Identity Verification Tool

Table of Contents

Author: ing. Rares Tohanean
rares.tohanean@gmail.com

1 Introduction

What is it?

"A software tool to verify the identification documents by checking the security elements, internal consistency and intra documents correlations."

What is it good for?

In the last years we see an increased amount of online services. The need to enhance security and mitigate the risk associated with online services is high and it will only increase as more and more people and companies are oriented to providing online services.

The classical way to identify persons is by requiring physical presence to a counter and present identity documents. The operator confirms visually that the user carries his own documents (photo match between documents and the person) then inspects the documents visually or using a document reader.

In order to receive services from a company, is mandatory for the person to pass the identification process. The services requested as a result may be:

- creating a bank account
- loans
- issue a credit card
- request to issue government document

- requests duplicate of a government document
- issue personal digital certificate
- request information for personal use (criminal record proof, financial reports related to fees, etc.)
- services associated with virtual documents such as mobile driver's license
- any other personalized and secure private or government service

Because of the offline identification process required, that kind of services can't be offered online, yet.

The product helps securely identify the person online and can also be a great tool for document verification when the document is physically present.

Banks or Electronic Money Institutions are the first that adopted the online verification process in order to create accounts and issue cards to people without requiring physical presence at the counter. Because the identification services are used mainly by financial institutions, all the players in this market offer services of identification and checks oriented toward financial preventing financial crime and meeting the regulations known by acronyms KYC/AML/CFT[i]. For other type of services, for example government services, the business landscape is not very crowded. Establishing a product like this, could lead the company to become a market creator in a *blue ocean*[ii] business configuration.

Who can use this service?

- Banks
- Financial institutions
- Government Institutions
- Any organization in need to securely identify its customers

"Reliable services such as video identification will undoubtedly become the backbone of a new generation of safer internet services. User identity and privacy will become especially relevant in strengthening an economy more conscious about the importance of personal information and identity".[iii]

Who can build it?

A small team of highly specialized software programmers having experience in:

- secure documents analysis or production
- dense algorithms
- artificial intelligence
- image processing and patterns recognition

- image segmentation algorithms

How this product is different?

- The last generation smartphones come with a very performant image acquisition system and with a powerful processor. From the hardware perspective those devices are ready to be used for document verification instead of the bulky traditional document readers. This opens up new possibilities where a document reader could not be used for example on the road by the police or at any bank office.
- The software can be offered as a service to analyze documents images sent by the persons who require online services. Using innovations in document image capture and image analysis employed by artificial intelligence and new machine learning algorithms, we can let the person make the photo remotely and still provide a secure identification.

2 System description and process

The system described in this document helps to identify the person and gives answer to the question:

-- Is this person who claims to be?

The answer could be "yes" or "no" or a probability between them.

A company needs the true identity of a user to be able to verify him on black lists.

In order to achieve this identification, we need to

- require identification documents from the user
- analyze the document for consistency
- analyze the document if it's genuine
- compare document photo with a different photo required from the client
- consolidate data with auxiliary documents

2.1 Reasons for forgery

What could be the reasons for persons to declare a different identity?

- If a person wants to make transactions knowing that it could brake the law he will try to hide his identity in order to protect himself.
- If a person is already under some form of surveillance or penalty, he will try to change his identity in order to be able to continue his illicit activity.

- A person steal the identity of another person and tries to get a loan

2.2 How identity can be faked

What are the ways a client can fake his own identity?

If a person wants to protect himself from future investigations he needs to assure that in case of an investigation, the authorities will not be able to reach the real person. This means that he will have to fake all the data from the identification documents: name, photo, birth date and place, document number, personal number, address.

If a person wants to assure is not found on black lists, it is enough to change only some data like the name, the document number and personal number.

Let's take the fake identity use cases in an online identification process from the simplest to the most complex, and find solutions to detect them. There are two actors in these scenarios: the client and the company.

2.3 Declared identity

The company requires from the client to declare his identity in a form of text input.

The client can declare a different identity, real or fake.

Problem: The company has no way to verify that client's data are correct.

The client has no responsibility if they declare a different identity. No one can accuse him because there is no link between the real person the the fake one. No proof that the real person did this. Anybody can declare my name on a website, that doesn't mean I'm responsible for it.

Example:

- creating an email address at google.com
- PayPal used to function like this in the past

Conclusion: It is not enough to let the client tell us who he is.

2.4 Upload identity document

In order to prevent the previous kind of forgery the company can require an image of an identification document.

The client must do an effort to produce an image of a different document or to fake his own document image.

Does not need to actually print the document, only to fake a digital image.

If a person obtains an identity document of a different person (for example by physically stealing one or just capture the image), he can present it to the company. This is a form of identity theft.

Problem: The company has no way to determine if the person behind the account is the person from the identity document.

The client can change his own identity document by changing the name, person number and document number. Changing data from an image of a document it's much easier compared to changing a printed document.

The present technology to change an image is available as free software (https://www.gimp.org) so anybody having basic computer operating knowledge can change an image. **Changing the image preserves the security features in the document image**. Any software checking those features will validate the document as genuine.

For example, we can check the hologram manual or automatically and confirm that this is a photo of a genuine document. Unfortunately, the company has no way to test if the text data was changed.

Genuine

Fake (the image of the hologram has not been tempered)

*Personal data have been erased in this example.

As you observe, the name has been changed in the image but the security elements, the hologram and the phantom image in the right are still present.

Problem: It's very hard for the company to detect if the text from the document has been changed.

Conclusion: Is not sufficient to get an image of an identity document.

2.5 Upload selfie

In order to prevent identity theft, the company could require the client's selfie.

The client can send to the company a selfie of the document's owner. This validation is not foolproof but theoretically it should be harder for the client to have a selfie with the person he is impersonating.

The client still can produce a selfie from social media or some other form of social engineering.

Problem: The company cannot test if the user sends his own selfie.

This method doesn't help at all if the client changes the document data and keeps his own face. He/she will not be found on black lists with the new name.

Conclusion: The uploaded selfie is not a strong security element.

2.6 Upload selfie with id document

In order to prevent the user stealing a selfie from a different person and thus consolidating the false identity, the company requires from the client a selfie showing his document in the same image. This way it is harder for the client to obtain a selfie with the document. It cannot steal it from social media or some other form of social engineering.

The company can compare the uploaded image of the document with the document from selfie and can validate that:

the client's photo from selfie match the photo from the uploaded document and with the photo from the selfie.

the uploaded document and the document from selfie must be identical

It is harder for the client to forge a document because now is not enough to forge the image of the document. The document must be produced (printed) and laminated in order to show it in the selfie. The legal implications of producing a false document are higher than just changing its image.

Problem: The company has no way to verify if the document from the selfie is authentic. The client will upload the changed image with the security elements intact and will show a printed version of that image knowing that is very hard (next to impossible) to validate the security elements of the document form the selfie.

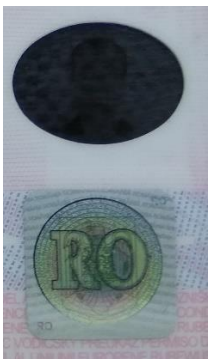
Conclusion: If we let the client upload the image of the document, the image might be digitally tampered.

2.7 A controlled image capture process

In order to prevent the client editing the image of the document, the company must control the document photo capture process and restrict the client access to the digital image. This way the client has no way to edit the image digitally and is forced to print the fake document, then photograph it. The printing process destroys the security elements and gives the company a chance to discover the counterfeit.

Photo of genuine document

Photo of a re-printed document



Security element in the genunine document

Security element in the reprinted document

One can observe the differences between the photo of a genuine document and the photo of a re-printed document.

Problem: The client can still digitally forge the selfie and steal the physical document.

2.8 Controlled selfie and photographic process

If the selfie photography is controlled by the company and the client has no way to interfere with the digital image, the document in the selfie has no use any more.

If the company controls the selfie and document photography process, then the client is forced to be present and to make a fake document almost identical with the original. The

fake document must contain the security features the original has. This is indeed a difficult process and can discourage many clients from doing so.

Controlling the selfie and photographic process is done via a secure mobile application. The client will use this application to make the selfie and scan the document. The application will encrypt the files and send them directly to the company.

Problem: The client can make the selfie of a paper photo or a mannequin with his face.

2.9 Live identification with live operator

In order to prevent the client making selfie of a mannequin the company can impose a live identification. In the live process, the company may ask the client to show his identification document and make a comparison between the photo on the document and the face of the client. Currently I don't know of any system able to do this automatically due to the low resolution of the photo from the document. The comparison can be done manually by an operator but that means higher costs for identification.

Even if we are sure about the face of the client, we cannot be sure about his identity because the document from the live session may be fake.

Advantages:

- the professionalism and experience of the human operator may prove valuable in detecting fraudsters
- the human operator can instruct the client to present the document in different angles to the camera. Some security elements change their appearance when changing the angle, the light falls on them. Still, being a low-resolution image (film has a lower resolution than photo on an mobile phone) it is easier to fake the document compared with real life. Other security elements (for example micro text or curve lines) cannot be verified due to the low resolution of the image.

Problems:

The client may use a fake document in the live session. The company cannot check all the security elements because it doesn't have a high-resolution image of the document.

The process is expensive for the company and cumbersome for the client. Statistics show that only 30% from the clients starting the identification process, actually finish it. Many of them can't hear or understand the operator instructions or his language. Many of them have problem holding the camera in one hand and the document in the other hand while fitting the head in the middle of the picture. Problems have been reported also from persons with handicap who see this process discriminatory.

Example: www.idnow.io

2.10 Live identification with a robot

In order to verify the security elements of the document, the company require the high-resolution upload of the document and controlled photographic process.

In order to ensure the selfie is from a live person (the client), the company controls the selfie process and instructs the client to pronounce out loud a unique generated code. This gives the opportunity to check the video with a software and able to correlate the sound with lips movement. For correlation between the client from the movie and the client from the uploaded document a face recognition software can be used.

Problem: If the company is doing all the verifications and correlations is quite difficult for a client to fool this process. Because the entire process can be a little harder for the client to complete the company has to make a business decision weighing on the one hand, the risk of enrolling false identities and on the other hand, the risk of losing a few clumsy customers.

Conclusion: This process seams the best we can do with the actual technology.

Example: www.onfido.com

2.11 Consistency checks on the document

Data on the document is printed in a specific way with some redundancies. The producer of the document wants to ensure that if the data is changed by unauthorized persons, there is a way to test that. Also, the producer intends to make the editing process harder using redundancies. This way the forger is required to make changes in several places thereby amplifying the probability of making a mistake.

Every document has specific features. Let's take the Romanian National ID document and see what are the consistency data checks:

The national personal identification number (CNP) has a checksum. Changes of the digits renders the whole number invalid.

The national personal identification number (CNP) contains personal information like sex, data or birth, birth location, etc. The information should be correlated with the text from this document.

The document has a machine-readable zone (MRZ). This zone contains information about person like name, data of birth, etc. The information should be correlated with the text from this document and also with the CNP number.

The document should be valid (not expired)

2.12 Security elements checks on the document

The Romanian national ID is a weak document in terms of security features but still has the essential elements to make the forger life harder. The tests could be:

Verify that the curve lines patterns under the characters is consistent with the model outside the text. The easier form of text editing is by copying digits or letters from the same document. Unfortunately, the curve pattern behind other digits is different. By simply copying them, the pattern is broken and this is observable.

Verify that the ghost image is similar with the user image and with the selfie image. No software to date can match that. This verification needs a human operator.

Verify the ghost image and it's background.

Check for strange artifacts around the photo or the text.

3 System Components

Security elements in a virtual world

The security features of the printed identification documents are conceived to be inspected on the physical document. Now, the evolution of industry is toward online that's why a lot of use cases involves verifying the documents online. Some of the security documents hold in this environment, some don't and we can start asking what other security elements could be invented to help authenticate a document after it's picture.

Security Element	Security for physical inspection	Security in controlled image capture	Security in uploaded image
Support: polycarbonate	medium - high	low	none
Laser printer	high	low	none
Microtext	medium - high	low	none
Transparent window	high	medium	none
Hologram	high	high	none
UV print	medium - high	none	none
Ghose image	high	medium	none
Light reflection elements	high	medium	none

Controlled image capture is the process of capturing the image of a document using a specialized software. The user cannot change the captured image.

Uploaded image is when the user makes his own photo of the document then uploads the file into the system. The person can do whatever he wants with the image even submit it to experts to change it.

Most of the security levels (Level 1 – on entry, Level 2 – back office, Level 3 – forensic lab, Level 4 – manufacturer) are irrelevant in a scanned image. It's obvios that a new security element is needed for online verification. A security element which can also be verified if the document is not physically present. The element must be readable and testable on the document's image sent over the Internet. A form of digital watermaker can be used. For example, a cryptographic hash of the text and photo. This hash can be encoded as a string of characters or as a computer readable geometrical shape and can be printed on the document.

The verification process will compute the hash from the document's image and then compare it with the hash engraved on the document. If they match, there is a good indicator for document integrity[1].

[1] https://www.researchgate.net/publication/245930357_Digital_watermarks_as_a_security_feature_for_identity_documents
https://www.locklizard.com/document-watermarking

How this system can be better than other similar systems?

From the tested systems we find that only the MRZ zone is read leaving room for a lot of improvement.

A full document OCR reader could correlate automatically data from the document with MRZ data and text data.

A face recognition software could automatically correlate the selfie face with the document face.

A face presence software could automatically validate the selfie

A new innovative process of mobile image capture is able to check the hologram, the transparent window and any metallic deposits on the document.

The principle is that we have to check using best of our knowledge, all the data from the client. Other systems evaluated test only a some of the security elements and do not make full correlations intra and inter documents.

Supported documents

Supported documents might be:

- Passport
- Driver's License
- National Identity Card
- Residence Permit Card

3.1 Electronic Identification Solution

The Electronic Identification Solution (eID) may contain the following components:

- Controlled document image and selfie capture
- Serverside analysis service

Captures the image of the identity document using a specialized app installed into the smartphone.

Type: mobile app and server service

Users: persons subjects to identification process

Beneficiaries: organizations in need to enroll and identify persons not physically present

Scope: Identity documents have to be scanned and sent in a secure manner to the organization.

Generally, a real time algorithm is used to detect and avoid glare and light burns when capturing the image of a document. A new method of document image capture using a mobile photo can be used to farther validate the document authenticity. Two pictures are taken in a short time interval, one without flash one using the built in LED flash. A comparison between photo taken without flash and photo taken using flash reveals the subtle changes of security elements in function of the angle the light falls over the document. A fake, printed and laminated document would be immediately set apart from the genuine document.

The online identification process:

- User uploads a high-resolution photo of his document in a controlled image capture.
- A video channel is opened between the user and a recording robot.
- The robot validates the liveliness presence of the user in the video.
- The robot asks the user to pronounce a unique generated code[iv]
- The robot automatically validates the code according with the sound.
- The robot confirms that lips movement match the sound within a reasonable probability.
- The system reads the machine-readable zone and human readable zone from the uploaded document and cross-validates.
- The system assesses the authenticity of the document inspecting the security features.
- The face from the document is matched automatically with the face from the movie (1 to 3 good quality frames).
- As an intermediary optional step, the system can validate the user phone number, the user email and the user street address using ancillary documents.

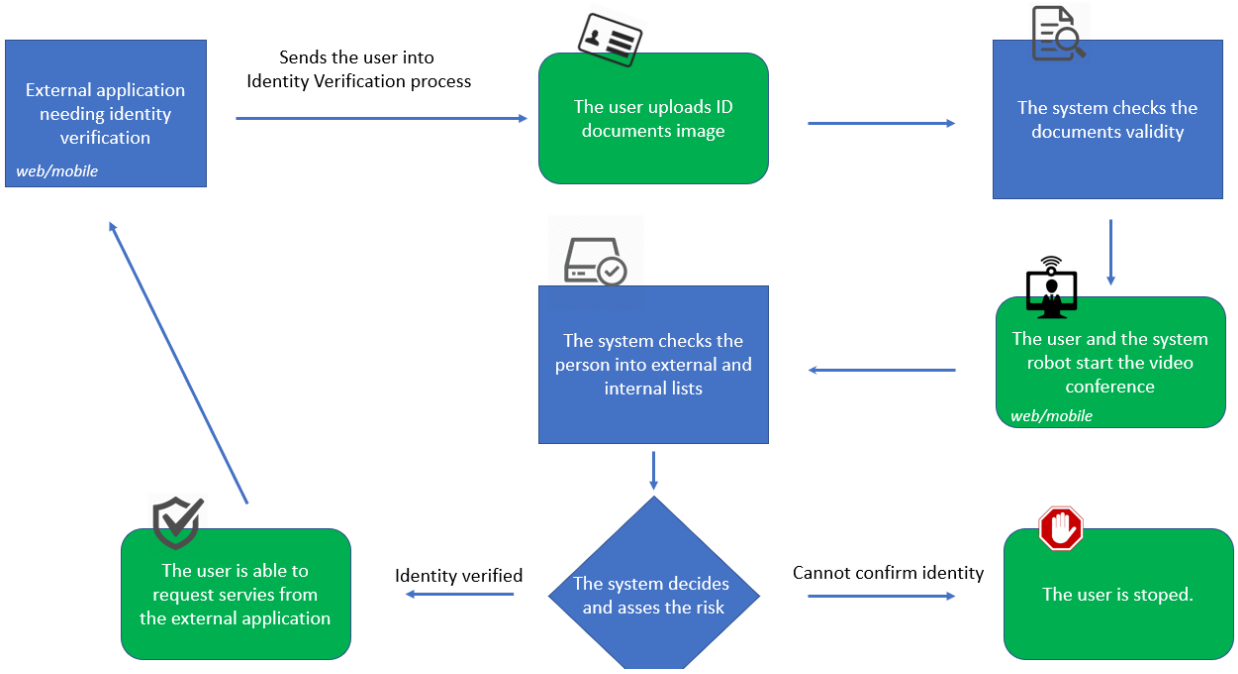


Diagram 1: Online identity verification use case

3.1.1 Selfie capture

Captures the person's photo in a secure manner. Don't let the user change the image in the process.

Type: mobile app

Users: persons subjects to identification process

Beneficiaries: organizations in need to enroll and identify persons not physically present

Scope: Match the photo from the document with the person's photo in order to prevent identity theft. Confirm the document belongs to the person on the other end.

3.1.2 Upload and verify digitally signed auxiliary documents

Receive and test digital copies of auxiliary documents like utility bill or bank statements. Usually the electronically issued invoices are digitally signed by the issuer using his certificate. This way, the document can be verified and any modification discovered. Unfortunately, not many eIDV providers do this.

Type: web app, mobile app

Users: persons subjects to auxiliary identification process like EDD (Enhanced Due Diligence)

Beneficiaries: organizations in need to verify the address of the person or to test additional documents for persons or companies.

Scope: Match the name of the person with address, confirm the address of a person or a business.

3.1.3 Live face capture

A video conference is opened with the person on one side and a software robot on the other side. The robot tests the live presence of the person giving instructions to slightly rotate the head and to pronounce a unique generated sequence of numbers or letters. The robot may correlate the image of the face with the photo from the document and may confirm that the sequence is correctly pronounced using a lips movement recognition algorithm.

Type: mobile app

Users: persons subjects to a high security identification process

Beneficiaries: organizations in need to enroll and identify persons in the most secure way.

Scope: if a controlled selfie capture process is not enough in order to prevent identity theft as the user may present a picture or a mold of a different person.

3.1.4 Physical document verification app for organizations

Organizations may need a tool to verify documents physically present and rely less or help the human. The operator can evaluate the document manually but as a supplementary verification, use the smartphone to test it. This product has all the functionalities of the document reader but is just a software. The hardware is a high-performance smartphone.

Type: mobile app

Users: physically present persons subjects to identification process

Beneficiaries: organizations in need to identify persons on site

3.1.5 Backend and API interface

Document image analysis is made in a secure environment. The organization submits the document image through a secure API call and the result of the verification is returned asynchronously. The application contains a document analysis algorithm.

Type: server-side application

Users: organizations that want to analyze the documents in depth in a secure environment.

Beneficiaries: organizations in need to check images of the documents

3.1.6 Document type enrollment back office app

The eID system should accommodate many document types. In order to accommodate a new document, the system has to learn the document features, location and any digital watermarking it might have. The software can use ML[2] techniques to enroll new document types.

Type: Backoffice web app

[2] Machine Learning

Users: security personnel, experts

Beneficiaries: owner of the eID system

4 Opportunities

- Current industry shift towards online services
- Current high demand for secure services and companies to trust
- Sophisticated technology is now available for counterfeiters
- Very high rate of development of online banking services
- With the advent of PSD2[v] in Europe and the correspondent trends in US, financial services will shift towards a mobile, more integrated solution. The role of security suppliers will become even more important than before.
- According to the EU electronic identification and trust services (eIDAS) Regulation, described as a pan-European login system, all organizations delivering public digital services in an EU member state shall accept electronic identification from all EU member states from September 29, 2018[vi]

5 Examples

idnow.eu

Provides live video identification, the user specifies personal data in a form. The data is validated and saved then a video conference is started with a live operator. In the video session the subject is asked to show his ID to the camera. In order to assess the risk of forgery, the user is asked to move the document in front of the camera with photo light on. The operator is trained to observe some security features on the document in those conditions.

sumsub.com

electronicid.eu

trulioo.com

Requires from the user the personal data, a high-resolution scanned ID document and a selfie with the ID document. Trulioo assess the forgery risk by analyzing the uploaded image. The match between person photo and document photo is automatically done using AI. Users are complaining that the process is cumbersome and only 30% of sessions are finished with a positive identification because of various technological problems.

onfido.com

Requires from the user the personal data including upload with high resolution scan of the ID. After saving the information, a short video session is started (mobile or web) where the user is asked to pronounce loudly and clearly four digits. Automatic systems asses the risk and confirm the match between photo from the document and live movie. The code is unique excluding the risk of user presenting a pre-recorded video.

complyadvantage.com

au10tix.com

Mitek (www.miteksystems.com)

Actuant (www.acuantcorp.com)

5.1 The prototype and the MVP

A prototype can contain only he document analisys algorithm because this is the essential part of the system and one that needs deep research efforts.

An MVP should contain:

- the mobile app for controlled document image capture and selfie
- backend API interface and document analisys algorithms
- minimal backoffice administration module

[i] https://en.wikipedia.org/wiki/AMLCFT
[ii] https://en.wikipedia.org/wiki/Blue_Ocean_Strategy
[iii] https://www.electronicid.eu
[iv] http://www.facebanx.com/product-liveness-detection.php;
https://www.researchgate.net/publication/228819905_Biometric_person_authentication_with_liveness_detection_based_on_audio-visual_fusion
[v] https://www.quora.com/Is-there-an-equivalent-of-PSD2-in-the-United-States
[vi] https://en.wikipedia.org/wiki/Electronic_identification