VP (%)	CV (%)
		SSIM	PSNR	fPSNR			
S	*Trans.* (16)	0.95	24.0	13.4	29.7	78.1	72.9
	Scal. (16)	0.97	28.5	14.8	28.6	71.3	62.5
	Rot. (16)	0.93	24.9	14.7	24.9	66.0	62.2
C	*Free.* (16)	0.93	24.8	14.1	20.0	63.8	59.4
	Illum. (16)	0.92	23.0	13.3	25.2	66.9	56.3
	Comb. (20)	0.94	24.0	13.8	28.9	72.5	53.3
Overall (100)		0.94	24.8	14.0	26.2	69.9	60.3

Table 2. Mean spatial similarity scores, FEA metric, VP and CV-based detection accuracy for simple (S) and complex (C) tampering factors in COVERAGE.

We firstly evaluate forgery quality in COVERAGE using the metrics proposed in Section 3. All image pairs in COVERAGE are annotated and categorized based on their tampering factors. Table 2 lists the mean spatial similarity scores, sparsity-based FEA values, VP, and CV detection accuracies for each category, as well as for the whole database. The CV-based CMFD accuracy is obtained by averaging the results from the SIFT [1], SURF [3] and dense-field [4] CMFD methods. Given that the aim of tampering is to escape human detection, the forgeries that are more difficult to detect



(a) Original, CoMoFod (b) Forged, CoMoFod (c) Original, COVERAGE (d) Forged, COVERAGE

Fig. 3. CMFD performance comparison: Results of SIFT [1] (top) and dense-field [4] (bottom) methods. Left to right: Original (a, c) and forged (b, d) pairs from CoMoFod [7] and COVERAGE respectively. Dots and red lines represent key points (or block centers) and matched pairs.

correspond to lower VP-based accuracies. We observe that the FEA metric (linearly) correlates best with VP-based accuracies ($\rho = 0.89, p < 0.01$) among the measures considered. Among similarity metrics, fPSNR optimally characterizes forgery quality, and correlates negatively ($\rho = -0.33$) with human performance as one would expect. Also, while human performance considerably exceeds CV-based CMFD for some tampering factors, CV methods perform as well or better than humans for others (rotation and free-form deformation). These results suggest that a collaborative framework could potentially work best for CMFD with SGOs.

Finally, to demonstrate the challenge that COVERAGE poses to popular CMFD methods, we performed CV-based CMFD on the CoMoFoD [7], Manipulate [8], and GRIP [4] datasets. From each original-forged pair, we randomly selected one image for testing for all datasets. Table 3 lists the corresponding CMFD accuracies. SURF-based CMFD is generally poor across all datasets. The SIFT and dense-field methods perform better over, with dense-field producing the best performance. Overall, considerably lower mean CMFD accuracy is obtained for COVERAGE.

Figure 3 qualitatively presents CMFD using the SIFT and dense-field methods on exemplar CoMoFod and COVERAGE pairs. CoMoFod and CMFD methods in general, do not account for SGOs, and the number of point/block matches is considered as a measure of copy-move forgery (Fig. 3(a,b)). In contrast, COVERAGE explicitly considers SGOs, and more matches are noted between SGOs in the original (Fig. 3(c)), than between the *duplicated* and *forged* (red can on the left with changed illumination) regions. Moreover, relatively few matches are obtained for complex tampering such as illumination-based (Fig. 3(d)), leading to low CMFD accuracies as in Table 3. COVERAGE is mainly intended to help develop robust CMFD methods addressing such limitations.

CMFD Database (# test images)		SIFT	SURF	Dense-Field	Aver.
COVERAGE	*Trans.* (16)	50.0	75.0	**93.8**	72.9
	Scal. (16)	56.3	56.3	**75.0**	62.5
	Rot. (16)	46.7	53.3	**86.7**	62.2
	Free. (16)	43.8	50.0	**68.8**	59.4
	Illum. (16)	43.8	**62.5**	**62.5**	56.3
	Comb. (20)	**55.0**	**55.0**	50.0	53.3
	All (100)	50.5	58.6	**71.8**	60.3
CoMoFoD (200)		**77.0**	51.5	72.0	66.8
Manipulate (48)		75.0	58.3	**95.8**	76.4
GRIP (100)		71.0	52.0	**82.0**	68.3

Table 3. CMFD detection accuracies using SIFT, SURF, and Dense-field methods on various datasets. For each dataset, the best CMFD result is marked in bold.

5. CONCLUSION

This paper presents COVERAGE, a novel CMFD database with annotations where the challenge is to distinguish the forged region from SGOs. Several metrics are proposed to estimate forgery quality, as well as CV and VP-based CMFD performance. Obtained results reveal that (i) the FEA metric correlates well with human performance, and (ii) automated CMFD methods perform poorly on COVERAGE.

6. REFERENCES

[1] A Jessica Fridrich, B David Soukal, and A Jan Lukáš, "Detection of copy-move forgery in digital images," in *Digital Forensic Research Workshop*, 2003.

[2] G. Muhammad, M. Hussain, and G. Bebis, "Passive copy move image forgery detection using undecimated dyadic wavelet transform," *Digital Investigation*, vol. 9, no. 1, pp. 49–57, 2012.

[3] Xu Bo, Wang Junwen, Liu Guangjie, and Dai Yuewei, "Image copy-move forgery detection based on SURF," in *MINES*, 2010, pp. 889–892.

[4] D. Cozzolino, G. Poggi, and L. Verdoliva, "Efficient dense-field copy–move forgery detection," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 11, pp. 2284–2297, 2015.

[5] T. Ng, S. Chang, and Q. Sun, "A data set of authentic and spliced image blocks," *Columbia University, Technical Report*, pp. 203–2004–3, 2004.

[6] I. Amerini, L. Ballan, R. Caldelli, A. D. Bimbo, and G. Serra, "A SIFT-based forensic method for copy–move attack detection and transformation recovery," *IEEE Trans. Information Forensics and Security*, vol. 6, no. 3, pp. 1099–1110, 2011.

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

[8] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou, "An evaluation of popular copy-move forgery detection approaches," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1841–1854, 2012.

[9] A. Quattoni and A. Torralba, "Recognizing indoor scenes," in *Computer Vision and Pattern Recognition*, 2009, pp. 413–420.

[10] L. Ballard, D. Lopresti, and F. Monrose, "Forgery quality and its implications for behavioral biometric security," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, no. 5, pp. 1107–1118, 2007.

[11] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou, "An evaluation of popular copy-move forgery detection approaches," *IEEE Trans. Information Forensics and Security*, vol. 7, no. 6, pp. 1841–1854, 2012.

[12] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electronics letters*, vol. 44, no. 13, pp. 800–801, 2008.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[14] J. Wang, G. Liu, B. Xu, H. Li, Y. Dai, and Z. Wang, "Image forgery forensics based on manual blurred edge detection," in *Int'l Conference on Multimedia Information Networking and Security*, 2010, pp. 907–911.

[15] G. Peterson, "Forensic analysis of digital image tampering," in *Advances in Digital Forensics*, pp. 259–270. 2005.

[16] D. Liu, Z. Wang, B. Wen, W. Han J. Yang, and T. Huang, "Robust single image super-resolution via deep networks with sparse prior," *IEEE Transactions on Image Processing*, 2016.

[17] Y. Li, K. Lee, and Y. Bresler, "Identifiability in blind deconvolution with subspace or sparsity constraints," *arXiv preprint arXiv:1505.03399*, 2015.

[18] Z. Wang, J. Yang, H. Zhang, Z. Wang, Y. Yang, D. Liu, and T. Huang, *Sparse Coding and its Applications in Computer Vision*, World Scientific, 2015.

[19] B. Wen, S. Ravishankar, and Y. Bresler, "Structured overcomplete sparsifying transform learning with convergence guarantees and applications," *Int'l Journal of Computer Vision*, vol. 114, no. 2, pp. 137–167, 2015.