FMIDV: Forged Mobile Identity Document Video dataset

This webpage presents a dataset for copy-move forgeries on the identity documents of MIDV-2020 dataset. The forged samples contain many Similar butGenuine Objects (SGO) which has been shown as a challenge for Copy-Move Forgery Detection (CMFD) algorithms and should be useful in many works in digital forensics research.

Any use of this dataset is required to cite the following reference:

M. Al-Ghadi, Z. Ming, P. Gomez-Krämer, and J.-C. Burie. Identity documentsauthentication based on forgery detection of guilloche pattern, arXiv, 2022.

https://doi.org/10.48550/arxiv.2206.10989

FMIDV size and access

The dataset has a size of 4,7 GB and is hosted on an FTP server of the University of La Rochelle. Please fill this **form** for getting access to the dataset. In case of any problem, please contact (musab.alghadi@univ-lr.fr or muhammad_muzzamil.luqman@univ-lr.fr).

Structure of genuine dataset (MIDV-2020)

MIDV-2020 dataset collects identity documents in four categories:

1- Template category : comprises 1000 dummy identity documents
The original sample images obtained from Wikimedia Commons andedited to remove non-persistent data (such as signature, photo, and text field values).

2- Photo category : comprises 1000 photos in resolution of 2268×4032 pixels

A photo was taken for each physical document sample, given variousconditions and using two smartphones. Half of the photos were captured using Apple iPhone XR, and the other half using Samsung S10.

3- Scan category : comprises 2000 scanned identity documents
Each physical document sample was scanned using Canon LiDE 220and Canon LiDE 300 scanners, in two different modes (1000 for each).
Scanning modes:
- upright position (scan_upright.tar)
- rotated to an arbitrary angle (scan_rotated.tar)

4- Video clips category : comprises 1000 video clips

For each physical document sample a video clip was captured vertically using Apple iPhone XR and Samsung S10, in a resolutionof 2160×3840 pixels, with 60 frames per second.

Structure of forged dataset (FMIDV)

FMIDV dataset consists of 28k forged identity documents for 10 countries basedon copy-move forgeries on the identity documents of MIDV-2020 dataset.

For each identity document in the template, photo and scan categories of MIDV-2020, we have generated 7 forged samples based on copy-move operation.

Copy-move operations were applied on zones of sizes 16×16 and 32×32 and 64×64 pixels; selected randomly.

For 16×16 and 32×32 pixels copy-move forgeries were applied 2 times for 2 different zones, 2 times for 4 zones, and 2 times for 6 zones.

For 64×64 pixels copy-move forgery was applied 1 time for only 2 different zones; because we don't have enough available zones for applying this action forsome countries e.g. Finland, Serbia and Slovakia. Moreover, applying copy-moveoperation on 64×64 zones are out of interest as they could be detected by manualinspection (naked eyes).

FMIDV is structed as follows:

1- forged_ templates: comprises 7K IDs for 10 countries.
2- forged_photo: comprises 7K IDs for 10 countries.
3- forged_ scan_rotated: comprises 7K IDs for 10 countries.
4- forged_ scan_upright: comprises 7K IDs for 10 countries.

The format of any forged identity document in FMIDV is as follows:

no._category_Px_Zy.png

no.: presents the sample number; $no. = [00 - 99]$

category: presents the category name that one sample belongs

Px: partition size; $x = \{16, 32, 64\}$, if $x = 16$ that means that copy-moveoperation done on zones of 16×16 etc.

Zy: presents number of selected zones; $y = \{2, 4, 6\}$

Demos of the samples

Contacts:

Dr. Musab Al-Ghadi musab.alghadi@univ-lr.fr
Dr. Zuheng Ming : zuheng.ming@univ-lr.fr
Dr. Muhammad Muzzamil Luqman : muhammad_muzzamil.luqman@univ-lr.fr
Dr. Petra Gomez : petra.gomez@univ-lr.fr
Pr. Jean-Christophe Burie : jcburie@univ-lr.fr