A dataset of 1050-tampered color and grayscale images (CG-1050)

Maikol Castro, Dora M. Ballesteros*, Diego Renza

Universidad Militar Nueva Granada, Colombia

A R T I C L E   I N F O

Article history:
Received 9 July 2019
Received in revised form 12 November 2019
Accepted 13 November 2019
Available online 21 November 2019

Keywords:
Fake image
Tampering detection
Forgery detection
Copy-move
Cut-paste
Retouching
Colorizing

A B S T R A C T

This paper presents the CG-1050 dataset consisting of 100 original images, 1050 tampered images and their corresponding masks. The dataset is organized into four directories: original images, tampered images, mask images, and a description file. The directory of original images includes 15 color and 85 grayscale images. The directory of tampered images has 1050 images obtained through one of the following types of tampering: copy-move, cut-paste, retouching, and colorizing. The true mask between every pair of original and its tampered image is included in the mask directory (1380 masks). The description file shows the names of the images (i.e., original, tampered and mask), the image description, the photo location, the type of tampering, and the manipulated object in the image. With this dataset, the researchers can train and validate fake image classification methods, either for labeling the tampered image or for forgery pixel-detection.

© 2019 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Data

The dataset is organized in four directories: Original images, Tampered images, Mask images, and a Description file [1].

* Corresponding author.
E-mail address: dora.ballesteros@unimilitar.edu.co (D.M. Ballesteros).

https://doi.org/10.1016/j.dib.2019.104864
2352-3409/© 2019 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Fig. 1 shows the structure of the dataset, which is explained below:

- The directory of Original images includes 15 color and 85 grayscale images. All original images are in JPEG format.
- The directory of Tampered images has 100 sub-directories (i.e., T_1 to T_100) with 1050 images obtained through one of the following types of tampering: copy-move, cut-paste, retouching, and colorizing. The first 50 sub-directories have 11 tampered images, each one; the last 50 sub-directories have 10 tampered images by directory. All tampered images are in JPEG format.
- The directory named Mask has 100 sub-directories (i.e., Mask_1 to Mask_100) with their true masks obtained by each pair of original and tampered image. In the case of color images, every of the 15 sub-directories has 11 folders and 3 masks by folder, that is, 495 masks. In the case of grayscale images, every one of the first 35 sub-directories have 11 masks; and every one of the last 50 sub-directories have 10 masks, i.e., 885 masks. To sum up, the entire CG-1050 dataset has 1380 masks. In the mask image, the manipulated pixels are black and the unmodified pixels are white.
Finally, the directory named Description has an excel file with information about the dataset details: original images (i.e., photo name, image description, photo place), tampered images (i.e., folder name, type of tampering, tampered photo name, object, location), and mask (i.e., folder name, mask photo name).

Fig. 2 shows an example of cut-paste manipulation for a color image, with the original, the tampered image, and the mask. Fig. 3 shows an example of a copy-move operation and associated images. Figs. 4 and 5 show an example of grayscale images for colorizing and retouching. Tables 1–4 describe the information in Figs. 2–4.

2. Experimental design, materials, and methods

The natural images were captured in the following places: street (51 photos), park (20 photos), touristic place (11), mall (8 photos), shop (4 photos), classroom (2 photos), parking lot (1 photo), room (1 photo), kitchen (1 photo), and playroom (1 photo). Size of the images are (3456 × 4608) or
(4608 × 3456) pixels. For every original image, 10 to 11 tampered images (i.e., with copy-move [3,4], cut-paste [2], retouching [5,6] and colorizing [7,8]) are obtained.

Fig. 2 shows an example of cut-paste modification of a color image. The left plot is the original image, the middle plot is the tampered image, and the right plot is its corresponding mask (G band). The light pole located in the right side of Fig. 2b is the object copied from another image. Table 1 shows the details of these images found in the Description directory.

Fig. 3 shows another example of tampered color images. In this case, a copy-move modification is applied to the original image, pasting twice a bush. The true mask is presented in Fig. 3c. Table 2 shows the details of the original, tampered and mask images for this example.

For the third example, a grayscale image of the CG-1050 dataset is selected. The intensity of the girl’s dress is changed as well as the intensity of the sidewalk. Fig. 4 shows the original image, the tampered image and its mask. Table 3 lists the details of these images.
The last example is shown in Fig. 5. The lines of the road are blurred, through a retouching effect. The tampered object is located in the middle right of the image (see Fig. 5c). Table 4 shows the details of this manipulation.
Acknowledgments

This work is supported by the “Universidad Militar Nueva Granada – Vicerrectoría de Investigaciones” under the grant IMP-ING-2936.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Castro, D.M. Ballesteros, D. Renza, CG-1050: Original and tampered images (Color and grayscale), Mendeley Data, v2. https://doi.org/10.17632/dk84bmnyw9.2, 2019.

[2] L. Zheng, Y. Zhang, V.L. Thing, A survey on image tampering and its detection in real-world photos, J. Vis. Commun. Image Represent. (2019) 380–399, https://doi.org/10.1016/j.jvcir.2018.12.022.

[3] H.Y. Huang, A.J. Ciou, Copy-move forgery detection for image forensics using the superpixel segmentation and the Helmert transformation, EURASIP J. Image Video Process. 1 (2019) (2019) 68, https://doi.org/10.1186/s13640-019-0469-9.

[4] Y. Li, J. Zhou, Fast and effective image copy-move forgery detection via hierarchical feature point matching, IEEE Trans. Inf. Forensics Secur. 14 (2018) 1307–1322, https://doi.org/10.1109/TIFS.2018.2876837.

[5] X. Liu, L. Xie, B. Zhong, J. Du, Q. Peng, Automatic facial flaw detection and retouching via discriminative structure tensor, IET Image Process. 11 (2017) 1068–1076, https://doi.org/10.1049/iet-ipr.2016.1095.

[6] H. Ortiz, D. Renza, D. Ballesteros, Tampering detection on digital evidence for forensics purposes, Ingeniería y Ciencia 14 (27) (2018) 53–74, https://doi.org/10.17230/ingciencia.14.27.3.

[7] Y. Guo, X. Cao, W. Zhang, R. Wang, Fake colorized image detection, IEEE Trans. Inf. Forensics Secur. 13 (8) (2018) 1932–1944, https://doi.org/10.1109/TIFS.2018.2806926.

[8] L. Zhuo, S. Tan, J. Zeng, B. Lit, Fake colorized image detection with channel-wise convolution based deep-learning framework, in: IEEE Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2018, pp. 733–736, https://doi.org/10.23919/APSIPA.2018.8659761.