Validation of Legal Documents
by comparing digital and printed photographs

Abstract - This paper discusses the comparison of the photos on legal documents. The documents have a chip photo and a printed photo. Those two photos should be the same. An algorithm is designed to compare whether the chip photo and the printed photo originate the same. Direct comparison is impossible, because a photo is made of the whole document instead of only a photo of the printed photo. Research was done to compare different methods for aligning the document photo with the chip photo and the comparison of the photos. Research shows that SURF feature detection performed best for aligning the document photo with the chip photo and to check whether the two photos are the same. Binary images performed best to compare the photos without being bothered by differences in brightness, contrast and sharpness. An experiment is designed to check how well both parts of the algorithm perform separately. Another experiment is designed to check how well the overall algorithm performs. All experiments are based on comparing the chip photo with the document photo of the same photo, same person but different photos (for instance passport and drivers license) and different persons. The results show that overall the algorithm works sufficient. Only alignment based on SURF is not performing sufficient if the document photo is taken at an angle or when the same chip photo and document photo of residence permits are compared. The printed photos on the document photo of residence permits are invisible, because of a glare caused by the angle of incidence of light.

I. Introduction

In the Netherlands, everyone is obliged to carry proof of identification. This can be a passport, an ID card, a drivers license or a residence permit. Each document has a chip photo and a printed photo. Sometimes the printed photo is changed. This way it looks like it is a legal document, but actually it is a fraud. To check if the document is legal, the chip photo and the printed photo on the document need to be compared. The photos should have the same source. If the two photos show the same subject, but do not share the same source, the algorithm must detect that it is different. Also if the photos are simply different, such as with two different people, the algorithm must detect that there is a difference.
The comparison starts by reading the chip photo and photographing the document photo. First of all there is a problem with the alignment. A photo is taken from the document and not only from the printed photo on the document, so the photos should first be aligned after which the photos can be compared. The difficulty at the comparison is that the chip photo and the document photo do not have the same quality and properties. The photos differ in brightness, contrast and sharpness. The goal of this research is to compare and evaluate different existing methods to design an algorithm that can detect whether the chip photo and the document photo originate the same. In Section II., the methods that can be used in the algorithm are discussed. In Section III., the methods chosen per part are described. This yields the following research questions. Research question one, does the algorithm work sufficient for aligning the document photo with the chip photo based on SURF feature detection? With sub question, should the photo of the document been taken from straight above or does photographing at an angle still give a sufficient result for alignment? Research question two, does the algorithm work sufficient to use binary images to determine whether the chip photo and the document photo have the same source? With sub question, what is the best set threshold value for the chip photo? Experiments must be done to answer those research questions.

The set up for the experiments are described in Section IV. Section V. will show the results from the experiments and the results will be discussed in Section VI. The conclusion is given in Section VII.

II. Related work

A method that is based on the comparison of two photos with the same source is reverse image search. Reverse image search will find the same image on the web. This image can be cropped, transformed or the lighting can be changed.[1]
The reverse image search can be done if the image is almost identical. In case of comparing the chip photo with the document photo, there is an extra difficulty, because first alignment needs to be done before the comparison can start. This means that when comparing the chip photo with the document photo, there are two difficulties, alignment and then the comparison. Alignment because a photo is taken of the entire document instead of just the printed photo on the document. Comparison, because there is a difference in brightness, contrast and sharpness. Research needs to be done to compare different methods for aligning the document photo with the chip photo and the comparison of the photos.

A. Alignment

For aligning the photos, there are multiple solutions. The two methods that will be discussed are facial landmark detection and SURF feature detection.

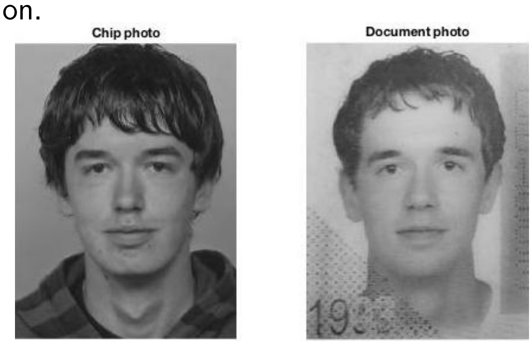
A.1. Facial landmarks

Facial landmark detection focuses first on detecting the face and facial regions. When the facial regions are found, facial landmarks should be extracted. Facial landmark detection is designed such that facial regions are always detected. It does not matter if the face is turned a bit left or right.[2] This means that if the same person is on the photo, the facial landmark detector gives enough matching landmarks to align the document photo to the chip photo, as in Figure 1.

For this assignment, the photos should be the same and not only the person on the photo. This means that, at for instance twins or the same person with a different photo on the document, the facial landmark detection would not detect a difference in the photo.

A.2. SURF

SURF (Speeded Up Robust Features) is often used for image matching. Images have a unique set of features which are exclusive per image, so if an image is identical or partly identical, identical features can be detected. By comparing different image matching techniques, SURF is most accurate, fastest, scale invariant and rotation invariant.[3] Because in this assignment, the images should be identical, SURF will be investigated as a possible method. SURF will be used to align and detect differences in the photos.
The SURF algorithm consists of detection, a descriptor and matching. SURF is partly inspired by SIFT, only SURF is faster and more robust for image transformations such as scale and rotation.



(a) Twins

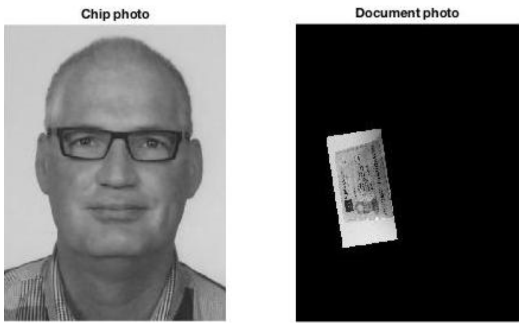
(b) Same person, different photos

Figure 1: Alignment using facial landmarks

At the detection part, SURF detects features in images. Those points of interest are found using a blob detector based on the Hessian matrix. A blob detector is aimed at detecting areas with

different characteristics compared to surrounding areas.[4] To measure the local change around the interest point, the determinant of the matrix is used and the points where the determinant is maximal are chosen. [5]
The descriptor provides a robust description of a point of interest. For instance this is done by describing the distribution of intensity of the pixels nearby. The descriptor is scale invariant and rotation invariant. The scale invariance is handled by sampling the descriptor over the window in proportion to the window size at which the point was detected. To handle the rotation, the dominant direction of the point of interest must be found. Next, the sampling window must be rotated to align with that angle. [6]
The matching part of the SURF algorithm compares the descriptors from the images in comparison. If the descriptors match, which means that the points of interest and the pixels nearby are the same, matching pairs can be found.[5]

(a) Twins



(b) Same person, different photos

Figure 2: Alignment using SURF

In this case the photos should be the same, so the points of interest and the neighbouring pixels in the photos should be the same. When using another photo of the same person or twins, there should be no match between the descriptors which results in not being able to properly align photos. This can be seen in Figure 2.

A.3. Conclusion

Because the alignment will also be used to check whether the photos are the same, SURF feature detection will be used for aligning the photos in the total algorithm.

B. Photo comparison

When the alignment is finished, the chip photo and the document photo need to be compared. To compare the images, they should have the same quality and properties which means that there should be no difference in brightness, contrast and sharpness. Three methods were investigated to get comparable images. Edge detection, histogram equalization and the use of binary images.

B.1. Edge detection

Edge detection identifies points in an image where the brightness changes sharply or has discontinuities. Those points are merged into a set of curved line segments, referred to as edges.[7]
In edge detection, there are several methods. Sobel edge detection is the one that is examined here. Prewitt and Sobel are both providing an efficient good edge mapping, but Prewitt is sensitive to noise, which makes it less suitable for this research.[8] Sobel processes the image in the horizontal direction and vertical direction. Two kernels of a 3x3 matrix, one for each direction, are used as filters for detecting the edges. This is done by scanning for large changes in the gradient.[9]
The chip photo and the document photo have different edges in the photos, because in the document photo there are some marks visible in the photo. This causes some extra edges, new discontinuities, or removes an edge, because there is no discontinuity anymore. Furthermore the alignment should work well, because the detected edges are thin lines as in Figure 3. It will be impossible to compare the photos if there is a mismatch in aligning the photos.

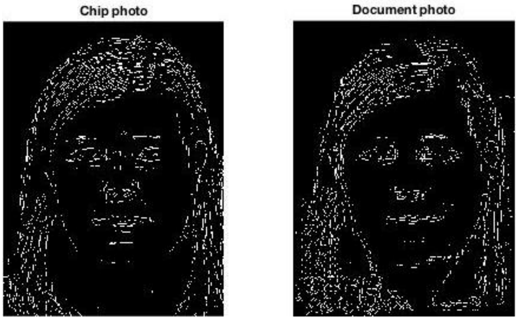


Figure 3: Same photo with edge detection

B.2. Histogram equalization

Histogram equalization is the processing technique to change the contrast in an image. A histogram is made of an image and by changing the histogram, the contrast of the image also changes. If the histograms of two images need to match, the contrast of one image will change to the other.[10]
Due to the markings on the printed documents, the histogram is not the same in every part of the image. This will result in too much difference, as can be seen in Figure 4. It will be difficicult to compare directly.

Figure 4: Same photo with histogram equalization

Figure 5: Same photo with histogram equalization per part

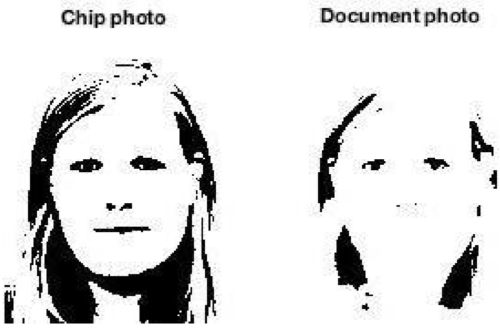
When the photo is cut into pieces and the histogram of a part of an image is matched to the same part of another image, the result for the equalization is better. Only the lines of the transition between each part is visible, as in Figure 5. This makes it still difficult to compare directly.

B.3. Binary images

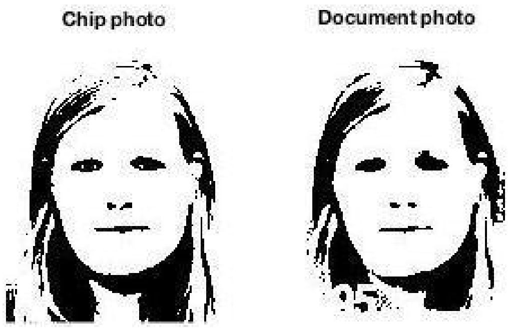
A binary image consists of two colors, usually black and white. Binary images are created using a threshold value. This threshold value sets a pixel on black if the intensity is lower than the threshold value or a pixel on white if the intensity is higher than the threshold value.
In case of comparing two images, the threshold can be set the same or separately. Due to the difference in intensity in the photos, separately seems to be better as can be seen in Figure 6.

B.4. Conclusion

Binary images will be used to compare the photos in the total algorithm. Compared to edge detection, binary images showed the least interference of the markings on the document photo and gave the best score by comparing the two processed images via Matlab. Binary images are also chosen over histogram equalization, because histogram equalization could not be chosen on its own.



(a) Same threshold



(b) Different threshold to get max score

Figure 6: Same photo as binary images

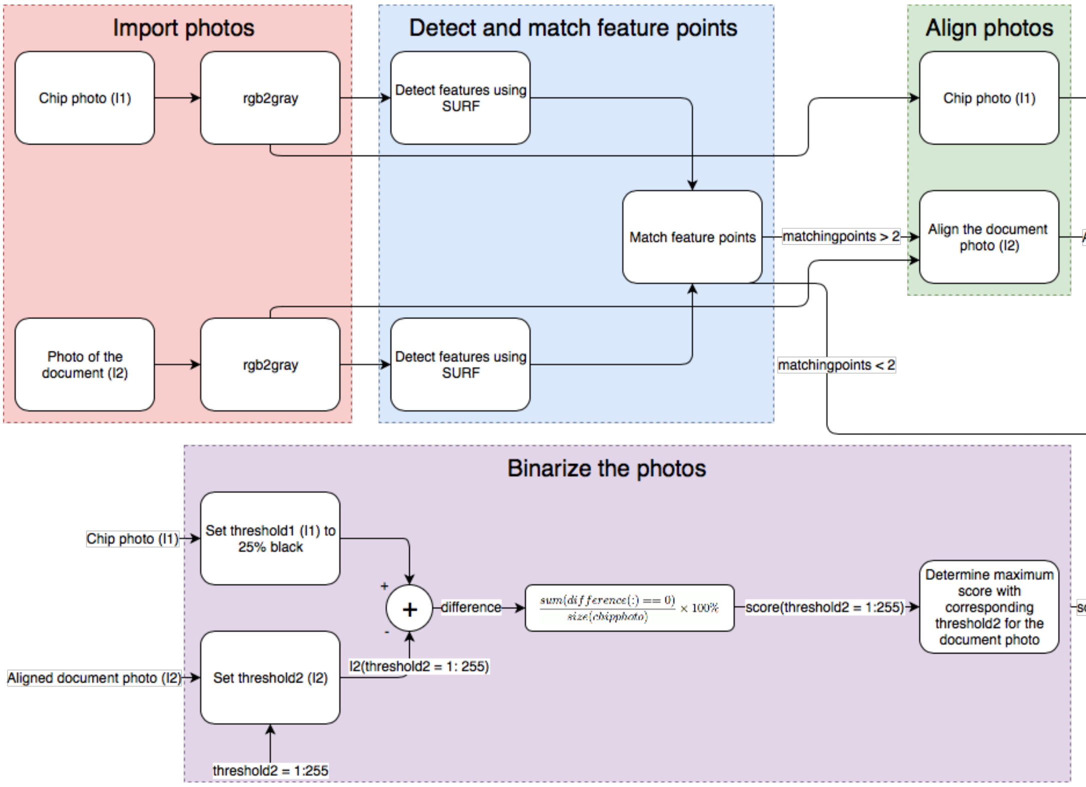


Figure 7: Total setup of the comparing algorithm

III. Method

After comparing the methods for the different steps in the algorithm, one method per step is chosen. In Figure 7, the total setup of the algorithm can be found. First the chip photo and the document photo are imported and then set to gray scale. These photos in gray scale are needed to use SURF. SURF will be used to align the photos and check whether the photos are the same. Alignment can only be done if there are at least two matching pairs, so when there are less then two matching pairs, the score will be set to 0%. After the photos are aligned, the photos are compared, this will be done using binary images. During the comparison a score will be calculate which will be displayed.

A. SURF feature detection

SURF is chosen for the alignment. It works properly for aligning the same photo. In case of two different photos, the alignment is wrong. This means that SURF is also detecting different photos, it does not matter who is on the photo. Different photos will be detected and misaligned by the SURF algorithm.

Before SURF can be used, the photos should be set to gray scale, because the patterns of edges are analysed. SURF starts with detecting the features in the photos. In Figure 8 all SURF points detected are shown.

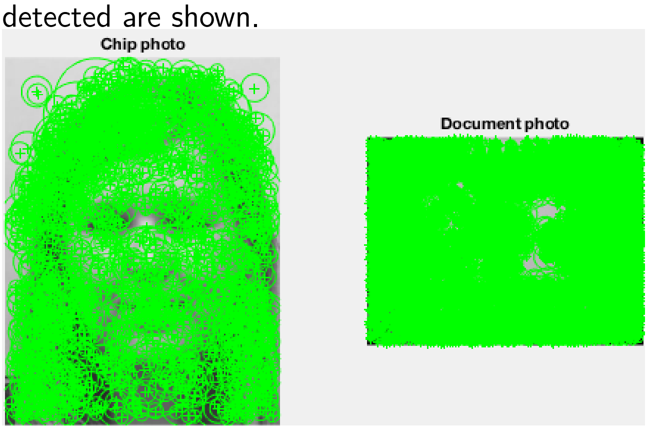
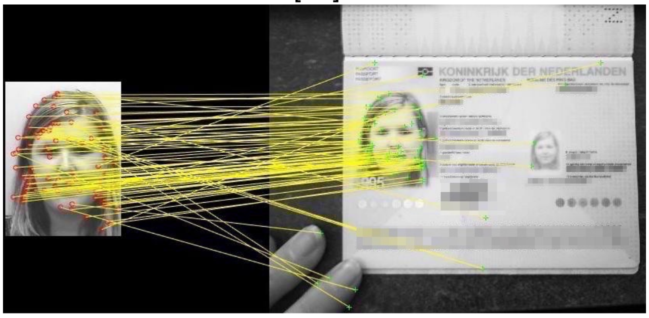


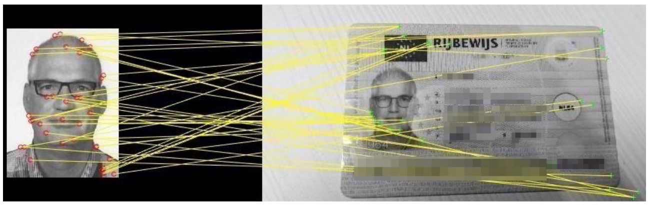
Figure 8: SURF points found per photo

After all SURF points are found, the corresponding points should match. Matlab is used to detect [11] and match [12] the SURF points. The matching points are shown in Figure 9. In Figure 9a, the photos are the same and the matching points ensure good alignment. In Figure 9b and 9c, the photos are different and the alignment will go wrong. The document photo will be rotated and scaled incorrectly to get all matching points in the same place.

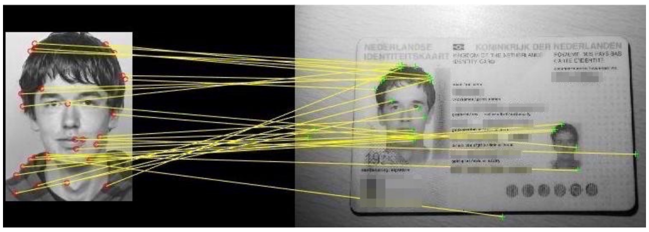
After the matching points are found, the alignment should be done. First the geometric transformation for the matching points should be estimated using a function in Matlab. [13] When the transformation is known, it should be applied to the document photo which is also done using a function in Matlab. [14]



(a) Matching points with the same photo



(b) Matching points with different photo of the same person



(c) Matching points with different photos of twins

Figure 9: Matching points using SURF

B. Binary Images

When the alignment is working properly, the images should be compared. The comparison is done using binary images. The binary images are made by using a set threshold value for the chip photo and a variable threshold value for the document photo. The set value for the chip photo should be a value that works sufficiently for every person on the photo. The value for the document photo will be defined after calculating the score in percentages as in Equation 1 and 2 for every possible threshold value. First, the difference between the chip photo and the document photo is calculated by subtracting the document photo of the chip photo. The number of zeros in the matrix of the difference is counted and then divided by the size of the chip photo. To get a percentage, it is multiplied by 100.

$$difference = chipphoto - documentphoto \tag{1}$$

$$score = \frac{sum(difference(:) == 0)}{size(chipphoto)} \times 100\% \tag{2}$$

The threshold value with the maximum score will be used for the document photo. The corresponding score is the result for the entire algorithm and will be displayed.

IV. Experiment

For the different parts in the algorithm, there are experiments designed. Also an experiment for the total algorithm is done. For the experiments per part, the documents of family and friends are used. For the experiment of the total algorithm, the documents of family, friends and InnoValor are used. In total, 51 chip photos and 69 document photos will be used. The various documents used are 24 passports, 21 ID-cards, 20 drivers licenses and 4 residence permits. The documents are from the Netherlands, China, Belgium, Norway, Finland, France, Spain, Deutschland, New Zealand and the United Kingdom of Britain and Northern Ireland. All experiments are based on comparing the chip photo with the document photo of the same photo, same person but different photos and different persons in order to check whether the algorithm or part of the algorithm works sufficient.

A. Alignment

For the alignment are two experiments designed. The first experiment examines how well the alignment works based on SURF feature detection. The second experiment examines whether it matters if the photo of the document is photographed with an angle.

To show how well the alignment works based on SURF feature detection, the aligned photos should be placed next to each other to check whether the characteristics of the face are on the

same height. This can be done using a ruler or by adding horizontal lines on top of the photos. If the characteristics are on the same line and the photos are the same, the alignment based on SURF works properly. If the characteristics are not on the same line, the alignment does not work properly.
To check if photographing at an angle still gives sufficient alignment, a photo taken from right above and a photo with an angle should be compared. This will also be done by checking if the characteristics of the face are on the same height using a ruler or by adding horizontal lines on top of the photos.

B. Photo comparison

There are also two experiments designed for the photo comparison. The first experiment examines how well the comparing works using binary images. The second experiment examines what value for the threshold of the chip photo performs best for comparing the photos.
The use of binary images will be tested using the score calculated in the algorithm. The scores of the same photo should be above 75% and the scores of different photos should be below 75%. The difference in the threshold values between the chip photo and document photo will also be looked at, to see if setting the threshold automatically for the document photo gives the maximum score. This can be done by plotting the threshold for the document photo against the score in percentage.
To know what value for the threshold of the chip photo is best to use, document photos will be used of people with different appearances. The different appearances that will be tested are, a person with white hair and a person with black-/dark hair. The values are percentages of the photo that will be set to black. The values tested are 20% black, 25% black, 30% black and 35% black. By comparing the outcome of the chip photos of those people, the standard value can be set. This standard value should work sufficient for every person.

C. Total algorithm

To test the total algorithm, a bigger test setup will be made. The procedure before the test can start will be as follows:

1. Take a photo from right above the document with a plain background.
2. Scan the document and read the RFID chip in the document with InnoValor's ReadID (NFC Passport Reader) app to get the chip photo out of the document.
3. Save the chip photo and the document photo with the same name.

After all chip photos and document photos have been saved, the test of the total algorithm can start. The test is performed per chip photo. One chip photo will be compared with every document photo. The scores of all comparisons are stored in a table in Excel. This happens for every chip photo, so in the end there are just as many Excel files with the results as the number of chip photos.
To show how well the total algorithm works, a ROC-curve will be made. A ROC-curve is made by plotting the false accept rate (FAR) against the true accept rate (TAR). To calculate FAR and TAR, the number of false accepts (FA), the number of false rejects (FR), the number of true accepts (TA) and the number of true rejects (TR) must be known. If the names are the same and the score is higher than the percentage tested, TA is counted. If the names are the same and the score is below the percentage tested, FR is counted. If the names are different and the score is higher than the percentage tested, FA is counted. And if the names are different and the score is below the percentage tested, TR is counted. When the counting per percentage is done, FAR and TAR are calculated as shown in Equation 3 and 4.

$$FAR = \frac{FA}{FA + TR} \tag{3}$$

$$TAR = 1 - \frac{FR}{FR + TA} \tag{4}$$

In some cases, the documents have the same photos. To ensure that the ROC-curve is correct and does not give a false result, the results comparing the documents of the same person with the same photo will be removed.

V. Results

The designed experiments of the previous section are executed and the results are shown in this section.

A. Alignment

The first experiment for how well the alignment works was done with the chip photo and the document photo from the same photo. The result is shown in Figure 10. Horizontal lines are added on top of the photos, and it shows that the characteristics are on the same height. For different photos of the same person or for twins, it can be seen in Section II., Figure 2. Here it is not necessary to add horizontal lines on top of the photos to see that the alignment is not working properly.

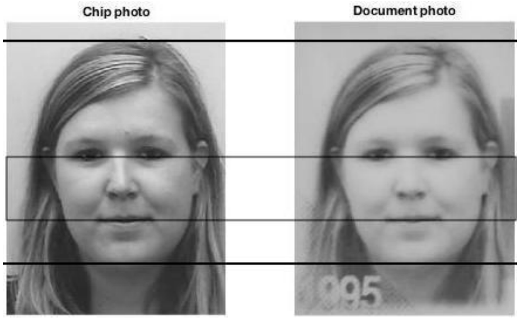


Figure 10: Same photos photographed from right above aligned

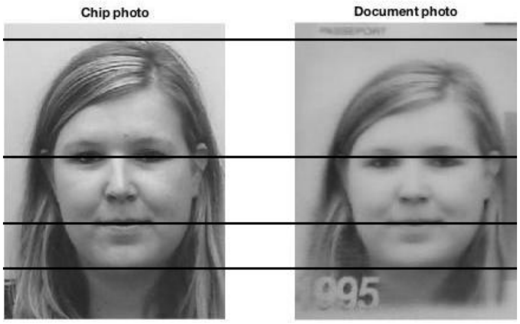


Figure 11: Same photos photographed at an angle aligned

For alignment with a photograph of the document at an angle, the result is shown in Figure 11. It can be seen that the alignment works properly when the photo is taken right above the document (Figure 10). If the photo is taken at an angle of approximately 45 degrees, it is hard to properly align the photos (Figure 11). The alignment does not work when photographing at an angle, because as transformation type similarity is used in the function in Matlab [13]. Which only enables forward and inverse transformations.

B. Photo comparison

By comparing two identical photos, Figure 12 shows the compared photos as binary photos and the comparison. In this case, the threshold value for the chip photo was set on 89. The pixels with a value lower than 89 will be set to black and the pixels with a value higher than or equal to 89 will be set to white. The threshold value for the document photo was defined on 121, because a threshold of 121 gave the highest score. This can be seen using the plot in Figure 13. The score corresponding to the threshold of 121 is 92%. This means that 92% of the pixels are the same in both photos.

Figure 12: Comparison of the binary images

In the plot showed in Figure 13 it can be seen that the score for the same photo will be at least 75%. The plot always starts at 75% and ends at 25%, also when comparing two different photos as is showed in Figure 14. This because the threshold value for the chip photo was set to 25%. So 25% of the chip photo is black and the other 75% is white. If a higher value is chosen, a person with white hair will set almost the whole face to black to reach the 30% or 35% which is hard to compare. If a lower value is chosen, a person with dark hair will have a white face, so no characteristics in the face. This is also hard to compare.

In the plot of Figure 13 it can be seen that when the threshold values would be the same for the chip photo and the document photo, a threshold value of 89, the score would be lower.

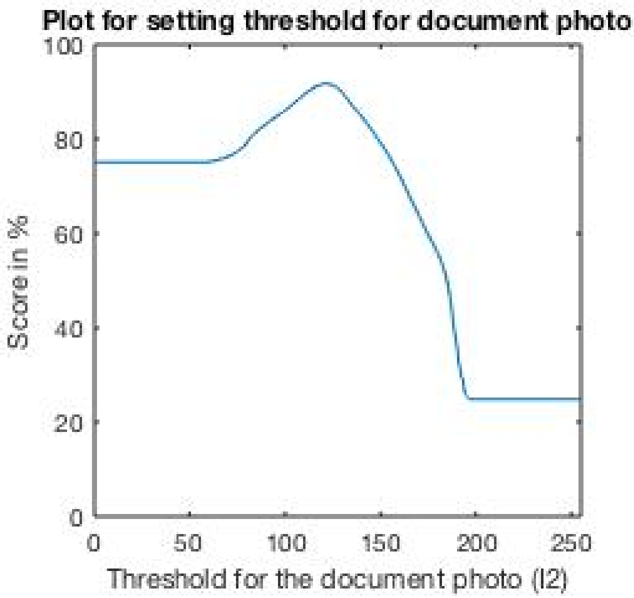


Figure 13: Plot for setting the threshold for the document photo to the highest score for the same photos

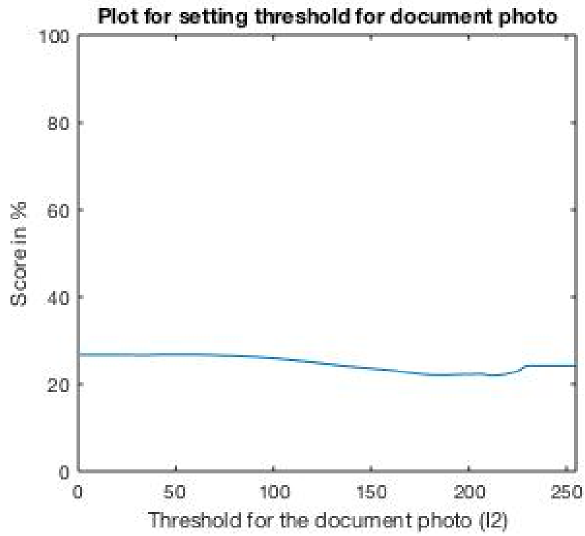


Figure 14: Plot for setting the threshold for the document photo to the highest score for different photos

C. Total algorithm

For the total algorithm, a test with 51 chip photos and 69 document photos were used. This gives a total of 3510 tests that were done, because the tests with the same photos on documents of the same person were not done. To see how well the entire algorithm works, a ROC-curve is made. The plot can be seen in Figure 15. This plot shows that the algorithm works quite well, because there are only a few false rejects. The plot also shows some false accepts, but that is only the case when looking at a passing score below 75%. It will not cause any problem in this algorithm, because the same photos always have a score higher than 75% as explained in Section V.B.

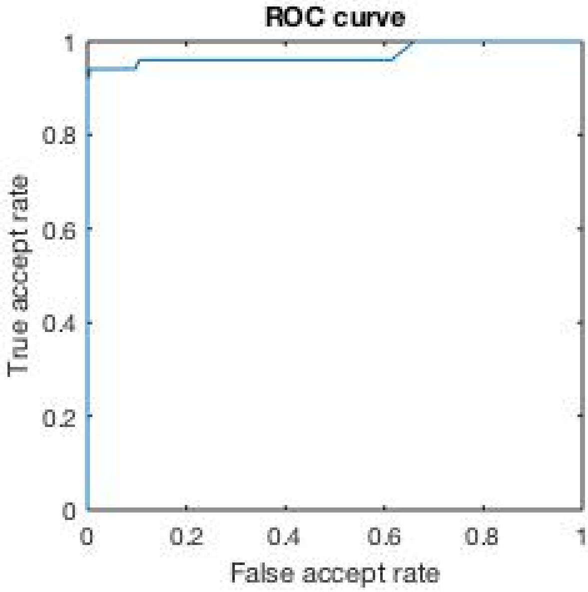


Figure 15: ROC curve of the results of the total algorithm

When looking at the scores per test, passports, ID cards and drivers licenses are working well when stating the minimum passing score to 75%. This means that if the algorithm gives a score higher than 75% percent, the photos are stated as the same and if the score is lower than 75%, the photos are stated as different. Only the residence permits are not passing the minimum score, so these are the false rejects. The score will not be above 75%, because the photo on the document was poorly visible as in Figure 16 by a glare over the photo. The glare can be caused by the angle of incidence of light.

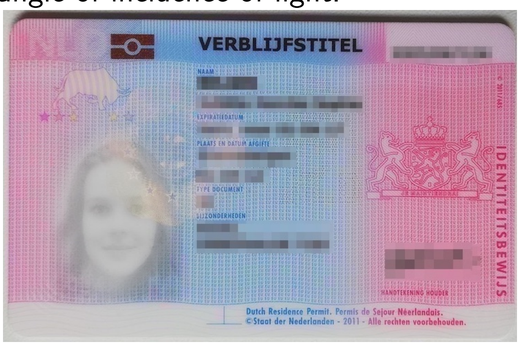
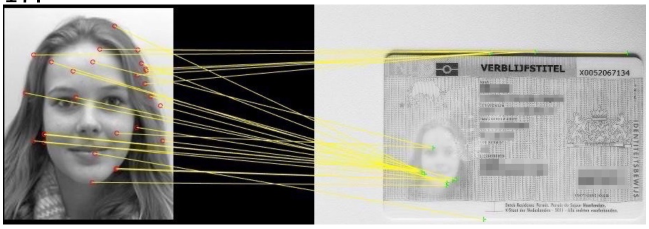
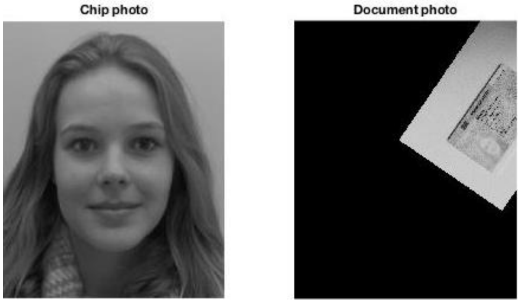


Figure 16: Residence permit

Because the printed photo on the document photo was poorly visible, the SURF feature detection could not detect and match the right points of interest which results in not being able to properly align the photos. This can be seen in Figure 17.



(a) Matching points with the same person on a residence permit



(b) Alignment for the same photo using SURF on a residence permit

Figure 17: Alignment with a residence permit

VI. Discussion

The results of the experiments will be discussed and suggestions will be made to improve the algorithm.

A. Alignment

The results show that the alignment works properly when using the same photos and take the photos from right above. If the photos differ, the alignment does not work properly. That is no problem, because only identical photos should pass the algorithm. This means that SURF will not only align the document photo with the chip photo, but also check if there is a difference between the chip photo and the document photo.
For photos taken from right above, the results show that SURF performs well, but not for photos taken at an angle. To avoid to have a small angle, it might be better to scan the document photo instead of photographing. The document will then always be from right above. Also when scanning the document, the glare over the document photos of the residence permits would be less visible. This will make the printed photo more visible, which results in better alignment for the residence permits.

B. Binary images

The threshold value for the chip photo was set on 25% and the threshold value for the document photo was defined to the value with the highest score. The plot in Figure 13 shows that this will give the highest score.
The results also show that the scores of the same photo are above the 75% and the scores of different photos are below 75%. Only the residence permits are not included, because they do not have a higher score than 75% when comparing the same photo.

VII. Conclusion & future work

The goal of this research was to compare and evaluate different existing methods to design an algorithm that can detect whether the chip photo and the document photo originate the same. To reach this goal, the following research questions were designed.
1. Does the algorithm work sufficient for aligning the document photo with the chip photo based on SURF feature detection? With a sub question, should the photo of the document been taken from straight above or does photographing at an angle still give a sufficient result for alignment?
2. Does the algorithm work sufficient to use binary images to determine whether the chip photo and the document photo have the same source? With sub question, what is the best set threshold value for the chip photo?
It can be concluded that the total algorithm works well, but some parts can be improved.
The alignment based on SURF works sufficient according to the results, but to get rid of the false rejections, it is suggested to use a different method to obtain the document photo of a residence permit. This can be done by scanning instead of photographing as explained in Section VI.A. Scanning all document photos would also ensure that there are no more document photos that have been photographed at an angle, because photographing at an angle does not give a sufficient result for alignment.
The results show that the use of binary images

in the algorithm to determine whether the chip photo and the document photo have the same source worked sufficient. But improvements can be made. An option is to add an extra step to the algorithm that will also compare the photos. Right now, the total algorithm is based on the alignment, so if SURF aligns two different photos correctly, then the comparison based on binary images can not see the difference. This will for instance happen when there is only a small change in eyes or eyebrows in the printed photo on the document, but the rest of the photo will remain the same.

References

[1] M. Gaillard1 and E. Egyed-Zsigmond, "Large scale reverse image search," pp. 127–142. [Online]. Available: https://pdfs.semanticscholar.org/9fe8/9b3bd51ab9f1fe9e883a646172afb9f3d8ba.pdf

[2] A. Rosebrock, "Facial landmarks with dlib, OpenCV, and Python." [Online]. Available: https://www.pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/

[3] N. Jayanthi and S. Indu, "Comparison of Image Matching Techniques," vol. 7, no. 3, pp. 396–401. [Online]. Available: http://www.ijltet.org/journal//147599146852.1167.pdf

[4] Wikipedia, "Blob detection." [Online]. Available: https://en.wikipedia.org/wiki/Blob_detection

[5] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," vol. 14, pp. 2–8, 2008. [Online]. Available: ftp://ftp.vision.ee.ethz.ch/publications/articles/eth_biwi_00517.pdf

[6] C. Woodruff, "Feature Detection and Matching." [Online]. Available: https://courses.cs.washington.edu/courses/cse576/13sp/projects/project1/artifacts/woodrc/index.htm

[7] Wikipedia, "Edge detection." [Online]. Available: https://en.wikipedia.org/wiki/Edge_detection

[8] O. Mavrantza, "What are the differences in first order derivative edge detection algorithms and second order edge detection algorithms?" [Online]. Available: https://www.researchgate.net/post/What_are_the_differences_in_first_order_derivative_edge_detection_algorithms_and_second_order_edge_detection_algorithms

[9] S. Sodha, "An Implementation of Sobel Edge Detection." [Online]. Available: https://www.projectrhea.org/rhea/index.php/An_Implementation_of_Sobel_Edge_Detection

[10] P. Bourke, "Histogram Matching." [Online]. Available: http://paulbourke.net/miscellaneous/equalisation/

[11] Mathworks, "detectSURFFeatures." [Online]. Available: https://nl.mathworks.com/help/vision/ref/detectsurffeatures.html

[12] ——, "matchFeatures." [Online]. Available: https://nl.mathworks.com/help/vision/ref/matchfeatures.html

[13] ——, "estimateGeometricTransform." [Online]. Available: https://nl.mathworks.com/help/vision/ref/estimategeometrictransform.html?s_tid=doc_ta

[14] ——, "imwarp." [Online]. Available: https://nl.mathworks.com/help/images/ref/imwarp.html?s_tid=doc_ta