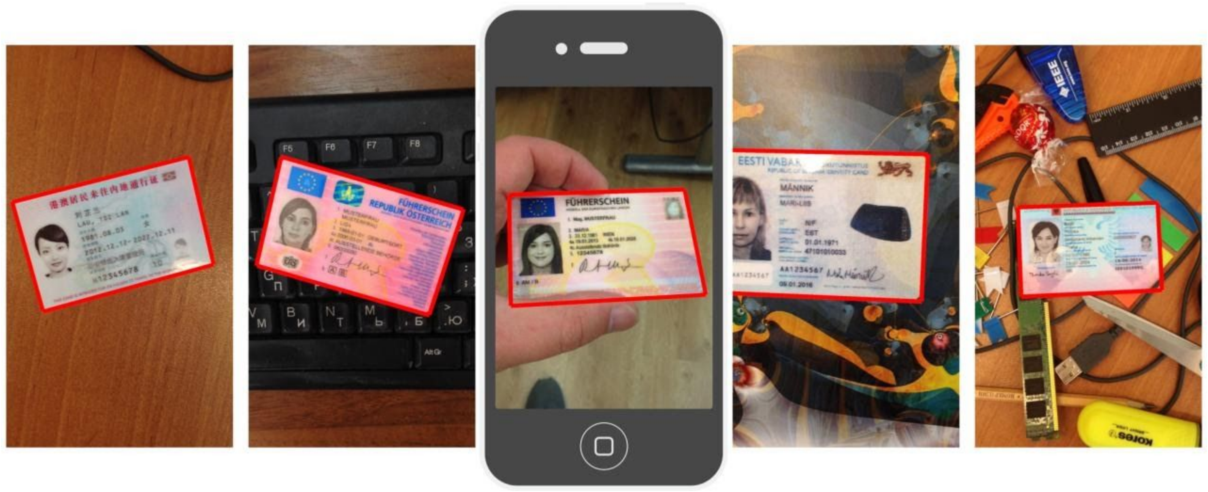
smart engines™

Whitepaper

# Datasets of ID documents:
# MIDV-500

## How to test ID recognition algorithms?

As you might already know, we at Smart Engines are developing computer vision and document recognition systems, and are engaged in scientific research in this field. Our main focus for many years has been placed on solving problems related to recognition of identity documents, the algorithms and approaches to which we have been gladly sharing with the community of interest in our scientific articles. One of the critical problems with the preparation of those though was a lack of public datasets which we could use to demonstrate our achievements. Thus, for research purposes, two years ago we started working on an open collection of video clips, and now would like to share the results of our work with you.



### What is this about?

The fact that smartphones and other portable devices have become a crucial element in any online service - especially when it comes to fintech, sharing economy, and the public sector, - has long been accepted. This is particularly relevant for those services that involve remote identification, where users have to provide their personal details from either a passport, driver's license or a bank card. Manual data entry on a mobile device, as we all know, is not exactly the most convenient process, which is the reason most large companies have shifted towards introduction of various document recognition and analysis systems to simplify this procedure. And that is not the only reason why they integrate such technologies: those systems not only accelerate and simplify the data entry processes; they automate the document validity verification, thus detecting users who might have malicious intent. The major challenge here is that recognition systems have to perform well in any challenging, real-life environment – be it in a well-lit office, or on a suburban train platform at dusk – the technology has to keep its proclaimed standard.

There are several research teams around the world that are engaged in solving problems related to mobile recognition and analysis of documents in challenging conditions. Aside from the fact that this task is not easy in itself, the main problem here is that there is no data to test the results on. This is especially true for identity documents, open datasets of which, for good reason, simply don’t exist. As a result, there is nothing left for researchers to do but work with synthetically generated data (which does not really reflect the reality), and test their algorithms on either their own identity documents or publicly available samples, which are insufficient.

There are, of course, several open databases with samples of identification documents (i.e. PRADO or Edison). The images in these databases though, as a rule, do not reflect real use cases; and that is not the only concern – in most cases the images are also protected by copyright and are restricted for reuse with or without modification. Another problem with document datasets publicly available today is that they are insufficient for benchmarking the full recognition of identity documents cycle, especially when it comes to on-mobile and in the video stream recognition. Some of those datasets are provided in the table below.

| **Task** | **Datasets** |
|---|---|
| Document detection and localization | 10.1109/ICDAR.2015.7333943 |
| Text segmentation | arXiv:1601.07140 |
| Document image binarization | 10.1109/ICDAR.2017.228 |
| Optical character recognition | MNIST, Uber-Text |
| Image super-resolution | 10.1109/ICDAR.2017.306 |
| Document forensics | 10.1109/EST.2017.8090394 |
| Document image classification | 10.1109/ICDAR.2015.7333910 |
| Document layout analysis | 10.1109/ICDAR.2009.271<br>10.1109/ICDAR.2017.229 |
| Document image quality assessment | 10.1007/978-3-319-05167-3_9<br>10.1109/ICDAR.2015.7333960 |

*Table 1. Some of the open datasets that can be used in certain cases for analysis and recognition of documents.*

Since we are particularly interested in the algorithms for processing and recognition of identity documents in the video stream received from the mobile device’s camera, we have decided to create a dataset that will contain short videos of several different ID documents shot under various conditions. Why are videos so important? Firstly, analysis of several consecutive frames allows using various filtering and refinement techniques when solving one of the main challenges for document recognition systems, which is document location and type classification. Secondly, to solve the problem of choosing the best representation of an object (the best full image of the document, for example, or the best personal photo), one needs a dataset that contains multiple frames. Thirdly, having multiple images of the same object helps to improve the recognition accuracy by combining the results from multiple frames.

We called the resulting dataset “MIDV-500”, where MIDV stands for “Mobile Identity Documents in Videostream”, and 500 represents the number of videos we have collected for it. The dataset is available for download **via this link**.

## Dataset structure

You might probably wonder where we got all these documents from. Well, to create a dataset, we went through Wikipedia in search of samples of ID documents that were either unprotected by copyright, or open source licensed. We managed to find 50 document types, including 17 types of ID cards, 14 types of passports, 13 types of driver’s licenses and 6 other types (**listed here**). We printed out each document, laminated them to simulate reflections, and shot 10 videos for each of them in 5 different conditions using the iPhone 5 and Samsung Galaxy S3 smartphones. A list of shooting conditions is presented in Table 2.

| **Identifier** | **Description** |
|---|---|
| TS, TA   Table | simplest case, the document lays on the table with homogeneous background |
| KS, KA   Keyboard | the document lays on various keyboards, making it harder to utilize straightforward edge detection techniques |
| HS, HA   Hand | the document is held in hand |
| PS, PA   Partial | on some frames the document is partially or completely hidden off-screen |
| CS, CA   Clutter | scene and background are intentionally stuffed with many unrelated objects |

*Table 2. Scanningting conditions for the MIDV-500. The first letter of the identifier encodes the condition, the second one - the device (A - Apple iPhone 5, S - Samsung Galaxy S3).*

Each of the 500 video clips was shot for at least 3 seconds, and the first 3 seconds were split into frames with 10 FPS frame rate. A total of 15,000 frames were received, all with a resolution of 1080x1920px.

Each of the 15,000 frames was accompanied by a JSON file with the coordinates of the corners of the document in the following format:

```
{

  "quad": [ [0, 0],      [111, 0],

            [111, 222], [0, 222] ]

}
```

In case the corners of the document were outside the frame, the coordinates were extrapolated so that the segments connecting the vertices of the quadrangle corresponded to the visible borders of the document.

For each of the 50 documents in the dataset there was a JSON file containing information about the location of the fields (in the aligned image of the document) and the values of its text fields, in the following format:

```
{

  "field01": {

    "value": "Erika",

    "quad": [ [983, 450],  [1328, 450],

              [1328, 533], [983, 533] ]

  },

  // ...

  "photo": {

    "quad": [ [78, 512],   [833, 512],

              [833, 1448], [78, 1448] ]

  }

}
```

In total, the dataset had 48 "photo" fields, 40 "signature" fields, and 546 text fields in different languages..

A year after the first publication of the MIDV-500 dataset, we began to receive the first feedback from researchers and, based on its results, prepared and published an extension of the dataset called MIDV-2019. 200 video clips of the same 50 documents were added, but shot in two new conditions: with strong projective distortions and in low lighting. The new video clips were shot at a resolution of 2160x3840px on two more modern smartphones: iPhone XS Max and Samsung Galaxy S10. The MIDV-2019 extension is available here.

## The experiments

It’s been almost two years since MIDV-500 dataset was first released with open access, and a number of works have been published on its basis since then - both by our employees and researchers from other scientific and business teams. In Table 3 we reference some of the publications involving the use of MIDV-500 and MIDV-2019 datasets to develop and benchmark ID documents recognition algorithms.

| Purpose of the experiment | Link |
|---|---|
| Face detection and text fields recognition on documents with noisy border detection.The main MIDV-500 publication | 10.18287/2412-6179-2019-43-5-818-824 |
| Recognition of text fields of the document projective distortions and low lighting conditions. Main publication of MIDV-2019. | 10.1117/12.2558438 |
| Document type identification based on keywords | 10.1109/RUSAUTOCON.2019.8867614 |
| ID document detection and quality assessment | http://www.informaticahabana.cu/es/node/5578 |

| | |
|---|---|
| Fast detection and identification of an ID-document | 10.1109/ICDAR.2019.00141 |
| Text field recognition | 10.1109/ACCESS.2020.2974051 |
| Detection of faces in images of ID documents | 10.1109/ICDARW.2019.30065 |
| Sensitive data detection and masking | 10.1109/SIBIRCON48586.2019.8958325 |
| Detection of vanishing points in document images with HoughNet | 10.1109/ICDAR.2019.00140 |
| Monospaced font detection | 10.1117/12.2559373 |
| Stopping the text field recognition in a video stream | 10.1007/s10032-019-00333-0 |

*Table 3. A selection of publications.*

We continue to work on expanding the dataset, paying particular attention to the text / image data variability, the background, and capturing conditions. We hope that the publication of such data will enable research teams to benchmark their image analysis algorithms and successfully publish their results.