**ÇUKUROVA UNIVERSITY**
**INSTITUTE OF NATURAL AND APPLIED SCIENCES**

**MSc THESIS**

**Hazem Abdullah**

# IDENTITY DOCUMENT IMAGE ANALYSIS USING ARTIFICIAL INTELLIGENCE TECHNIQUES

**DEPARTMENT OF COMPUTER ENGINEERING**

**ADANA-2019**

**ÇUKUROVA UNIVERSITY**
**INSTITUTE OF NATURAL AND APPLIED SCIENCES**

# IDENTITY DOCUMENT IMAGE ANALYSIS USING ARTIFICIAL INTELLIGENCE TECHNIQUES

**Hazem Abdullah**

**MSc THESIS**

**DEPARTMENT OF COMPUTER ENGINEERING**

We certify that the thesis titled above was received and approved the award of degree of the Master of Science by the board of jury on

| | | |
|---|---|---|
| ……..…… | ……..……..… | …..……..……..……. |
| Assist. Prof. Dr. B. Melis OZYILDIRIM | Prof. Dr. S. Ayşe ÖZEL | Assist. Prof. Dr. Onur ÜLGEN |
| SUPERVISOR | MEMBER | MEMBER |

This MSc Thesis is written at the Computer Engineering Department of Institute of Natural and Applied Sciences of Çukurova University.
**Registration Number:**

**Prof. Dr. Mustafa GÖK**
**Director**
**Institude of Natural and Applied Science**

**ABSTRACT**

**MSc THESIS**

**IDENTITY DOCUMENT IMAGE ANALYSIS USING ARTIFICIALINTELLIGENCE TECHNIQUES**

**Hazem Abdullah**

**ÇUKUROVA UNIVERSITY**
**INSTITUTE OF NATURAL AND APPLIED SCIENCES**
**DEPARTMENT OF COMPUTER ENGINEERING**

Supervisor : Assist. Prof. Dr. B. Melis ÖZYILDIRIM
Year: 2019, Pages: 55
Juries : Assist. Prof. Dr. B. Melis ÖZYILDIRIM
: Prof. Dr. S. Ayşe ÖZEL
: Assist. Prof. Dr. Onur ÜLGEN

Identity document understanding is one of the most important parts in the document recognition systems, even though that many researches are done in "document image analysis", however this research still has many challenges. This research aims to find a better solution to analyse the identity document image using artificial intelligence techniques. Hence, proposed method detects the location of the important information in the identity document, classifies the text to many categories (date of birth, last name, first name...etc.), and detects the key objects in the identity document image like the face photo and signature and even the logos.

Methods used in this research are divivded into two categories which are document type based and machine-readable zone (MRZ) based approaches. In the document type based method, first the document image is classified, then pre-trained model for each class is utilized to derive the information needed such as location and text segments for that class on the query image. In the MRZ based approach, first MRZ location is detected, then Visual Inspection Zone (VIZ) area zone with the MRZ information are matched and small Natural Language Processing (NLP) engine is utilized to detect text.

The accuracy of the research was 94.8% with accurate detection for all segments in the document, and 3% with good detection. However some information in the document image was missing, resulting in 1% of wrong detection as false positive, and the last 3 was not detected at all. Consequently, the overall accuracy of the research was 97.8% of the test samples.



**Keywords:** Identity document, Text Segmentation, feature extraction, document classification, Document understanding

# ÖZ

## YÜKSEK LİSANS TEZİ

**YAPAY ZEKA TEKNİKLERİ KULLANILAN KİMLİK BELGESİ GÖRÜNTÜ ANALİZİ**

**Hazem Abdullah**

**ÇUKUROVA ÜNİVERSİTESİ**
**FEN BİLİMLERİ ENSTİTÜSÜ**
**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

Danışman : Dr. Öğr. Üyesi B. Melis ÖZYILDIRIM
Yıl: 2019, Sayfa: 55
Jüri : Dr. Öğr. Üyesi B. Melis ÖZYILDIRIM
: Prof. Dr. Selma Ayşe ÖZEL
:Dr. Öğr. Üyesi Onur ÜLGEN

Kimlik belgesi anlamlandırılması, belge tanıma sistemlerinin en önemli alt alanlarından biridir ve birçok "belge görüntüsü analizi" çalışmaları yapılmasına rağmen, bu araştırmanın hala birçok zorluğu vardır. Bu çalışmada, yapay zeka teknikleri kullanılarak kimlik belgesi görüntüsünü analiz etmek için daha iyi bir çözüm amaçlanmıştır. Bu nedenle, bu araştırma, kimlik belgesindeki önemli bilgilerin yerini tespit etmeyi, metni birçok kategoriye sınıflandırmayı (doğum tarihi, soyadı, ad, vb.) ve kimlik belgesindeki yüz fotoğrafı, imza ve logolar gibi anahtar nesneleri tespit etmeyi hedeflemektedir.

Bu çalışmada, kimlik belgesi görüntüsünü analiz etmek için belge tipi tabanlı ve makine tarafından okunabilir bölge (MRZ) tabanlı iki ana yöntem kullanılmaya çalışılmıştır. Belge türüne dayalı yöntemde, önce belge görüntüsü sınıflandırılmaya çalışılmış, sonra her sınıf için önceden eğitilmiş modele bağlı olarak o sınıf için gereken bilgi, konum ve metin bölümleri sorgu görüntüsünde işaretlenmiştir. MRZ tabanlı yöntemde ise MRZ konumunu tespit etmeye çalışıldıktan sonra Görsel İnceleme Bölgesi (VIZ) alanı ile MRZ bilgisini küçük ölçekli Doğal Dil İşleme (NLP) motoru kullanarak eşleştirilmektedir.

Önerilen çalışma ile, belge tipi tabanlı yöntem kullanılarak dokümanların %94.8'i yüksek doğrulukta, %3'ü iyi doğrulukla analiz edilmiştir. Yüz ve bazı logolar hariç herhangi bir örüntü bilgisi tespiti bu çalışmada yapılmamıştır. Test örneklerine göre %97.8 oranında başarılı analiz elde edilmiştir.

**Anahtar kelimeler:** Kimlik belgesi, Metin Bölümlendirme, özellik çıkarma, belge sınıflandırma, Belge anlamlandırılması

## EXTENDED ABSTRACT

Everyone in this world has an identity document card to prove their identity and their personal information such as name, address, date of birth and their national identity number. Identity document reading systems will help to extract the information from the identity document and digitize the document information to the computers. These systems will have a record for each person, and this can help to do the required action for criminal people who has a criminal record.

The aim of the study is to help the identity document systems analyze the identity document layout all over the world.

First, it needs to analyze all textual information in the identity document image, detect the textual information location and the textual information area, then lable the detected text and information to the expected lables that usually exist in the identity documents like first name, last name, document number, date of birth, expiry date, issue date, address and personal number.

To detect the text in the document the reseach depends on Connectionist Text Proposal Network (CTPN) that accurately localizes text lines in natural image, some customization has been used on CTPN method in order to use it in the identity document images. Even the text detection results were promising, but it has some limitation to detect the text on the identity document images such as detecting some false positive cases. This issue can be solved using image processing to filter the detected shapes and remove them depending on their positions and colors. Another issue from the text detection is to merge the textual information belonging to more than one segments together. After text detection, filtering the false positive detected text, and spliting the regions, it searches for one character in the image. Hence, an OCR and NLP engine are required on the detected text to classify the text.

Another proposed approach utilizes two main methods to analyze the document: identity document predefined model, and ICAO generic model.

The test data set contains 1858 samples collected from over 800 different document types, different light conditions and different orientation that cover most important documents all over the world (more than 100 countries).

The overall accuracy was 97.8%. In the first experiment 1534 sample images out of 1858 have been detected correctly, while 324 of the samples were not detected at all.

After applying the rotation method, adjusting the threshold and adding the MRZ module 1817 samples detected correctly. Hence, the over all accuracy is 97.79%.

According to the new data protection regulation, data protection and privacy laws prevent the test results to be compared with other systems.

**ACKNOWLEDGEMENT**

First of all I would like to thanks my supervisor Assist. Prof. Dr. B. Melis ÖZYILDIRIM for her kindness and her support.

I would like to thanks GBG for the assistance with the samples used in this research, and sure big thanks for GBG-Turkey management team for their support and motivate me when I need help.

And sure big thanks to Prof. Dr. Selma Ayşe ÖZEL and Assist. Prof. Dr. Onur ÜLGEN for their information and supports.

And finally, I wish to thank my family for the continued support.

**CONTENTS** **PAGE**

**LIST OF TABLES** **PAGE**

**LIST OF FIGURES** **PAGE**

**ABBREVIATIONS**

| | |
|---|---|
| AI | :Artificial Intelligence |
| VIZ | :Visual Inspection Zone |
| MRZ | :Machine-Readable Zone |
| NLP | :Natural Language Processing |
| DL | :Driving License |
| ID | :Identification Card |
| ICAO | :International Civil Aviation ORG |
| OCR | :Optical Character Recognition |
| United States of America | :USA |
| United Kingdom | :UK |
| European Economic Area | :EEA |
| DAR | :Document Analysis and Recognition |
| MSER | :Maximally stable extremal regions |
| CTPN | :Connectionist Text Proposal Network |
| T | :Template |
| CNN | : Convolutional neural network |
| SIFT | : Scale-Invariant Feature Transform |

# 1. INTRODUCTION

## 1.1. What is the identity document?

According to the oxford dictionary the Document is a piece of written, printed, or electronic matter that provides information or evidence or that serves as an official record. The identity document or also known as identity card is an official document or card consisting of the holder's name, date of birth, photograph, or other information. It's a piece of certification issued by the authorities for providing evidence of the identity of the person carrying it. Some of documents considered as identity documents are passport, driving license, student card, national identification card, and many others. A modern identity document contains biometric information, such as fingerprints, photographs, and face, hand, or iris measurements. An identity document also provides additional information such as full name, parent's names, address, profession, nationality in multinational states, blood type, and Rhesus factor.

## 1.2. Types of identity documents

The identity cards these days come in a variety of materials, thicknesses, and sizes. The main two categories are the documents compliant with the International Civil Aviation Organization (ICAO) standard and the documents that are not compliant with ICAO 9303 standard.

The identity documents compliant with the International Civil Aviation Organization (ICAO) 9303 (ICAO, 2015) standard, such as passports and visas which contain machine readable zone area (MRZ) usually at the bottom of the identity card. MRZ contains some special characters known as fillers "<<". Figure 1.1. shows MRZ area and Figure 1.2. represents VISUAL INSPECTION ZONE (VIZ) area.

The second type of document are not compliant with ICAO 9303 standard and these documents do not contain the MRZ area like the driving license. An example is given in Figure 1.3.

Identity documents which are compliant with the ICAO 9303 standard comprise an MRZ usually at the bottom of a page of the document. MRZ comprises identity information. The MRZ may include different numbers of lines, depending on the type of the document. For example, a passport compliant with the ICAO 9303 standard will comprise a MRZ having 2 lines and some identification cards compliant with the ICAO 9303 standard will comprise a MRZ having 3 lines. The MRZs comprise machine readable information contained in lines of text capable of optical character recognition (OCR). Each line consists of several characters, depending on the type of document, representing information such as document type, country code, primary and secondary identifiers (names), identity document number, nationality, date of birth, sex, date of expiry and check digits. The information is spaced with one or more filler characters, such as <. Figure 1.4. shows examples compliant with ICAO 9303 standard. In some documents the front side of the document is not an ICAO standard where the back side is an ICAO standard like the national identification card in Turkey. An example is given in Figure 1.5.

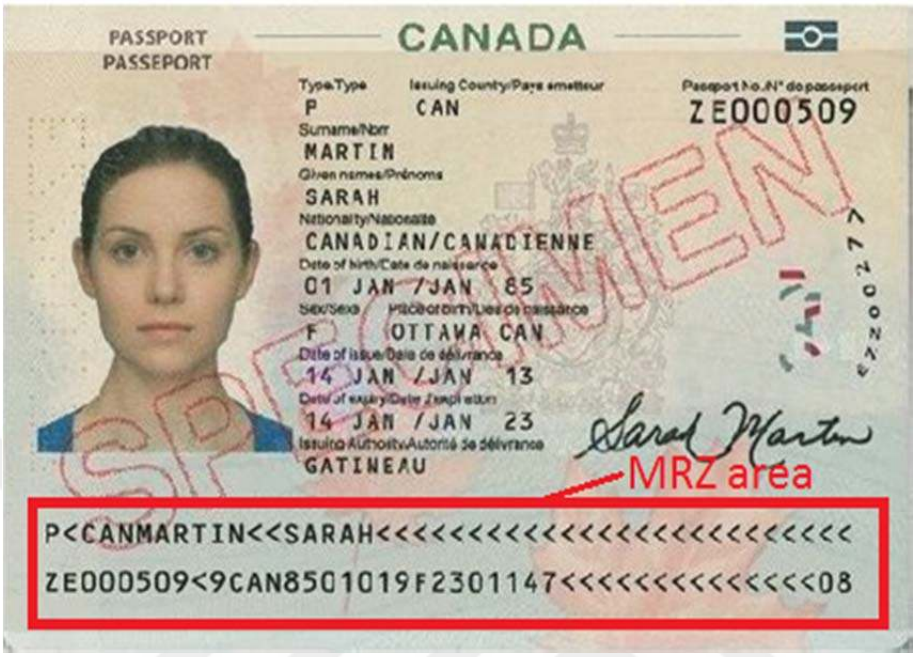


Figure 1.1. The identity document which shown the MRZ Area


Figure 1.2. The identity document which shown the MRZ area and VIZ area


Figure 1.3. An example about non-ICAO document type – Turkish Driving license.


Figure 1.4. Two examples of 2 lines MRZ and 3 lines MRZ


Figure 1.5. An example on non-ICAO document type front side and ICAO document back side.

### 1.3. Research Aims and Objectives

Today, many digital A4 scanners, cameras, and other types of document scanning systems are available to capture hundreds of identity documents every day. Extracting the important information from these documents is still challenge for most of the document extraction/recognition systems.

Most of the current document extraction/recognition engines still depend on manual segmentation and localization, that lead to a lot of efforts, wasting of time, and opened to human error. Hence, the previous approach needs a lot of manual effort to analyze the document image, find the textualinformation, and find the key areas in the query image like face and signature to analyze the data by another system or manually. The human efforts cost time and increases the errors.

All the previous issues affect the accuracy of the important information that will be extracted from the system later.

The purpose of this study is to fill the gap in the previous approach, by detecting the important and key information from the document image and labeling

these information and areas in the document image such as names, date of birth, passport number. The process can be different from one document type to another i.e. some documents contain specific information that may not exist in other documents like T.C. number in Turkish national identification card. Moreover, other kind of information may exist in the document and it's important to know for validation and authentication purposes such as face, signature, some logos and unique areas.

All that may help the identity document authentication and extraction systems to read the detected information from the document, authenticate the document image and even classify the document.

### 1.4. The data set

This research uses some image samples from GBG-Idscan document data set. These images were scanned by specific scanner to generate good quality images, the width and the height of these images are not less than 1100px and not more than 2200px.

This data set contains three main categories identification documents (national identification cards, and foreign identification cards), license (driving licenses, job licenses), and travel documents (passports, travel permits, and visas).

The sample set contains over 100,000 document samples, over 1000 document types, and it covers more than 150 countries all over the world including USA documents, UK documents, EEA documents, Turkey documents, Arab documents, Russian documents and Chinese documents.

The training set includes one image only from each document type or class (around 1000 different classes), the test set contains 2 to 10 images from each class. The total samples image set for testing are 1845 images more than 1000 document types, and it covers more than 150 countries and different lighting conditions.

# 2. RELATED WORKS

## 2.1. Background

Document understanding became one of the most important research in the document extraction, recognition and validation systems in the past few years, some researches were conducted to analyze the documents, detect the text and understand the layout of the documents. This literature review includes some researches that have been done for A4 document in Figure 2.1. MRZ document processing review, and few researches for Identity card processing.


Figure 2.1. An example of A4 document

### 2.2. A4 processing

The identity cards are considered as structured documents due to their predefined and stable layout. For instance, first name location in all Turkish passports that are issued in 2017 will be relevantly in the same location. Another type of document can be A4 documents that are used as proof of address like utility bills, electricity bills and other types of documents. A top-down data-driven approach was introduced to segment the structured A4 document consisting of simple background, mostly black text, signature and format tables (Cesarini, 2000). They present the X-Y tree where the regions splatted by meaning, the leaves describe the regions and the adjacency links among leaves of the tree describe local relationships between corresponding regions. This approach covers the invoice documents, utility bills and other types of documents that are mostly used as proof of address not proof of identity or identity cards.

In another approach Document Analysis and Recognition known as (DAR) system was built (Marinai, 2008). They aimed to extract the information presented on A4 paper and the outcome of the system presented by the ASCII format. Their approach mainly depends on analyzing the layout of the document image. They divided the document image into different regions which were classified then as signature area, graphic item area, and text area. They built a graphical shape detector, signature detector and text detector.

### 2.3. MRZ processing

One type of document known as the identity documents are compliant with the (ICAO) 9303 standard. These kind of documents such as passports contain MRZ. In some researches, document layout was analyzed, the MRZ data and VIZ area data were extracted.

One of the common approach is utilizing the OCR for the whole image then classifing the text in the image to label document. The text on the identity document containing MRZ was analyzed by detecting and extracting the

information from MRZ using OCR. Then, extracted results were matched with the information given in VIZ area. Regarding to passport processing, the researchers proposed an approach to analyze the passports layout, by processing MRZ area and VIZ area (Young, 2005). The approach recognizes the characters in the image using a predefined document template, then reads the MRZ fields using OCR. Since the OCR may cause some errors, some OCR corrections may be required in the extracted fields. Another OCR operation is applied on the VIZ area to read information and extracted information from MRZ and VIZ area were matched. Because of the complex background of the VIZ area, method was able to recognize only a few passports. Hence, in another study noise removal was proposed to increase the recognition rate.

### 2.4. Identity Card Processing

Regarding identity card document processing, some few researches have been done. Some of these researches aim to detect the text in the identity document which may help the OCR to recognize the information in the identity document image. A generic model was proposed to detect and recognize the text in the identity documents over the cloud computing. It combines the Maximally Stable Extremal Regions MSER, a locally adaptive threshold method for text segmentation, and a rectification correction using the Hough transform algorithm. That approach enhances the result of OCR, which will help to get better results for word recognition (Rodolfo, 2016). Their recognition process has some inaccurate results for images with complex backgrounds and different typographies.

# 3. MATERIALS AND METHODS

## 3.1. The text detection and Deep learning

In this study, the first approach was based on text detection algorithms to detect the text in the identity document image, classify the text in the image using text understanding (NLP) and analyze document layout. OCR was required to detect characters. Connectionist Text Proposal Network (CTPN) proposed is utilized to detect the text in the identity document image (Zhi, 2016). An example is given in Figure 3.1.

The text detecting results were promising however, it require an OCR engine and an NLP engine to classify the text in the identity document image.

Even the text detection results were promising, but it has some issues when it was applied on the identity document images. One of the issues was detecting some false positive cases such as some graphical shapes as text, given in Figure 3.2. This issue can be solved by applying image processing layers and filtering the shapes depending on their positions and colors. Another issue from the text detection is to merge the textual information that belong to more than one segments together, an example is given in Figure 3.3. Merging the textual information between labels and segments together is given in Figure 3.4. This issue may be solved by splitting the textual areas using image processing techniques such as color and space based splitting. Some texts also may not be detected depending on the current text detection especially if the text consists of only one letter such as gender field in passports. An example is given in Figure 3.5.

After the text detection and filtering the false positive detected text, the regions are splitted, searching is applied to detect one character fields in the image, and then OCR is applied on the detected text and NLP engine is build to classify the text.


Figure 3.1. The results of CTPN text detection applied on Turkish driving license

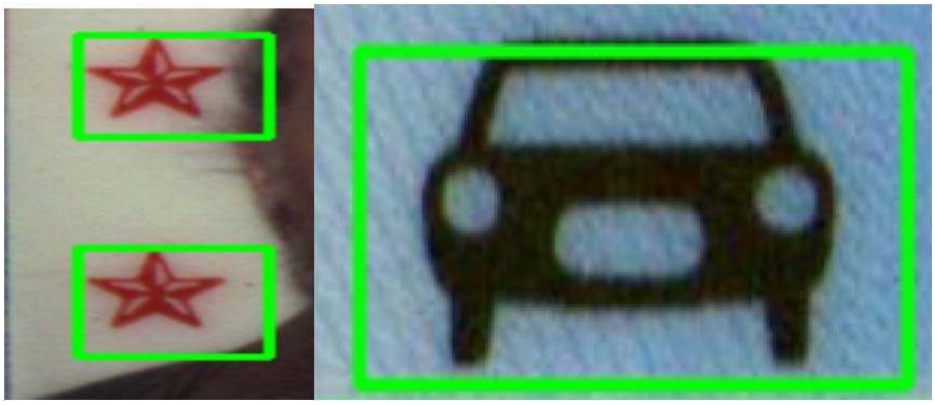

Figure 3.2. Some false positive cases on text detection method


Figure 3.3. An example of merging document number label with the document number segment as one area on the text detection method.

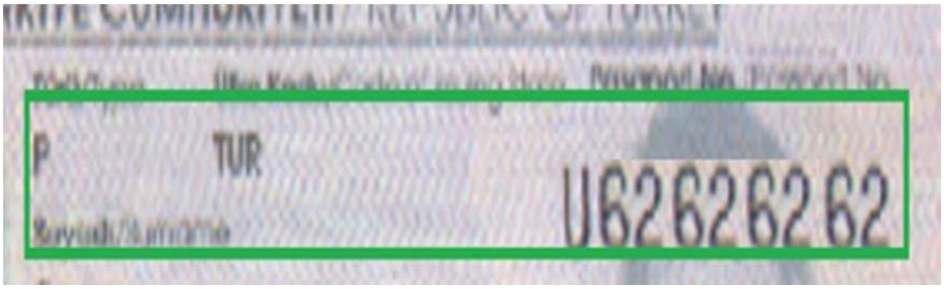


Figure 3.4. An example of merging more than one segment as one area on the text detection method.

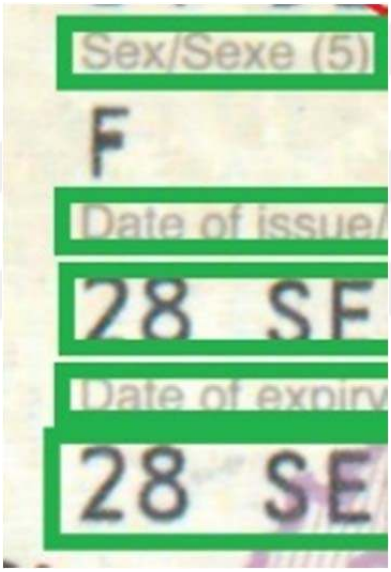


Figure 3.5. An example of not detecting one letter text in the text detection method.

Due to the huge processing time, the limitation in the text detection and the limitation in the NLP engine, some enhancement was required to reduce the processing time.

Deep learning has become useful approach for classification and segmentation problems. Hence, Convolutional neural network technique is utilized in this thesis. A pretrained CNN model providing high accuracy for real life images, deep face library, was utilized for face detection. However, its accuracy on

the identity document images has some limitations to detect the faces in case of reflection, blurriness, and for people who are wearing something on their heads like headband in some countries like Arab, Indian, and Chinese countires.

### 3.2. The identity document module and the MRZ module

The final approach depends on using two main methods: identity document predefined model, and ICAO generic model. To analyze the query image in this approach, the system expects a identity card image with good quality and scanned by a special scanner, as an input. The proposed method detects the important information such as first name, last name, document number, date of birth, issuing authority...etc., and also the face of the person, the signature and the logos.
The main steps are:

1. Calling the identity document trained module to find the correct predefined model that matches with the query image.
2. If the trained model does not exist, then the ICAO module will be executed.
3. The ICAO module will detect the MRZ in the image, then it will analyze both MRZ and VIZ area.

The next paragraphs will describe each process with some details:

### 3.3. The identity trained module

According to the described previous pipeline a trained model/template had been used for each document type. One sample from each identity document class is added to the class list, named as “typical image sample” and used for training. The key points are the top-left corner in the scanned image, which is *(0, 0)*, and the image width and height aer shown as *(W, H)*.

The model contains relevant location for each textual information in the document like the first name location, the last name location, document number location…etc., it also contains relevant location for the graphical information in the identity document like the location for logos, the face of the person, and the location of the signature. Recently, the non-English text location was added to the predefined model as well. The locations of each area are presented as relative to the location of the top-left corner, the width of the area and the height of it. All area's locations are stored by calculating the relevant distance between the top-left corner of the typical image sample *(0, 0)* and the top-left corner of the interested area that contain the required information *(x, y)*. The width of the interested area *w*, and its height is represented with *h*. If the top-left corner of the typical image is *(X0, Y0)* and the total width of the typical image is *W*, and the height of the typical image is *H*, the location of the first name relevantly will be *(x, y, w, h)* where *x* is the relevant location of the top-left point for the first name area on *X* axis, *y* is the relevant location of the top-left point for the first name area on *Y* axis, *w* and *h* are proportion of the relevant width and height of the first name area and the total width and height of the typical imag. Following equations show calculation of *(x, y, w, h)* for the first name area.

$$x = \frac{the\ horizontal\ distance\ between\ the\ topleft\ corner\ of\ the\ area - X0}{W} \tag{3.1}$$

$$y = \frac{the\ vertical\ distance\ between\ the\ topleft\ corner\ of\ the\ area - Y0}{H} \tag{3.2}$$

$$w = \frac{the\ real\ width\ of\ the\ area}{W} \tag{3.3}$$

$$h = \frac{the\ real\ height\ of\ the\ area}{H} \tag{3.4}$$

Horizontal - vertical distances and width-height for each area are calculated manually. An example is given in Figure 3.6. All areas in that document type are calculated in the same way and restored in a model belonging to the document type.

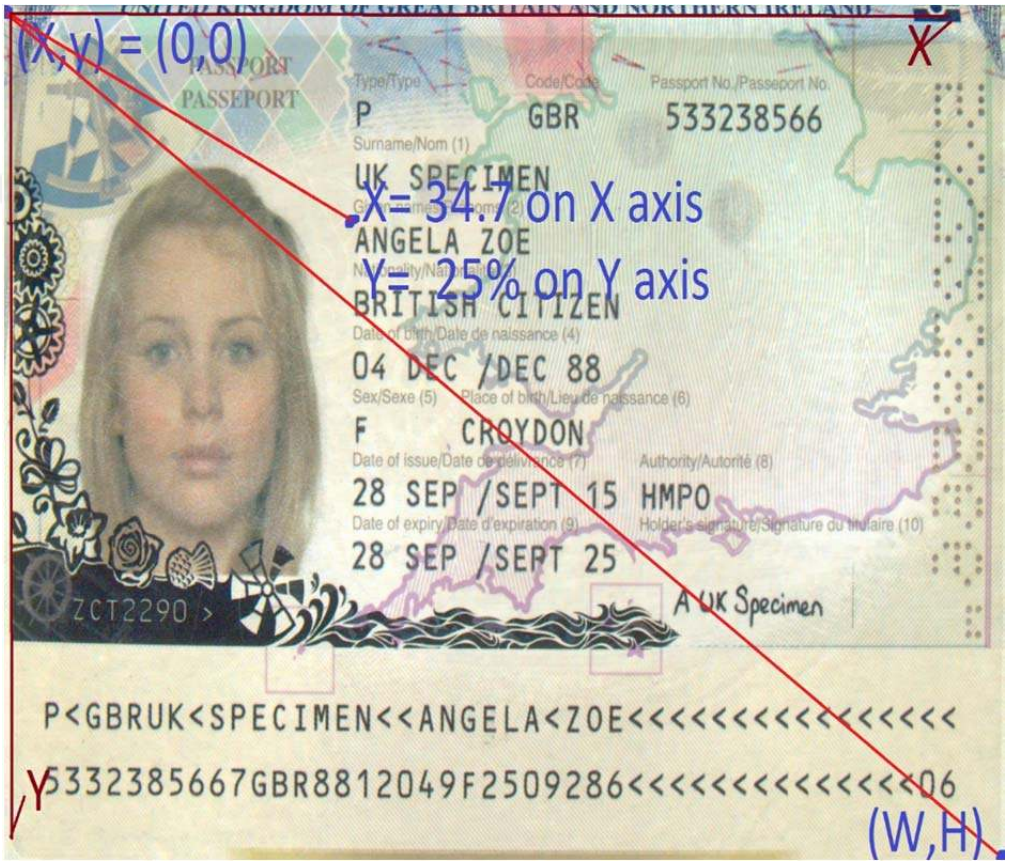


Figure 3.6. An example about how to calculate *(X, Y)* in the identity document image.

After generating the trained models, the system receives a query image. Using mathematical algorithms and trained model belonging to the document type of query image. For example, assume that query image is a Turkish passport, the width of the image is *Wq*, and the height of it is *Hq*. Trained model for Turkish

passport containing typical image, width *Wt,* height *Ht*, top-left corner and width-height data is uploaded to the system. If the last name area is presented as *(xt, yt, wt, ht)*, the last name location and size of the last name on the query image *(xq, yq, wq, hq)* are calculated as in 3.5-3.8.

$$xq = xt * Wq \quad (3.5)$$

$$yq = yt * Hq \quad (3.6)$$

$$wq = wt * Wq \quad (3.7)$$

$$hq = ht * Hq \quad (3.8)$$

Figure 3.7. shows calculation of the *(xq, yq, wq, hq)* for last name area in the Turkish passport.

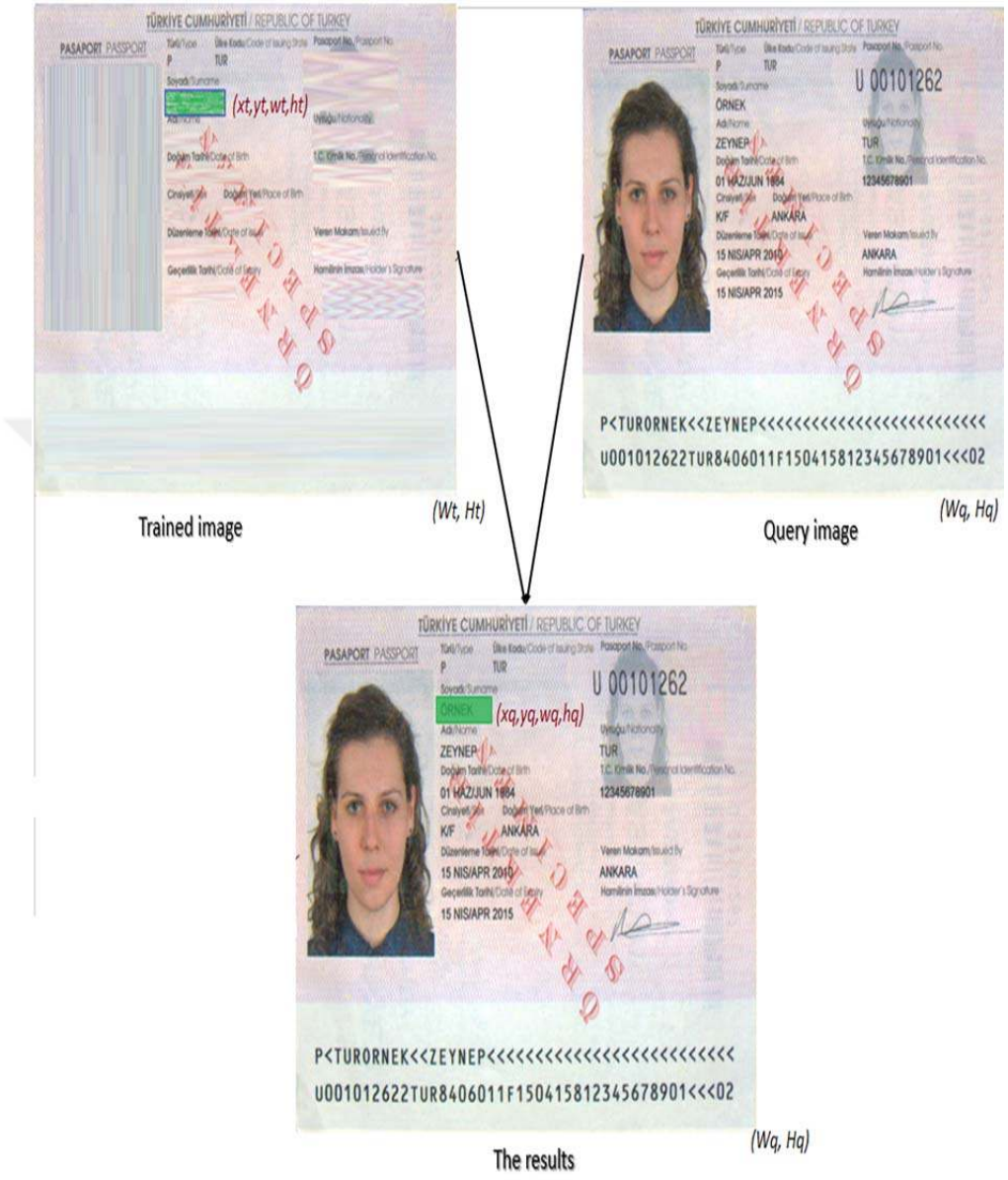


Figure 3.7. An example on how to calculate the *(xq, yq, wq, hq)* for last name area in the Turkish passport document image.

The previous approach has some limitations, and one of the limitations is when the query image is cropped from one of the parts especially the top part of

the document, the derived area in the query image will be shifted down according to the cropped part from the top. Figure 3.8. shows the issue occured due to the cropped part. This issue is solved by defining a key point in the typical image *(Xk, Yk)*, which is detected manually. Hence, instead of depending on the top-left corner in the typical image *(0, 0)*, *X0 = Xk* and *Y0 = Yk* are considered and all the areas and the segments are calculated according to the key point in the typical image after locating key point in the query image *(Xkq, Ykq)*. Figure 3.9. shows an example for the proposed solution.

Proposed study uses template matching algorithm to find key point of the template in the query image. A window with the size of template T is slided on the query image one pixel at a time and at each point a metric is calculated to decide if the sliding window contains the template or not (https://docs.opencv.org).

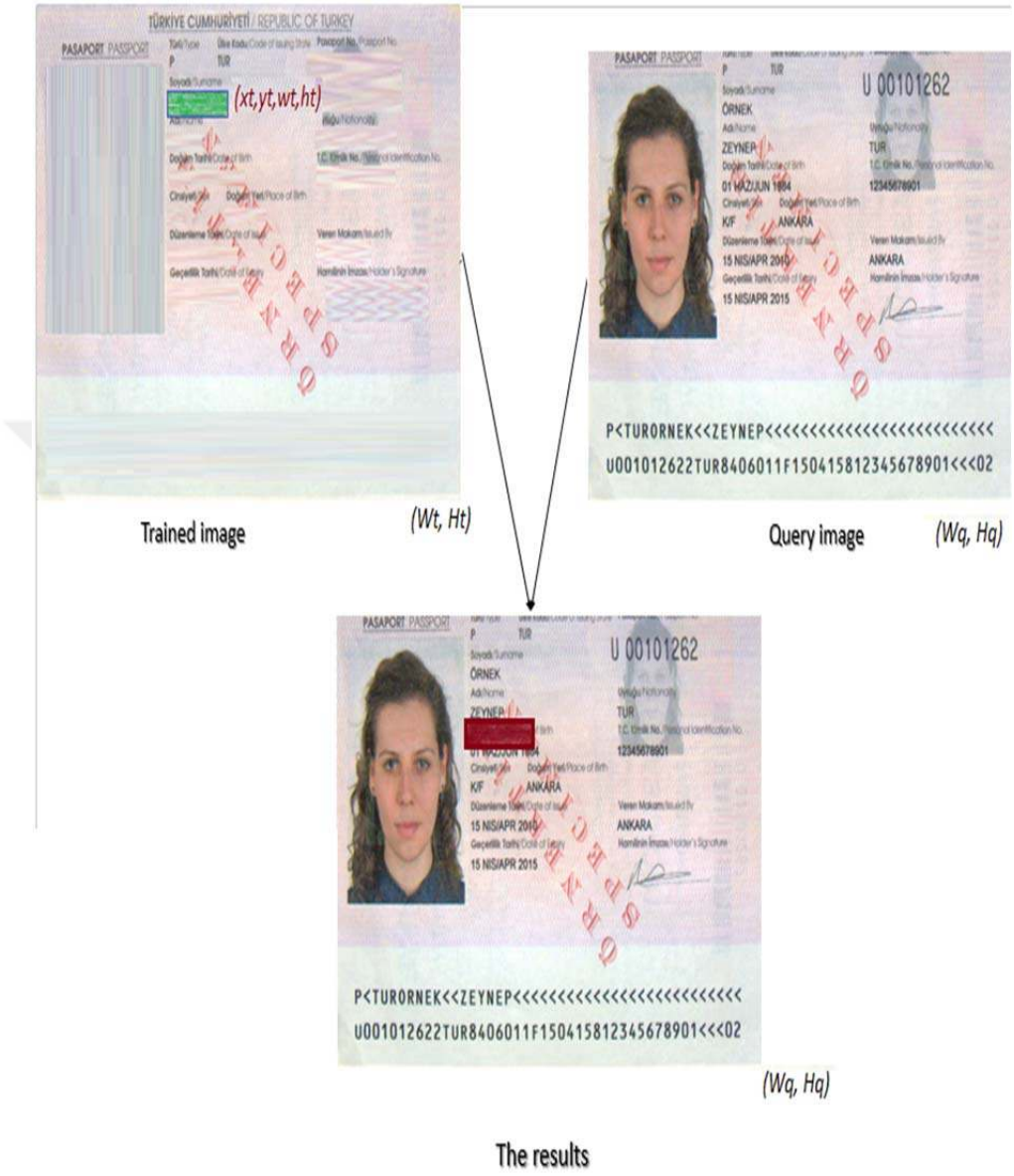


Figure 3.8. An example on incorrect area detection when the query image is cropped.

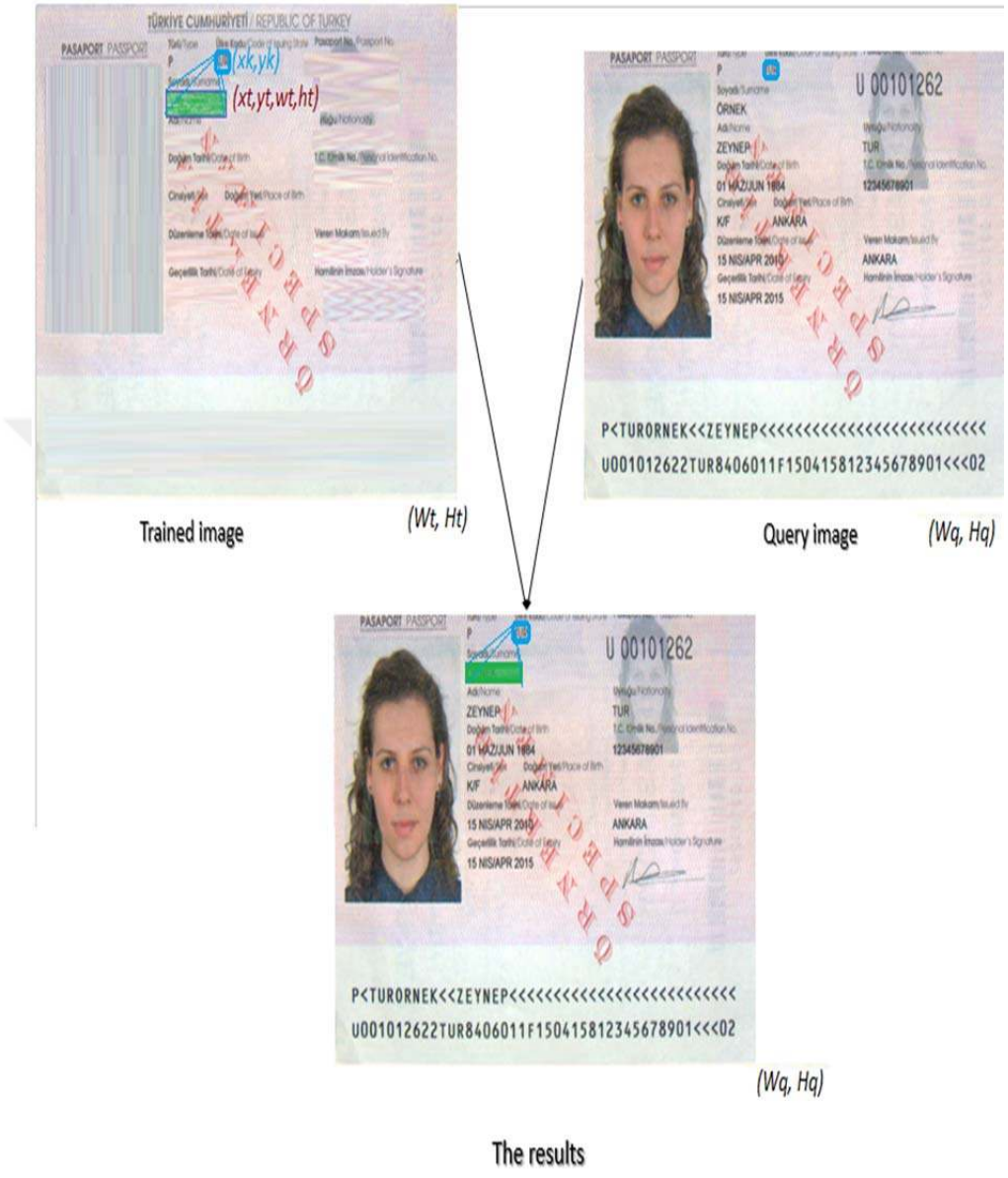


Figure 3.9. The results after applying the template matching method between the typical image and query image.

### 3.4. Feature extraction using SIFT

Objects have many features to be extracted for describing it. The scale-invariant feature transform (SIFT) is used to detect and describe the features in the images (David, 2004). It locates certain key points and then furnishes them with descriptors. The descriptors are supposed to be invariant against various transformations which might make images look different although they represent the same object. SIFT image features provide a set of features of an object that are not affected by many of the limitations experienced in other methods, such as object scaling and rotation. There are mainly four steps in SIFT algorithm:

1. Scale-Space Extrema Detection: Search over multiple scales and image locations

to find “characteristic scale” for features. This stage of the filtering attempts to identify those locations and scales that are identifiable from different views of the same object. This can be efficiently achieved using a "scale space" function. Further it has been shown under reasonable assumptions it must be based on the Gaussian function. The scale space is defined by the function:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (3.9)$$

Where * is the convolution operator, $G(x, y, \sigma)$ is a variable-scale Gaussian and $I(x, y)$ is the input image.

Various techniques can then be used to detect stable keypoint locations in the scale-space. Difference of Gaussians is one such technique, locating scale-space extrema, $D(x, y, \sigma)$ by computing the difference between two images, one with scale k times the other. $D(x, y, \sigma)$ is then given by:

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (3.10)$$

To detect the local maxima and minima of $D(x, y, \sigma)$ each point is compared with its *8* neighbours at the same scale, and its *9* neighbours up and down one scale. If this value is the minimum or maximum of all these points then this point is an extrema.

2. Keypoint Localistaion: Fit a model to detrmine location and scale. Select based on a measure of stability. This stage attempts to eliminate more points from the list of keypoints by finding those that have low contrast or are poorly localised on an edge. This is achieved by calculating the Laplacian, value for each keypoint found in stage 1. The location of extremum, $z$, is given by:

$$z = -\frac{\partial^2 D}{\partial \vec{x}^2}^{-1} \frac{\partial D}{\partial \vec{x}} \tag{3.11}$$

If the function value at $z$ is below a threshold value then this point is excluded. This removes extrema with low contrast. To eliminate extrema based on poor localisation it is noted that in these cases there is a large principle curvature across the edge but a small curvature in the perpendicular direction in the defference of Gaussian function. If this difference is below the ratio of largest to smallest eigenvector, from the *2x2* Hessian matrix at the location and scale of the keypoint, the keypoint is rejected.

3. Orientation Assignment: Compute best orientation for each keypoint region. This step aims to assign a consistent orientation to the keypoints based on local image properties. The keypoint descriptor, described below, can then be represented relative to this orientation, achieving invariance to rotation. The approach taken to find an orientation is:

- Use the keypoints scale to select the Gaussian smoothed image $L$, from above

- Compute gradient magnitude, *m*
-

$$m(x,y)=\sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2} \quad (3.12)$$

- Compute orientation, $\theta$

$$\theta(x,y)=\tan^{-1}\left(\frac{(L(x,y+1)-L(x,y-1))}{(L(x+1,y)-L(x-1,y))}\right) \quad (3.13)$$

- Form an orientation histogram from gradient orientations of sample points
- Locate the highest peak in the histogram. Use this peak and any other local peak within 80% of the height of this peak to create a keypoint with that orientation
- Some points will be assigned multiple orientations
- Fit a parabola to the 3 histogram values closest to each peak to interpolate the peaks position

4. Keypoint Descriptor: The local gradient data, used above, is also used to create keypoint descriptors. The gradient information is rotated to line up with the orientation of the keypoint and then weighted by a Gaussian with variance of *1.5 * keypoint scale*. This data is then used to create a set of histograms over a window centred on the keypoint. Keypoint descriptors typically uses a set of *16* histograms, aligned in a *4x4* grid, each with *8* orientation bins, one for each of the main compass directions and one for each of the mid-points of these directions. This results in a feature vector containing *128* elements. Use local image gradients at selected scale and rotation to describe each keypoint region.

These resulting vectors are know as SIFT keys and are used in a nearest-neigbours approach to identify possible objects in an image.

### 3.5. Searching for the correct model

Proposed study requires identification of the query image type. Image processing and machine learning techniques are used for identiciation process. In this step, features are extracted from the typical images, and then these features are used to train a Support Vector Machines (SVM) classifier. Feature extraction steps are shown in Figure 3.10 and the steps to train the classifier described in Figure 3.11. Proposed system extracts the features to produce a feature vector for each document type, extracted features are variant from document type to another. Hence, in this top 1000 features of documents have been considered as a feature vectors. Selected features are inputs for SVM classifier. Query images are applied to trained SVM classifier to detect the model.

Figure 3.12. shows the overall pipeline. First, the query image is classified by SVM classifier. If SVM cannot classify the document, it means unkown document type is applied and the process ends. Otherwise, the model belonging to that document type is loaded from the models pool. Using template matching and proportion calculations, all the locations and sizes for all defined areas are derived on the query image.


Figure 3.10. The steps to extract the features.

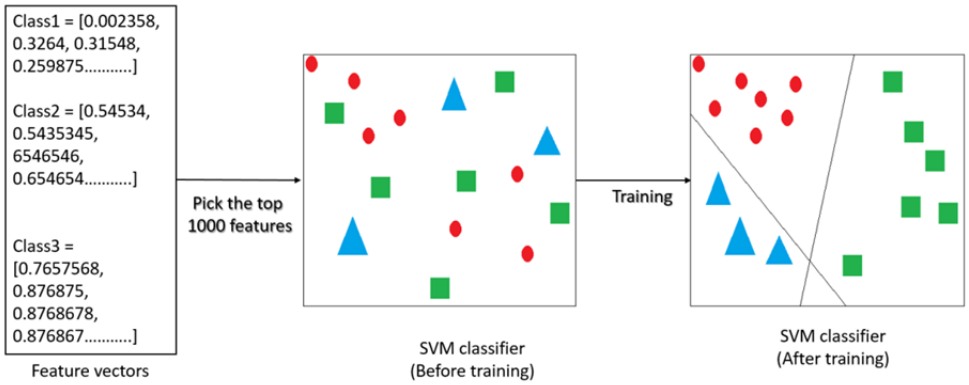


Figure 3.11. The steps to train the classifier.

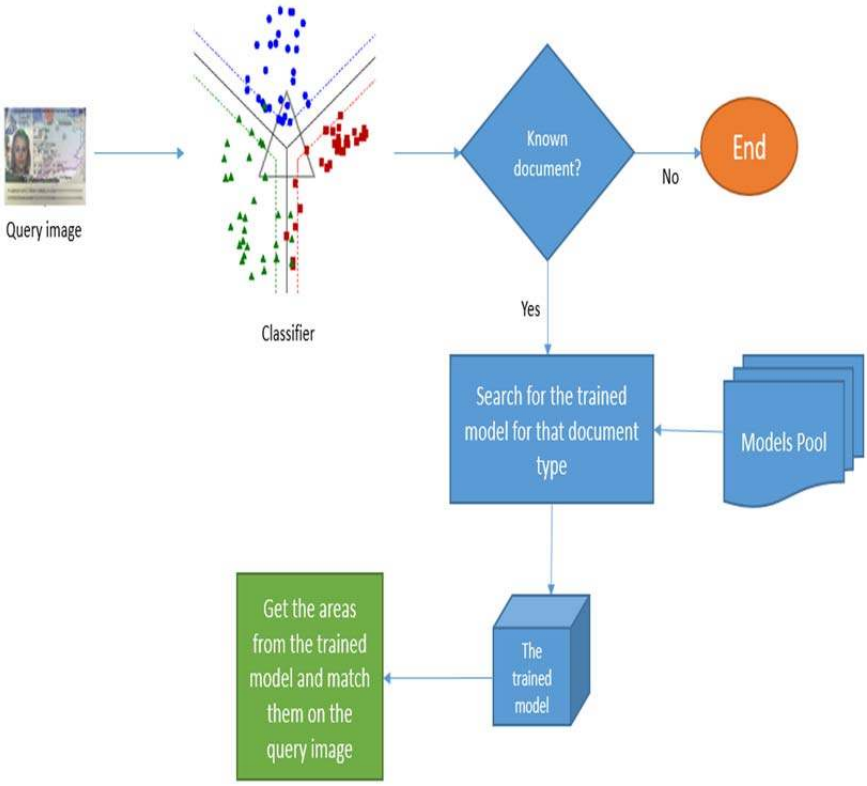


Figure 3.12. The steps on a query image.

### 3.6. Rotation

The document scanning systems usually scans around 11% of the images upside down.

The used feature extractor is transformation invariant, the feature extraction and the SVM classifier are not affected by the rotation of the image. However, the template matching is rotation variant approach. Hence, system does not work properly. This problem is solved by generating two templates for each typical image, the first one is the normal template and the second one is rotated 180 degree (upside down). If the rotated template matches, then the system calculates the coordinations for all areas in the documents and rotates them areas 180 degree

to match with the rotated query image. Figure 3.13. shows one of the results after applying the rotation template in the trained model.

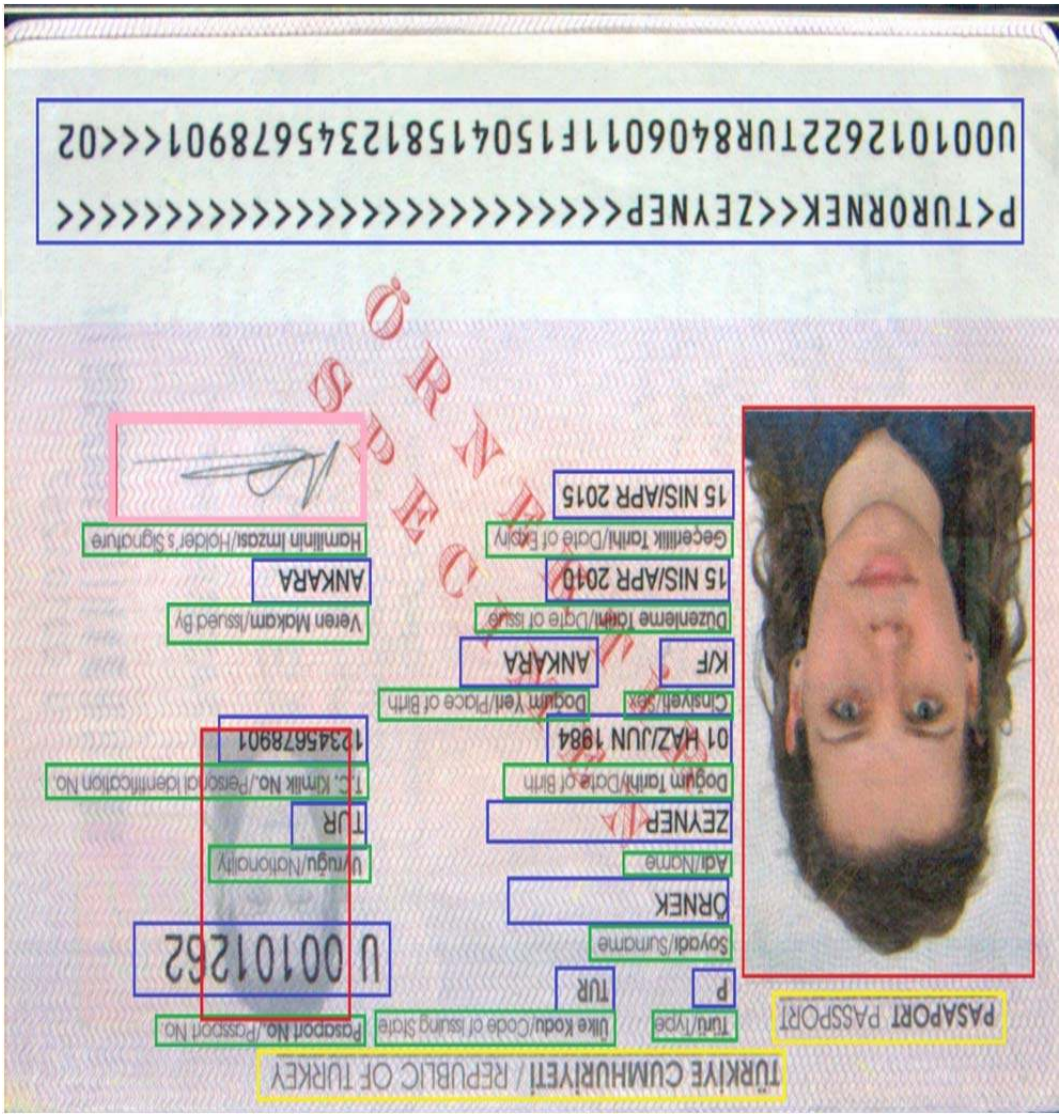


Figure 3.13. One of the results that after applying the rotation template in the trained model

### 3.7. The confidence of the template matching

The template matching is a method that checks how much the template *(T)* with size *(w, h)* and the query image are similar to each other. To find the template *(T)* in the query image *(I)* it will slide the window with the size of template T on the query image, then it compares the overlapped patches of size *(w, h)* in the query image against template *(T)*. The percentage of the overlapping between the patch with size *(w, h)* and the template *(T)* is called the confidence.

The low confidence matching can lead to some false positive results as given in Figure 3.14.

Figure 3.14. wrong detecting - false positive area detection

### 3.8. MRZ Processing

Some identity documents such as passports include MRZ at the bottom of consisting of only letters from A to Z, numbers from 0 to 9 and the filler character ‘<’.

The MRZ in passports usually have two lines with 44 characters in each. The first letter represents the document type and 'P' stands for passport, 'I' for ID and 'V' for Visa.

The first name and last name of the person usually exist in the first line of the MRZ in passports, and there will be two fillers '<<' to split between the first name and last name.

The second line usually contains the passport number (The first 9 characters of the second line), birth date, expiry date and the gender of the person.

The birth date and expiry date usually exist in YYMMDD format, so 190101 refers to 01/Jan/2019.

M or F in the second line of MRZ in passports usually refer to Male or Female.

Most of the MRZ information are duplicated in the VIZ area as well, and help cross match between the MRZ information and the VIZ area information.

In non-English passports, the transliterations are necessary since only letters from A to Z are allowed in the machine-readable zone. However, all characters are allowed in the VIZ area, hence there may be a mismatch between MRZ and VIZ areas. For these cases, the transliterations are used. Some of the letters and their transliterations are given below.

| | | |
|---|---|---|
| å → AA | lj → IJ | ß → SS |
| ä, æ→ AE | ñ → NXX | þ → TH |
| ð → DH | ø, œ, ö→ OE | ü → UE (German) |

Cyrillic and Arabic names are needed to be transliterated into their Latin versions. Hence, another solution is proposed to analyze and process the ICAO 9303 standard compliant identity documents (passports, identification cards and visas) in a general manner based on the specification offered by the ICAO 9303 standard and some other properties that can be deduced indirectly out of it.

Proposed solution recognizes type of document and issuing country, extracts the bio-information and document information from any image of an ICAO-compliant identity document with high quality.

A method of analyzing the identity document comprises acquiring at least one image of the identity document, analyzing the image to detect MRZ, processing the MRZ to extract information, detecting a VIZ area of the identity document, processing the VIZ to extract information, measuring matching score for the VIZ area and MRZ, and detecting the other information from VIZ that do not exist in the MRZ.

MRZ detection algorithm searches for at least one MRZ filler character "<" to find line of MRZ by using template matching trained on filler shape "<". After detecting at least one filler in the image, the system tries to detect another filler near to it by sliding the same template up and down. After detecting at least one filler from each line of the MRZ, a text detector is applied to the image to detect all the text that aligned with the location of the detected filler. Figure 3.15 and Figure 3.16. show the steps of MRZ detection and the result image, respectively.

After detecting the MRZ region, the system extracts all the information from the MRZ area and reads the MRZ text from the detected MRZ region by applying OCR. Figure 3.17. shows MRZ and text detection steps.

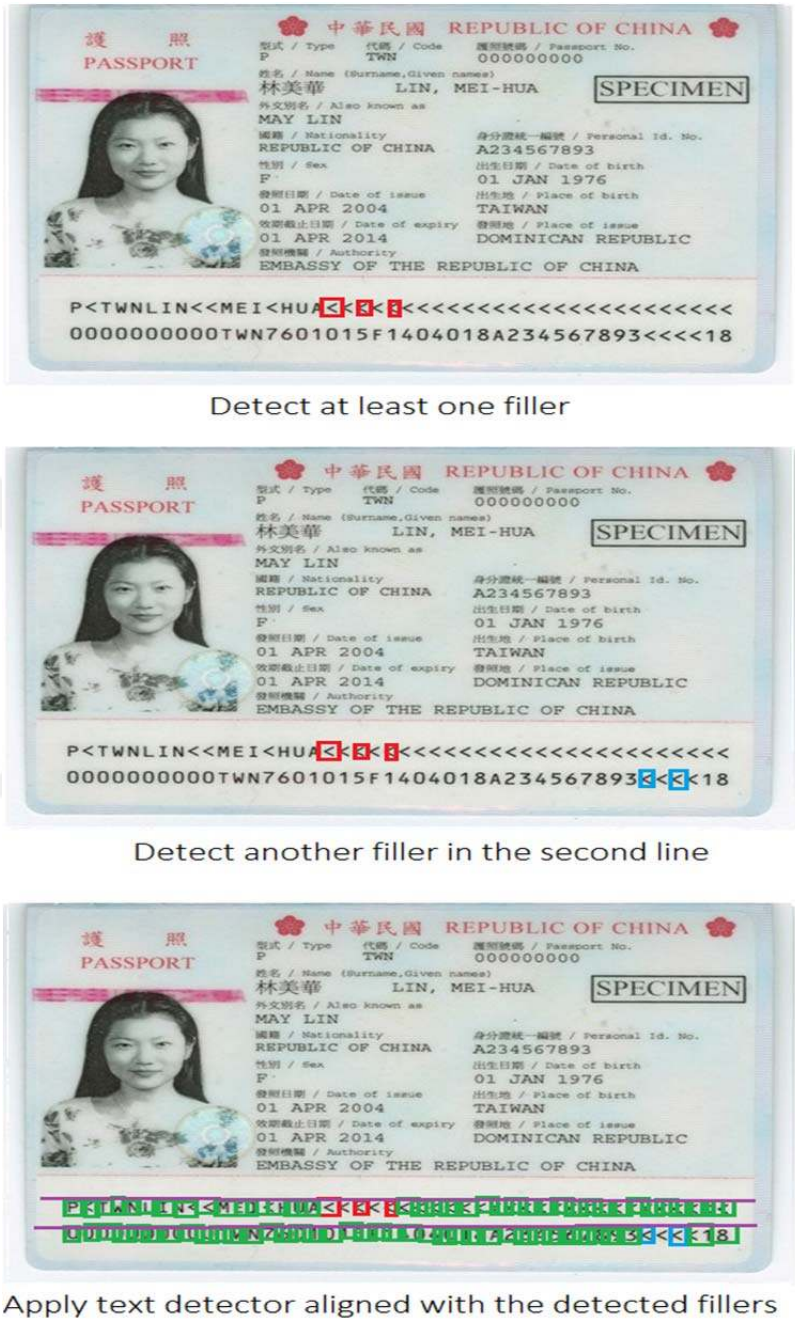


Figure 3.15. MRZ detector steps.

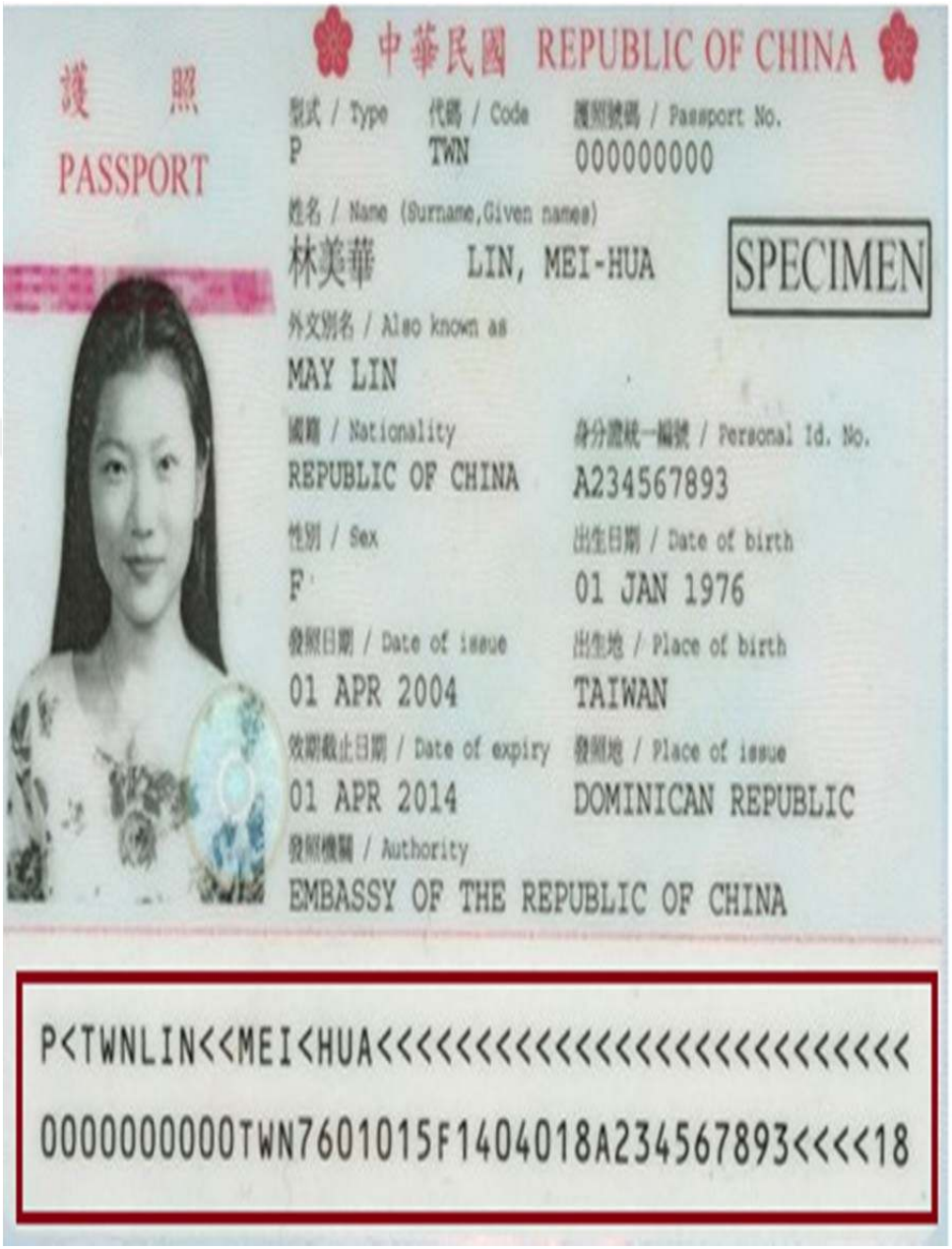


Figure 3.16. MRZ detector result.

Reading the MRZ helps to recognize the type of the passport, the issued country, first name, last name, date of birth, expiry date, gender and the passport number. Different rules are required to be applied on MRZ with two lines and MRZ with three lines. Rules have been applied to parse the MRZ text information to fields information such as first name, last name, passport number…etc shown in Figure 3.18. Figure 3.19 shows rules for detecting MRZ with three lines.

Parsing and analyzing the MRZ give all the information existing in the MRZ. According to the MRZ location, VIZ area's location of the document can be detected depending on the ICAO standard (Malcolm, 2010). Figure 3.20. shows ICAO layout.

There are two types of information in the VIZ area that are also existing on the MRZ region and VIZ area specific data. Text detection is applied on VIZ area, filtering is applied on detected text to eliminate labels and non-text regions. At last, OCR is applied on the detected VIZ area.

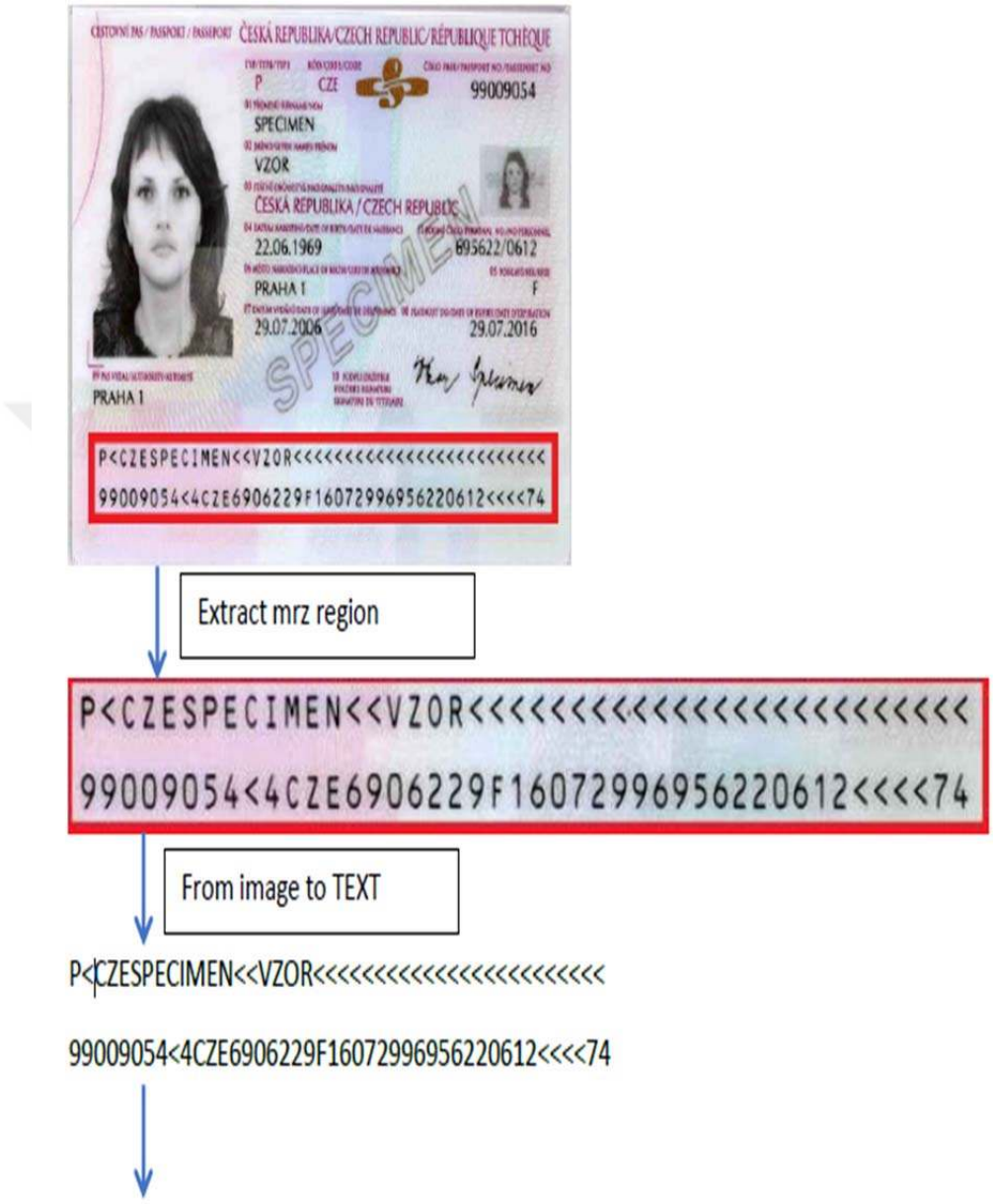


Figure 3.17. MRZ detection, MRZ extraction and OCR steps.

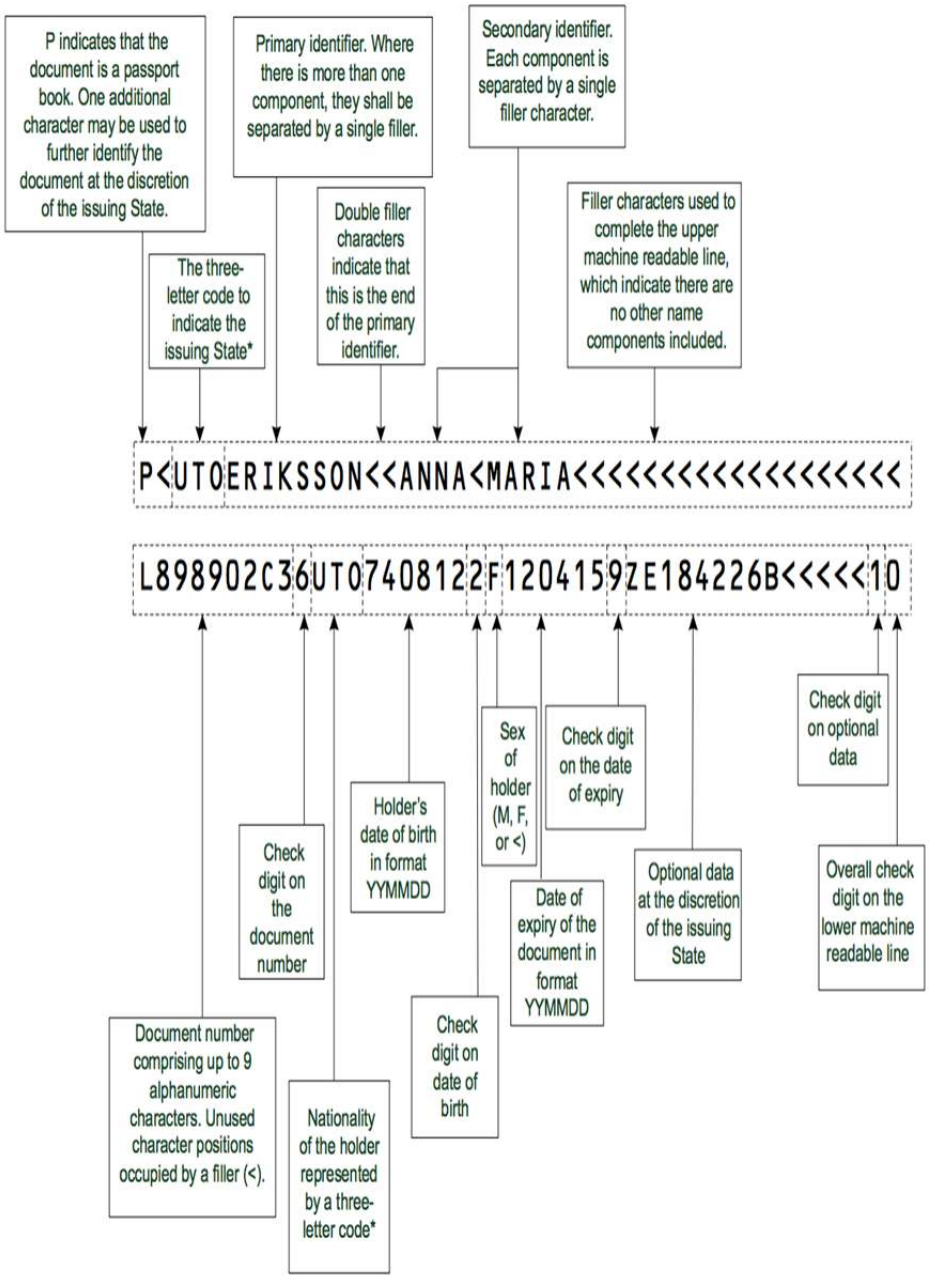


Figure 3.18. MRZ two lines format.

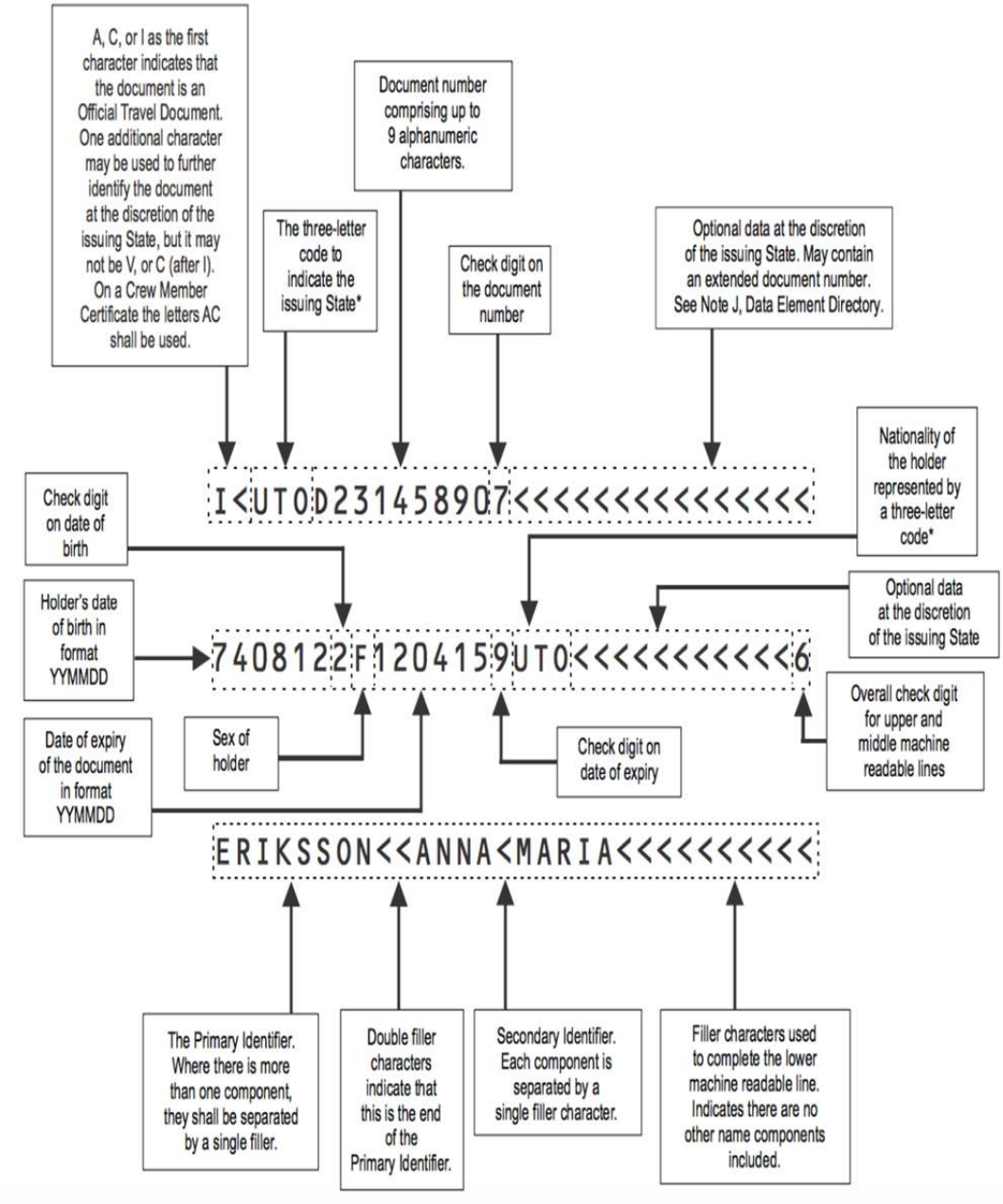


Figure 3.19. MRZ three lines format.

The information existing on both VIZ area and MRZ are used to label the detected text in the VIZ area. Figure 3.21. shows an example case. The remaining detected text in the VIZ area which did not match the information from the MRZ are mostly one of these options in the ICAO 93003 standard (Issue Date, Issuing authority, or Birth Place). After applying OCR for all detected text in the VIZ area, the dates are detected by using pretrained NLP module. Unlike MRZ, VIZ area includes issue date which is one of the required information. Since other dates are detected in MRZ, the remaining date is labeled as issue date. In addition to using ICAO layout matching, NLP model trained on city names and issuing authority may be applied for detecting birth place and issuing authority. Finally, last element to be detected on the document is face of the person.

The face detector uses Histogram of Oriented Gradients (HoG) feature-based cascade to train the classifiers (SVM). This classifier training uses many positive images (with faces) and negative images (without faces). Face on the query image is detected by sliding a window across the image. In each window, it computes the HOG then checks the classifier results to detect the face.

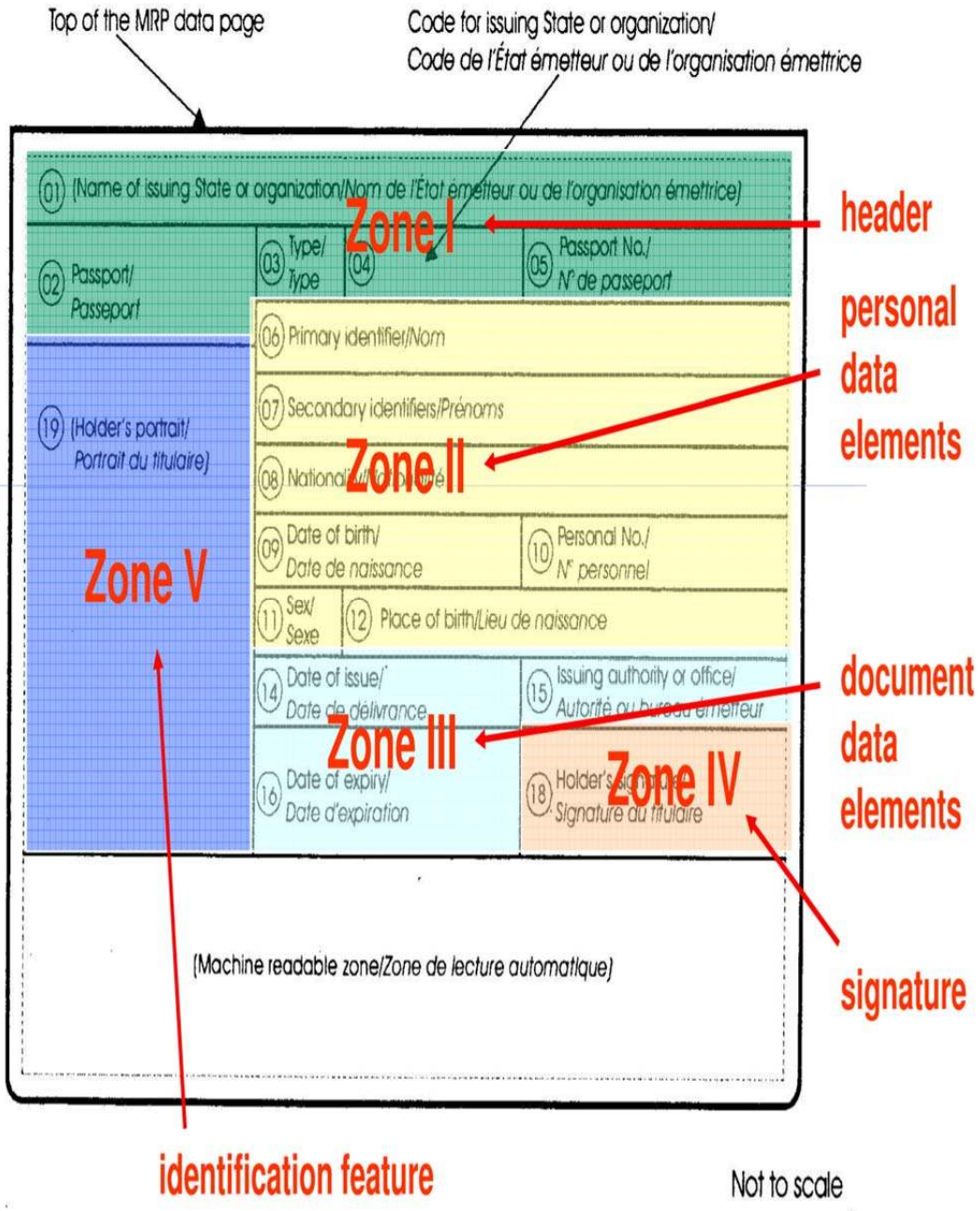


Figure 3.20. The expected layout on ICAO document with two lines of MRZ like passports.

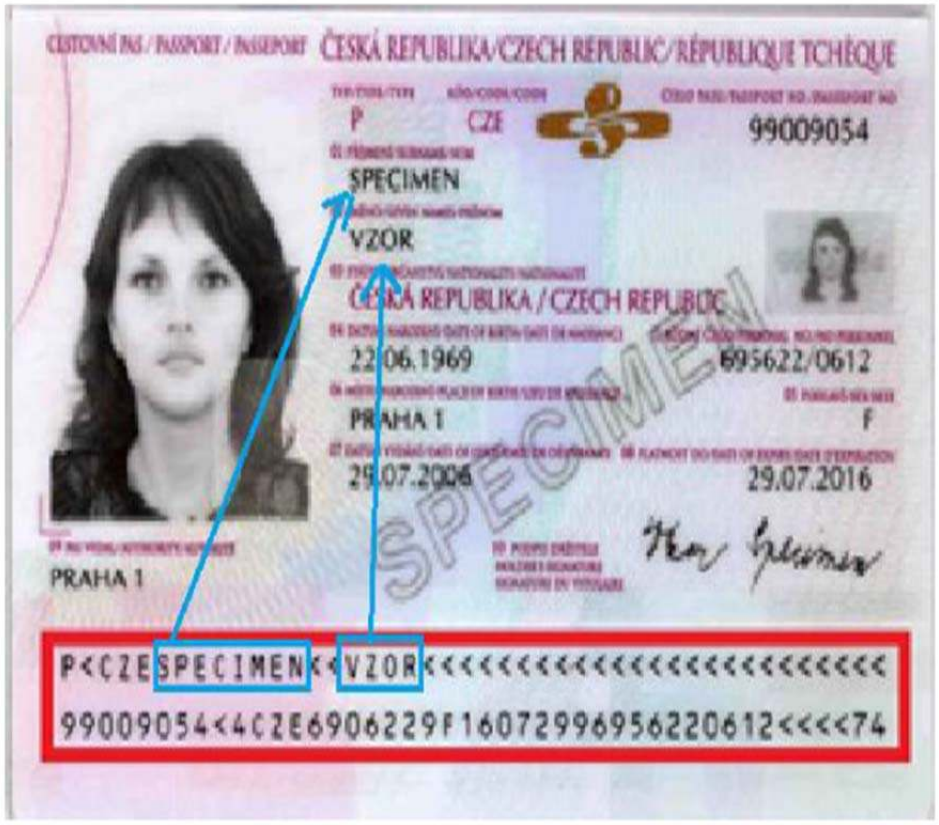


Figure 3.21. First name and last name information exist in both MRZ area and VIZ area as well.

Although MRZ processing results provide good understanding of the document, there exist other information to be detected in t in the ICAO document.

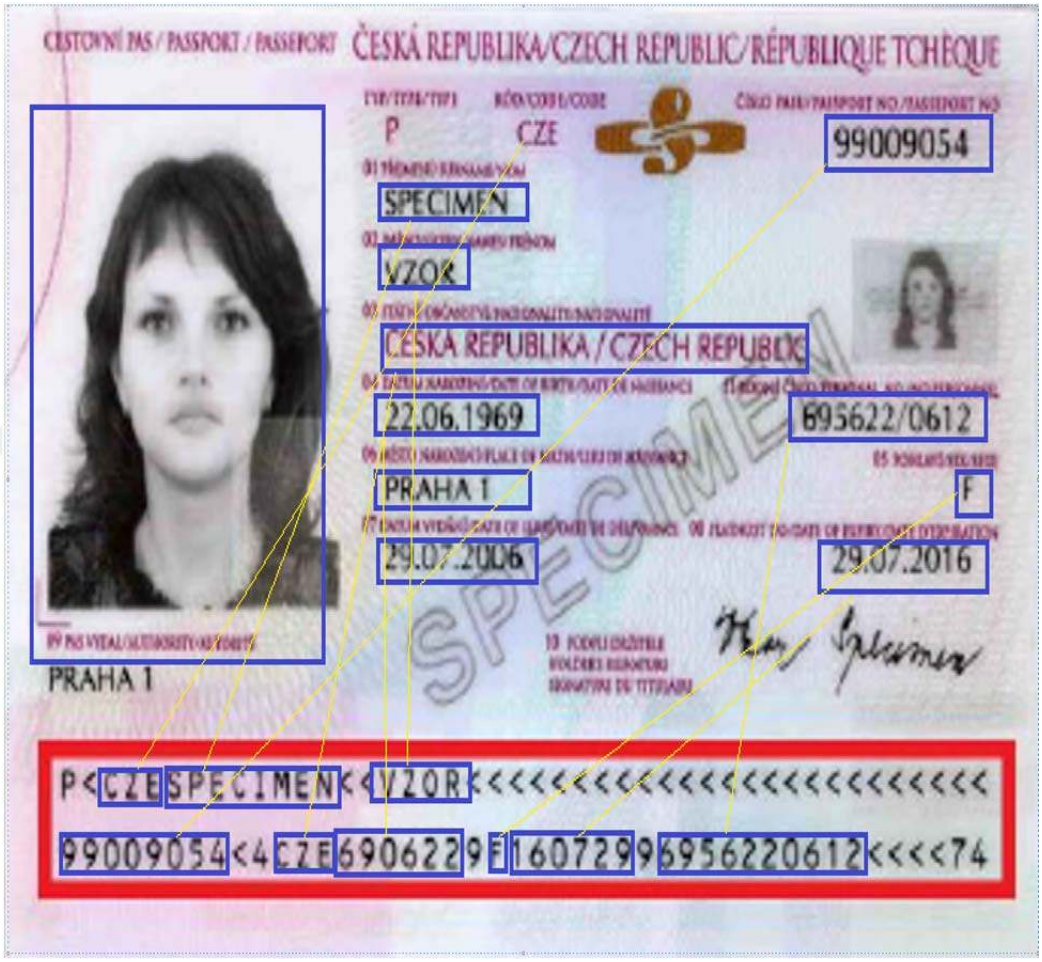


Figure 3.22. The matching between MRZ area and VIZ area.

# 4. RESULTS AND DISCUSSIONS

## 4.1. The testing results

Tests were performed on computer consisting of Intel core i7 2.60 GHz CPU, 32 GB of RAM, and Nvidia GTX1060. In the first experiment, CTPN, text detection, and OCR steps were tested and the average processing time was 11 seconds, and the average processing time for the OCR step was 5 seconds. Hence, the total average processing time was around ~15 - ~16 seconds with these steps.

The second experiment was done to train the CNN predefined structure using the dlib library, and the training set was over 100 of United Kingdom passport 2015 and the input was all segments and text in the United Kingdom passport 2015 like first name rectangles, last name rectangles, date of birth rectangle, and all other textual information in that document type. After training the model and use of more than 50 images for United Kingdom passport as test set, the test results detected 80% of the segments correctly, and the model need to increase number of samples used in training. After increasing the number of samples in the training set from 100 to over 200 the test results enhanced from 80% to 85%.

Using CNN deep network can help to provide good and high accuracy aligned with the objectives, however it needs more than 200 samples from each identity document type. Hence, more than 200 different images from each document type which is 1000*200=200,000 samples are required to be able to train CNN with good accuracy. Analyzing the identity documents consisting of over 1000 types requires over 250K training images which is not easy to get according to the new General Data Protection Regulation law.

Initial approach using SVM classifier was tested on trained model with over 800 different identity document types from all over the world. The test set contained 2 to 5 samples from each document type with 1585 sample image for testing which was different to the samples chosen for training the engines. Figure

4.1. shows one of the results that are getting using this approach using defined model for Turkish passport.

The accuracy of the this proposed structure was 82%. It detected correctly with all the expected areas from each document type, while 18% of the sample images were not classified or system could not detect any area on the sample image.

After analyzing the first test results, it shows that around 11% of the samples were rotated upside down. With the roation method and after applying the rotated template the accuracy was 93%, where the accuracy without roation method was 82%.

After analyzing the test results of the prevoius methods, it seems that the classifier was able to classify more than 96% of the test images, where only 93% of them was correctly detected. One of the reason behind the failures is the template matching. It had some limitations like rotation which has been solved, and by increasing the matching threshold more than 70%, number of matched samples was reduced 9% less, and by decreasing the matching threshold less than 40% the false positive matches increased 5%. The first threshold has been chosen as 50% and 93% of the test samples were matched with 0% false positive matches. Since matching threhold affects accuracy, threshold searching was applied to balance the percentage of the samples matched and the false positive cases. Incorrect detection tolerance value was defined as 0.5% and the experiments show that optimal value for the matching threshold is 45% for the tested data. This threshold value increased the accuracy of the correct matches from 93% to 94.8; however, number of false positive cases were also increased to 0.2%.

The accuracy of the template matched approach was 94.8% of the test samples, and by analyzing the remaining 5.2% failure cases, it was found that 3.1% of the test samples are passports or identity documents with two lines of MRZ or three lines of MRZ.

The accuracy for the overall pipeline including both template match and MRZ processing provided 3% increase on the number of correctly matched samples, however there were still 0.2% of the samples detected incorrectly and 2% of the samples were not detected at all. Addition of the MRZ processing increased the overall accuracy to 97.8%.

The results of the final approach which uses two main methods, document models and MRZ processing, to analyse the identity document image show that proposed system was able to detect the textual information from the the identity documents image, label the detected text to (first name, last name, document number,…etc.), and also detect the graphical shapes in the documents like face, signature and logos.

Using the 1858 test samples collected from over 800 different document images under different lighting conditions and different orientations and covered the most important documents all over the world (more than 100 countries) provided 97.8% accuracy and 2.2% false positive undetected documents.

In the first experiment with the identity module the initial results were 1534 sample images out of 1858 of the samples were detected correctly which is around 82%. However, 324 of the samples were not detected at all which is the remaining 18% of the samples.

After applying the rotation method, the results showed that 1728 sample images were detected correctly out of 1858 of the samples which is around 93%. However, there exist undetected 130 of the samples which is the remaining 7% of the samples.

By decreasing the template matching threshold to 0.45, 1761 sample images were correctly detected out of the 1858 of the samples in addition to previously correctly classified 94.8%. However, 93 of the samples were not detected at all which is 5% of the samples and 4 sample regions was detected in the wrong location (false positive) 0.2%.

With the MRZ processing, another 56 samples were detected correctly in addition to the previous 1761 which makes 1817 correctly detected samples and the overall accuracy of this research increased to 97.79%. Table 4.1. shows the results of experiments and the overall results of the research. Figure 4.1. shows the experiment results.

Table 4.1. The test results for each method and overall test results.

| | **Identity module** | **MRZ module** | **Low confidence** | **Wrong detection** | **Not detected** |
|---|---|---|---|---|---|
| **First test** | 82% | 0% | 0% | 0% | 18% |
| **With rotation** | 93% | 0% | 0% | 0% | 7% |
| **Decreasing the threshold** | 93% | 0% | 1.8% | 0.2% | 5% |
| **After ICAO** | 93% | 3% | 1.8% | 0.2% | 2% |
| **Final results** | **97.8%** | | | **2.2%** | |

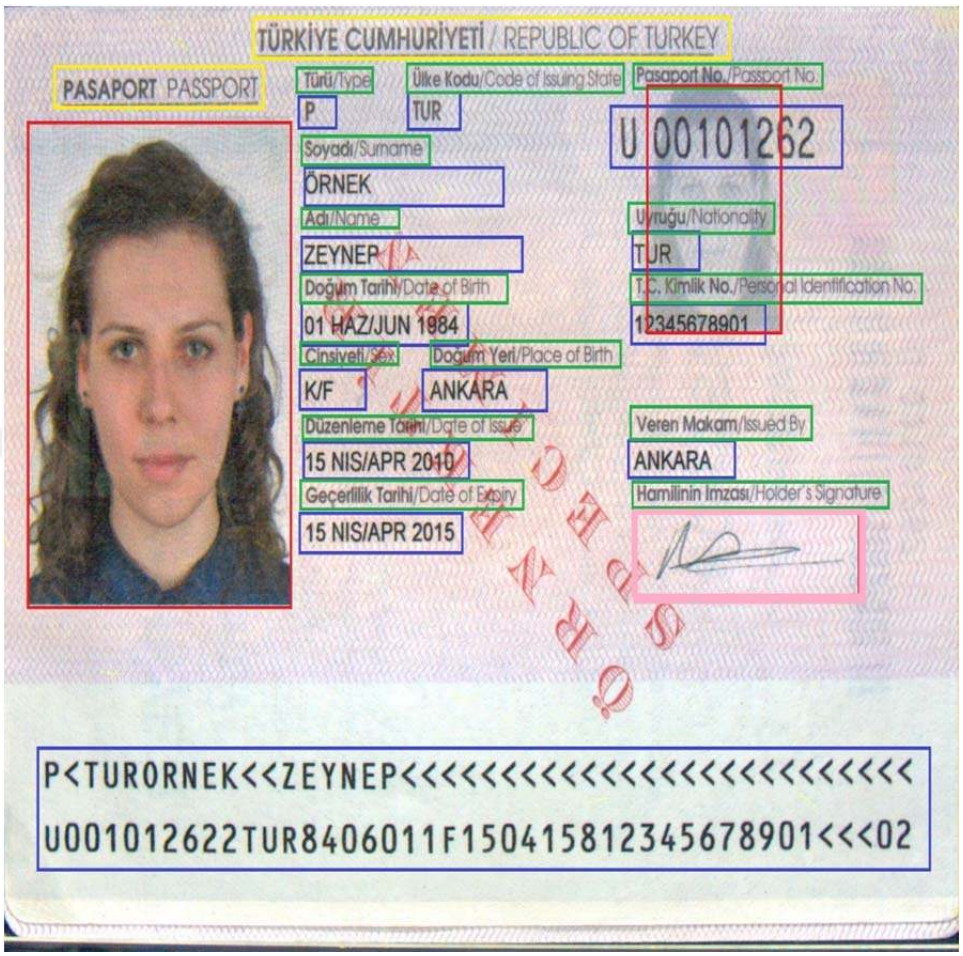


Figure 4.1. The results on Turkish passport query image using defined model for Turkish passport.

The processing time of this approach was tested on Intel Core i7 2.60 GHz, 32 GB. of RAM, and Nvidia GTX1060 and the performance time was given in Table 4.2.

Table 4.2. The processing time.

| | Identity module | MRZ module | Not detected | Total processing time |
|---|---|---|---|---|
| **Min time** | 0.2 Second | 0.5 Second | 0.6 second | 0.5 second |
| **Max time** | 3 Seconds | 7 Seconds | 6 Seconds | 7 Seconds |
| **Average time** | 0.5 Second | 3 Seconds | 2 Seconds | 0.6 Second |

### 4.2. The Benefits

This study helps to analyze the identity document image, with very high accuracy. Locating and analysis of the layout of the identity document save time and efforts. Moreover, proposed study reduces human errors and the OCR systems..

The MRZ processing also provides a generic way to analyze any kind of ICAO documents containing MRZ, whether it's a passport, visa or identity document with three lines of MRZ - mostly used in the back side of national identification cards around the world.

### 4.3. The limitations

The previous methods follow a structured document analysis approach where each specific identity document is predefined and trained in the system with information about the location and size of all elements. Although this approach provides accurate information analysis and retrieval capabilities, it has the shortcoming of the need to train every specific issue (layout) of any candidate document and when new issues of these documents are published, they will not be supported unless the system is updated with information about them.

The MRZ processing help partially to solve the previous issue if the identity document type contains MRZ, but it will work with other types of documents which can be conducted in future studies.

# 5. CONCLUSION AND FUTURE WORK

Using a trained model for each identity document type and the classifier model classify an identity document type which helps to analyze and localize the information in the identity document type with very high accuracy. Moreover, experiements showed that these results may be enhanced by using some methods such as a rotation method used to detect if the identity document image is rotated or not.

The MRZ processing provides a generic way to analyze any kind of ICAO document with two lines or three lines of MRZ. The MRZ processing detects the MRZ area and extracts all the information from the MRZ. While using similarities between the VIZ area and and MRZ area provide many required information, other parts are obtained by using the NLP engine and analysis of the ICAO layout. The MRZ processing provides good accuracy for the location of the textual information in the document type compliant with the ICAO 9303.

The accuracy of the research was 94.8% with accurate detection for all segments in the document using identity trained modules, and 3% with good detection using the MRZ module.

Some of limitation of the previous methods which can be conducted for future studies has the shortcoming of the need to train every specific issue (layout) of any candidate document and when new issues of these documents are published, the system needs to be retrained again to support that document. The MRZ processing will be able to help on this, but it will not work on documents that do not have an MRZ.

**REFERENCES**


Beattie, Charles, 2016. DeepMind Lab. arXiv:1612.03801v2.

Cesarini, F., 2000. Structured Document Segmentation and Representation by the Modified X-Y tree. Universit`a di Firenze

David G. Lowe, 1999. Object Recognition from Local Scale-Invariant Features, Computer Science Department University of British Columbia Vancouver, B.C., V6T 1Z4, Canada.

David G. Lowe, 2004. Distinctive Image Features from Scale-Invariant Keypoints, University of British Columbia Vancouver, B.C., Canada

Girshick, Ross, 2018. Facebook open sources Detectron. https://research.fb.com/facebook-open-sources-detectron.

Gordo, Albert, 2016. Scene-Text Localization, Recognition, and Understanding. DAS 2016.

Gori, Marco, 2003. Artificial Neural Networks for Document Analysis and Recognition. University of Florence.

https://docs.opencv.org/2.4/doc/tutorials/imgproc/histograms/template_matching/template_matching.html

https://www.icao.int/Meetings/AMC/MRTD-SEMINAR-2010-AFRICA/Documentation/5_CuthbertsonChalmers%20_MRTD-eMRTD.pdf

ICAO, 2015. Machine Readable Travel Documents Part 3: Specifications Common to all MRTDs Seventh Edition,

Jaderberg, Max, 2015. DEEP STRUCTURED OUTPUT LEARNING FOR UNCONSTRAINED TEXT RECOGNITION. ICLR 2015.

Malcolm Cuthbertson, 2010. ICAO MRTD and eMRTD Standards, ICAO Regional seminar, ISO.

Marinai, Simone, 2008. Introduction to Document Analysis and Recognition. University of Florence

Nagy, George, 2000. Twenty years of document image analysis in PAMI. IEEE.

Rodolfo Valiente, Marcelo T. Sadaike, 2016. A process for text recognition of generic identification documents over cloud computing, Laboratory of Computer Architecture and Networks, Escola Politénica da Universidade de São Paulo, São Paulo, SP, Brazil,

Strouthopoulos, C., 1997. Text identification for document image analysis using a neural network. Elsevier Science.

Young-Bin Kwon, and Jeong-Hoon Kim, 2005. Recognition based Verification for the Machine Readable Travel Documents, Departmen Computer Engineering, Chung-Ang University Seoul, Korea,

Zhi Tian, Weilin Huang, Tong He, Pan He, Yu Qiao, 2016. Detecting Text in Natural Image with Connectionist Text Proposal Network, cornell university

**CURRICULUM VITAE**

Hazem Abdullah was born in 1988. He has completed university education at department of Informatic Engineering at Aleppo university in 2011. Since 2011, he has been working as an A.I. research and development engineer in multiple lab researches.