UNIVERSITY
OF TWENTE.

GAN Generated Morphing Attack Detection by Analyzing the Geometry of Pupils

Christian Jerkovic
c.jerkovic@student.utwente.nl

University of Twente, The Netherlands

July 8, 2023

Abstract

Facial morphing is the process in which 2 images of 2 distinct faces are merged in order to create a new image that is a blend of both. Recent innovations in the field have resulted in facial morphs that are difficult for humans to identify visually. The type of morphs that will be evaluated in this research are GAN (Generative Adversarial Network) morphs. In order to combat the unreliability of humans' ability to distinguish between authentic images and generated ones, numerous methods for morphing attack detection have been developed and are being improved. Face morphing attacks have proven to be a challenge that severely hampers border control agencies' capabilities in detecting and preventing passport forgery. Due to the morphing procedure, however, the eyes of the resulting morphed image tend to have artefacts. These artefacts in many cases can distort the shape of the pupils and cause them to deviate in shape from a circle. Firstly, the eyes will be extracted from the source image in order to remove eyelashes and any other elements that could impede the segmentation of the pupil. After adjusting the image to improve the definition of the pupil, a contour will be drawn around it. The roundness of the pupil can then be used to evaluate whether or not a pupil is the result of a GAN Morph. In addition to the roundness of the original pupil contour, the roundness of the convex hull of the pupil was utilized for the detection. Finally, the ratio between the areas of the original pupil contour and its convex haul was used as a feature. The classification was done using a decision tree classifier model. This method correctly classified 66.6% of eyes as either morphed or bonafide.

Keywords— GAN, Morphing Attack, MAD, Pupil, Circularity

1 Introduction

Face biometric recognition is widely deployed in border control by comparing a photo of the document holder and their electronic passport or a national identity card. This is particularly prevalent in airports in the EU. Although many countries require the photos of passports or ID renewals to be taken by an authorized person in a government-controlled building, some countries allow for the applicant of the renewed document to provide their own pictures. For instance, The Netherlands and Spain both allow the applicant to provide their own photos in order to renew IDs or passports [1, 2].

This situation presents an opportunity for criminals to take advantage of the documents of users that have no criminal background. For instance, an individual with a criminal record could provide the governing body that is responsible for issuing passports a morphed image of themselves and an individual without a criminal record. Many modern morphing algorithms have proven capable of fooling both trained and untrained humans as well as multiple ABC (Automated Border Control) systems [3, 4]. If modern ABC systems cannot accurately detect whether a face on a document has been morphed, this would provide a gateway to criminals who are not permitted to travel to gain access to countries in which they would not be able to under normal circumstances. This could constitute a national security risk. The goal of this research is to detect whether a provided image has been morphed before being printed onto a travel document. This will be done by determining if the roundness of the pupils can be used as a metric to determine whether or not an image is the result of a GAN-generated morph. Specifically, the aim is to answer the following research question:

To what extent can the geometry of the pupils be used for morphing attack detection?

Although significant research has been done in this field, not much has been done in morphing attack detection by analyzing the geometry of pupils of GAN morphs. Effective and robust ways for the detection of morphing attacks are necessary due to the potential threat that morphing attacks pose. A convincing morph can potentially fool border personnel allowing individuals with criminal intentions, access to nations and areas they would otherwise not have access to, and commit fraud by assuming the identity of another individual. GAN-generated morphed images have artefacts present in them that cause deformation of the pupil and deviate from the circular shape of regular pupils by analyzing the geometry of a subject's pupil the photo can be classified as either a real image or a morphed image.

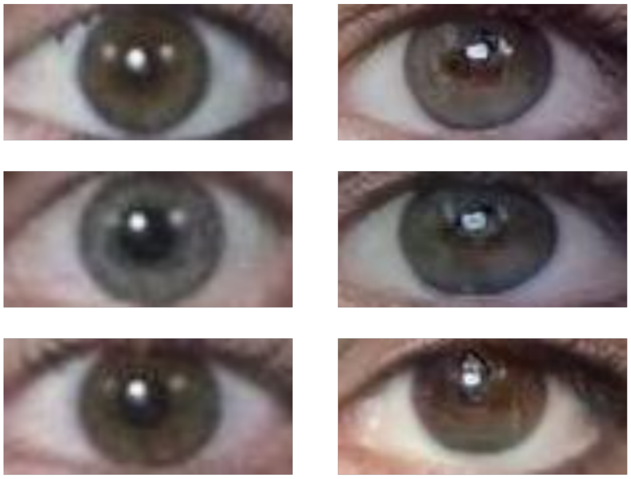


Figure 1: Bonafide vs Morphed

Following this introduction, a section discussing related work in this field and some background about the methods and technologies involved in pupil segmentation and morphing detection will be outlined. Further on, a section detailing the methodology followed in conducting this research will be described. Then, we will explain the experimental setup and the results that the aforementioned methodology obtained. Finally, we will discuss the results and suggest some methods for future work or improvements.

2 Backgroud and Related work

GAN-generated morphs differ from the landmark-based approach in the fact that they do not involve the alignment of landmarks in order to generate a morph and instead use 2 machine learning agents, a generator, and a discriminator, where the generator produces samples that should be accepted by the discriminator. Landmark-based approaches use predefined landmarks as a reference when creating morphs which preserve key features of the subjects used in the morph. Typically this is done by determining the Delaunay triangles of each face, averaging them and then blending them together. GAN-based morphs on the other hand produce an image that is similar to the overall appearance of the subject image without placing specific attention to landmarks. Although GAN-generated face morphs do not currently pose as substantial a threat as landmark-based morphs, improvements in GAN-based methods could improve their effectiveness [5].

There are multiple state-of-the-art MAD (Morphing Attack Detection) methods that use a variety of techniques, one such method is via texture analysis [6, 7, pp. 146–153]. Another approach is via residual noise analysis [8]. And finally, there are deep learning approaches [9]. The Handbook of Digital Face Manipulation and Detection [10], in addition to explaining the core concepts of face morphing, provides multiple sources and insight regarding this field of research.

Although studies have been done to detect whether non-morphed synthetic GAN image has been generated by analyzing the geometry of pupils [11]. not much research has been done into identifying if an image is morphed by primarily using the roundness of the pupil.

IrisParseNet provides a variety of models for pupil segmentation using machine learning methods. The top 3 models, however, are required to be trained locally with a Cuda-enabled GPU, which was not available for this study. The one exception is eyecool [12], which hosts the pre-trained model on Baidu and requires the installation of a 3rd party download manager application for access to the dataset.

3 Methodology

Section 3.1 will outline the datasets that have been chosen to conduct this research as well as the reasons they were selected. Section 3.2 will proceed in describing the process in which the eyes are extracted from the provided images. Section 3.3 will discuss the image preprocessing required for pupil segmentation. Section 3.4 will outline the steps taken in pupil segmentation. Section 3.5 describes the metrics that will be extracted following the segmentation. And finally Section 3.6 will discuss the experimental setup.

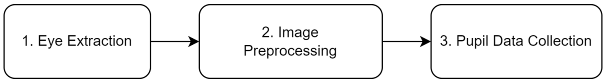


Figure 2: Methodology Pipeline

3.1 Dataset

Two datasets have been used to conduct this research, one of the morphed images and one of bonafide. The dataset chosen for the bonafide images was Face Research Lab London Frontal_Neutral which contains 102 images with a resolution of 1024 x 1024 pixels. For the morphed images, the Face Research Lab London morph_stylgan set was used.It contains 1122 Morphed faces based on the faces provided in Lab London Frontal_Neutral Using the Stylegan Method.

3.2 Eye Extraction

In this step of the process, the aim is to extract the eye from the image. In Figure 3 we can see a graph with every step taken in this process. First, we need to read the image in step 1. Then, we move on to find some landmarks that will locate the eye on the image. Later, dlib's landmark dataset is used to locate the convex hull of each eye in step 3. This step is later explained in more detail. In step 4 we fill the area outside the landmark point with the color white. Essentially to "erase" any other areas that are not the eyes. Finally, we save the image of each eye in step 5.

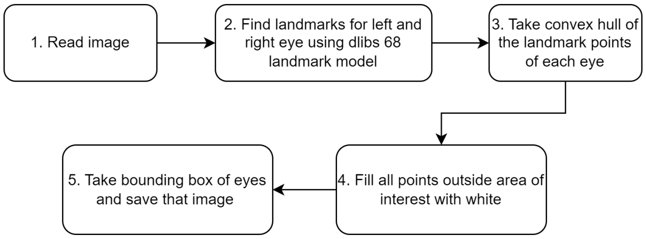

1. Read image
2. Find landmarks for left and right eye using dlibs 68 landmark model
3. Take convex hull of the landmark points of each eye
4. Fill all points outside area of interest with white
5. Take bounding box of eyes and save that image

Figure 3: Eye extraction pipeline

Step number 3 is the key part of the Eye Extraction process. In this step, we use dlib's 68 facial landmark dataset in order to locate the eyes [13]. After the 6 points per eye are found a contour of the convex hull is drawn. This is shown in the lefthand picture in Figure 4. After, all areas outside of the contour boundary are filled with white. The resulting image can be seen on the right hand of Figure 4. This method was chosen to minimize the effect the eyelashes could have when drawing the contours. We did see in previous attempts that the eyelashes were being confused as being part of the pupil which in turn affected the step to calculate its circularity.

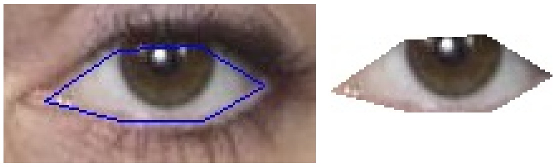


Figure 4: The convex hull of eye and extracted eye

3.3 Preprocessing

Following extraction, the images are cropped to the central third of the eye. This was done using Numpy and removing one-third of the image from the left and the right. Upon cropping the image the brightness is set to -127, the lowest level that openCv allows. This was done in order to maintain the darkness of the pupil while increasing the contrast between it and the rest of the eye. Following the lowering of brightness, the contrast is increased by a factor of 4. This procedure is done using the openCV convertScaleAbs function and setting the alpha parameter to 4 and the beta parameter to -127. In the first image of Figure 7, we can see the original cropped image. In the second image, after increasing contrast and lowering brightness we can see that the pupil has become much more defined. Finally, in preparation for contour selection, the image is converted to grayscale and its colours are inverted.

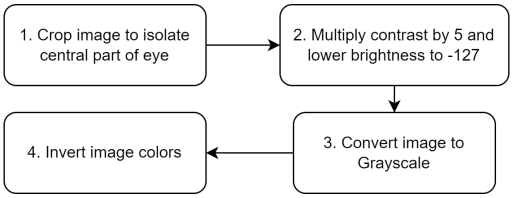

1. Crop image to isolate central part of eye
2. Multiply contrast by 5 and lower brightness to -127
3. Convert image to Grayscale
4. Invert image colors

Figure 5: Preprocessing Pipeline

3.4 Contour Selection

A binary threshold is manually applied to the image in order to get the best outline of the pupil. Using OpenCV's built-in drawContour method, a contour of the pupil is drawn. The following restrictions are placed on the contour selection: the centre of the contour must be within 50% of the image centre, and the area of the contour must be within 1% and 30% of the original images' resolution. Furthermore, the contour must be between 5% and 20% of the original eyes' area. When the contour is found, its circularity, as well as the circularity of its convex hull and the ratio between their areas is added to the dataset for further testing. Images 5-7 of Figure 7 show the result of the contour selection process.

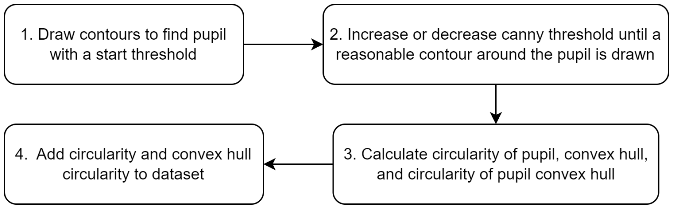

1. Draw contours to find pupil with a start threshold
2. Increase or decrease canny threshold until a reasonable contour around the pupil is drawn
3. Calculate circularity of pupil, convex hull, and circularity of pupil convex hull
4. Add circularity and convex hull circularity to dataset

Figure 6: Contour Selection Pipeline

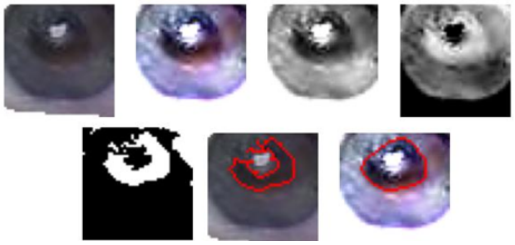


Figure 7: Preprocessing and Contour Drawing

3.5 Pupil Metrics

Following contour selection, multiple metrics are calculated. The one most closely investigated is circularity, which is defined as follows:

$$\frac{perimeter^2}{4 * \pi * area}$$

This is a common measurement in computer vision in determining the roundness of a subject [14]. The closer circularity is to 1 the more circle-like the contour is. Conversely, as it increases further from one in the positive direction the less round an object is considered to be.

If the object in question is a perfect circle the circularity will be 1. This is due to the fact that for a circle the formula would be as follows:

$$\frac{(2 * \pi * r)^2}{4 * \pi^2 * r^2}$$

In this case, the numerator and denominator are equivalent resulting in a circularity of 1

Metric	Definition
Circularity	The roundness of the original contour extracted.
Convex Hull Circularity	The roundness of the convex hull of the original contour extracted.
Ratio	The ratio between the area of the original contour and the area of its convex hull

Table 1: Metric Definitions

The circularity of the convex hull of the pupil contour is an important metric, especially in eyes where the pupil is occluded by reflection. In such cases, the circularity of the original contour is higher because the reflection hides the true shape of the pupil. This phenomenon is shown in Figure 8. The eyes that were chosen for analysis were the ones in which the pupil was either mostly or fully visible. In addition, the pupils that had very low contrast between the iris and the pupil were not considered for analysis as drawing a contour around the pupil using the aforementioned methodology proved unfeasible.

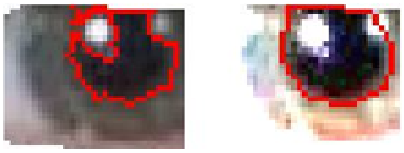


Figure 8: Original contour and convex hull

3.6 Experimental setup

In order to determine if circularity can indeed be used as an indicator for morphing attack detection, statistical data was collected on the aforementioned metrics. The statistical metrics collected are the mean and standard deviation for both bonafide and morphed eyes. Further on, histograms of the distribution of bonafide and morphed eyes are plotted. The final experiment is classifying the eyes in the dataset using a Decision Tree Classifier from which accuracy, precision, recall and F1 score will be calculated. This will be done using Sklearn's Decision Tree Classifier model. Finally a confusion matrix will be plotted to illustrate the results of the classification.

4 Experiment Results

In total, data from 48 bonafide and 46 morphed pupils were extracted and processed. All the steps in section 3 were done using Python in a jupyter notebook. The metrics defined in Section 3.5 were collected from each image and stored in a dataset. From the dataset, the mean and standard deviation of the original pupil contours circularity as well as the mean and standard deviation of the convex hull of the pupil were calculated. Finally, the mean and standard deviation of the ratio between the original pupil contour area and the convex hull area of bonafide and morphed eyes were calculated. All results will be shown below.

4.1 Original Contour Circularity

Metric	Value
Mean of bonafide pupil circularity	3.47
Standard deviation of bonafide pupil circularity	1.60
Mean of Morphed pupil circularity	6.42
Standard deviation of Morphed pupil circularity	3.28

Table 2: Circularity of Pupils

The distribution of morphed and bonafide pupil circularity can be seen in Figure 9. We have also found that out of the 46 morphed eyes present 86% of them were found above the mean and 41% were found above the maximum bonafide circularity which in this case was 7.07. Furthermore, the overlap between bonafide and morphed circularity was 58%, these are the morphed pupils that had a circularity below the maximum bonafide circularity.

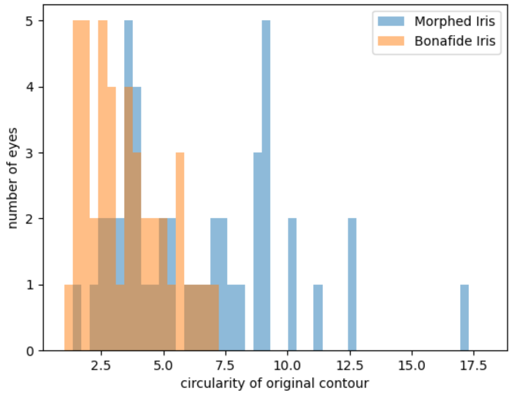


Figure 9: Pupil Circularity Distribution

4.2 Convex Hull Circularity

Metric	Value
Mean of bonafide pupil convex hull circularity	1.1170
Standard deviation of bonafide pupil convex hull circularity	0.0507
Mean of morphed pupil convex hull circularity	1.2266
Standard deviation of morphed pupil convex hull circularity	0.0985

Table 3: Convex Hull Circularity of Pupils

89% of morphed images had a greater convex hull circularity than the mean convex hull circularity of the bonafide images and 32% of morphed pupils had a greater convex hull circularity than the maximum convex hull circularity of the bonafide images. Finally, the overlap between bonafide and morphed convex hull circularity was 67%.

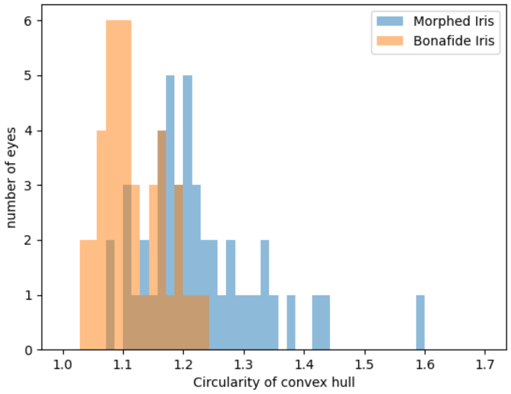


Figure 10: Pupil Convex Hull Circularity Distribution

4.3 Ratio of Original Pupil Contour to Convex Hull Contour Area

Metric	Value
Mean of bonafide original to convex hull area ratio	0.7011
Standard deviation of bonafide contour to convex hull area ratio	0.1363
Mean of morphed contour to convex hull area ratio	0.5616
Standard deviation of morphed contour to convex hull area ratio	0.1282

Table 4: Ratio of original pupil contour area to convex hull contour area

82% of morphed images had a smaller ratio of original pupil contour area to convex hull contour area than the mean of the bonafide images. 6% of morphed pupils had a smaller ratio of original pupil contour area to convex hull contour area than the minimum ratio of the bonafide images. Finally, the overlap between bonafide and morphed area ratio was found to be 91%.

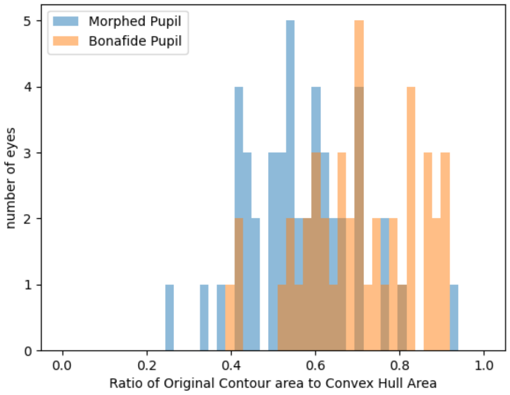


Figure 11: Ratio of original pupil contour area to convex hull contour area distribution

4.4 Classification

Metric	Value
Accuracy	66.66%
Precision	65.71%
Recall	71.87%
F1 score	68.65%
Type I error rate	31.65%
Type II error rate	38.70%

Table 5: Classication Results

The classification was done with a 30/70 % train-test split using a Decision Tree classifier. Figure 12 represents the confusion matrix of the results presented above. On both axes, 0 represents morphed eyes and 1 represents bonafide eyes. For instance, the top left corner corresponding with coordinates (0,0) is the quadrant where the actual eyes are morphed and the predicted result is also morphed. Of the 31 morphed images 12 images were classified as bonafide resulting in a false positive rate (Type I error rate) of 31.65%, and of the 32 bonafide images 9 were classified as morphed resulting in a false negative rate (Type II error rate) of 38.70%.

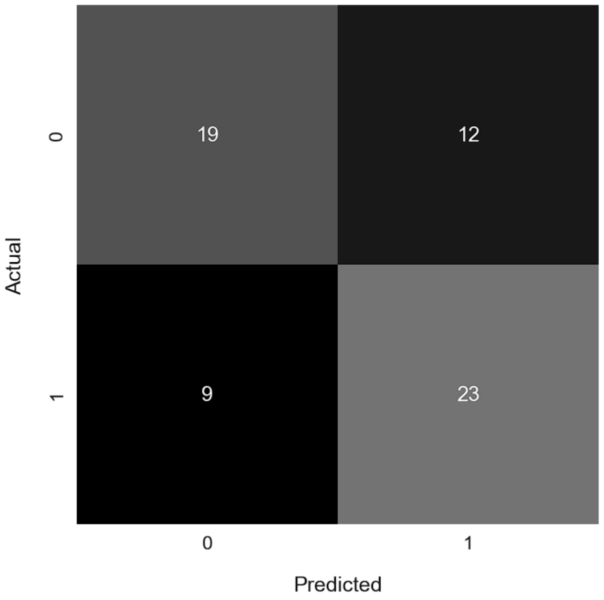


Figure 12: Confusion Matrix

5 Discussion

The primary reason for the high average circularity in the bonafide dataset is primary due to reflections present in some of the images in which a contour cannot be fully drawn. Furthermore, there are many morphed eyes that also appear quite round which results in some morphed eyes having a low circularity. With an improved iris segmentation method there might be a reduction in circularity which might yield a higher precision score. Both the original contour circularity and convex hull circularity had significantly less overlap than the ratio between their areas. Overall the accuracy of the classifier performs as expected due to the significant overlap present in the features. However, it does indicate that morphs that contain pupils that are particularly non-circular can be detected with this method. One limitation of using the geometry of a pupil as a predictor for morphing attack detection is that it does not account for bonafide pupils that have naturally present deformities such as ones occurring due to disease. With the current pupil segmentation method, we have found that up to 71.87% of eyes of GAN-generated morphs can be detected, and overall 66.66% of eyes can be correctly classified using the accuracy metric.

6 Conclusion

The main focus of this research was to answer the following question:

To what extent can the geometry of the pupils be used for morphing attack detection?

Throughout this research, we have seen what aspects of an image can affect the contour circularity of a pupil. The main contributors to an increased circularity in bonafide eyes are the full pupil not being visible, reflections present that obscure the pupil, and the colour of the iris. These factors have skewed the circularity of pupils to be higher than originally anticipated. Despite this, we have demonstrated that there is a measurable difference between the circularity of bonafide and morphed pupils. With the proposed methodology, we can conclude that up to 71.87% of eyes of GAN-generated morphs can be detected, and overall 66.66% of eyes can be correctly classified.

7 Future Works

The primary bottleneck in the methodology is the need for manual threshold selection. Although automatic thresholding techniques like otsus [15] thresholding exist, it was not applicable here as the contour selection was too inaccurate and in no cases would it correctly draw a contour around the pupil. One of the main issues of the contour selection process is that it is highly dependent on the difference in brightness in a given group of pixels. So even shadows cast by the eyelashes that are not part of the pupil can tend to creep into the area of the pupil as the threshold is decreased. On possible solution to this issue is to decrement the binary threshold value starting from 255 and collect all contours that are found outside the boundary of the pupil and fill them with the binary colour of 0 in the binary image to eliminate those points from any contour selection. Finally, when there is only one contour in the area of interest the iteration can stop. The main issue with this approach was properly centring the iris in order to determine the area of interest which is also an ongoing field of research. Another way to improve the iris would be to forgo algorithmic pupil segmentation and proceed with a deep learning approach such as the methods proposed by Wang et al. [12]. One more aspect that could be improved is the sample size of the dataset, currently, it contains 90 eyes of which 46 are morphed and 44 are bonafide.

References

[1] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. *Paspoortuitvoeringsregeling Nederland 2001.* nl. ministeriele-regeling. Last Modified: 2022-04-26. URL: https://wetten.overheid.nl/BWBR0012811/2022-02-05#HoofdstukIII_Paragraaf4_Artikel38 (visited on 05/03/2023).

[2] Ministerio del Interior. *Real Decreto 896/2003, de 11 de julio, por el que se regula la expedicion del pasaporte ordinario y se determinan sus caracter-*

isticas. July 2003. URL: https://www.boe.es/eli/es/rd/2003/07/11/896 (visited on 05/03/2023).

[3] Robin S. S. Kramer et al. "Face morphing attacks: Investigating detection with humans and computers". In: *Cognitive Research: Principles and Implications* 4.1 (July 2019), p. 28. ISSN: 2365-7464. (Visited on 05/03/2023).

[4] David Robertson et al. "Detecting morphed passport photos: a training and individual differences approach". In: *Cognitive Research: Principles and Implications* 3 (June 2018). DOI: 10.1186/s41235-018-0113-8.

[5] Sushma Venkatesh et al. *Can GAN Generated Morphs Threaten Face Recognition Systems Equally as Landmark Based Morphs? – Vulnerability and Detection*. July 2020. URL: http://arxiv.org/abs/2007.03621 (visited on 06/30/2023).

[6] Raghavendra Ramachandra, Kiran Raja, and Christoph Busch. "Detecting Morphed Face Images". In: September 2016. DOI: 10.1109/BTAS.2016.7791169.

[7] Christian Kraetzer et al. *Digital Forensics and Watermarking: 16th International Workshop , IWDW 2017, Magdeburg, Germany, August 23-25, 2017, Proceedings*. January 2017. DOI: 10.1007/978-3-319-64185-0.

[8] Sushma Venkatesh et al. "Morphed Face Detection Based on Deep Color Residual Noise". In: *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. November 2019, pp. 1–6. DOI: 10.1109/IPTA.2019.8936088.

[9] Clemens Seibold et al. "Detection of Face Morphing Attacks by Deep Learning". In: July 2017, pp. 107–120. DOI: 10.1007/978-3-319-64185-0_9.

[10] *Handbook of Digital Face Manipulation and Detection*. en. URL: https://www.springerprofessional.de/en/handbook-of-digital-face-manipulation-and-detection/20085002 (visited on 05/03/2023).

[11] Hui Guo et al. *Eyes Tell All: Irregular Pupil Shapes Reveal GAN-generated Faces*. en. September 2021. URL: https://arxiv.org/abs/2109.00162v4 (visited on 05/03/2023).

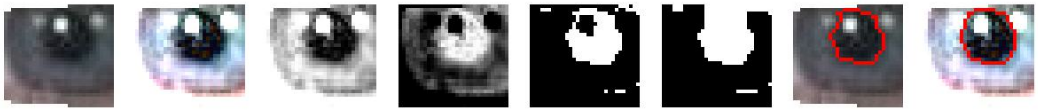
[12] Caiyong Wang et al. "NIR Iris Challenge Evaluation in Non-cooperative Environments: Segmentation and Localization". In: *2021 IEEE International Joint Conference on Biometrics (IJCB)*. August 2021, pp. 1–10. DOI: 10.1109/IJCB52358.2021.9484336.

[13] Davis E. King. "Dlib-ml: A Machine Learning Toolkit". In: *Journal of Machine Learning Research* 10.60 (2009), pp. 1755–1758. ISSN: 1533-7928. (Visited on 06/24/2023).

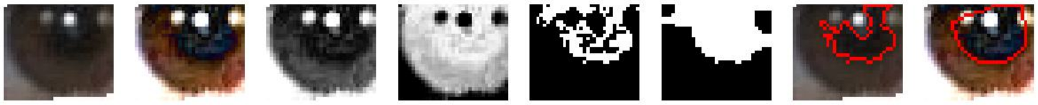
[14] Yasuhiro Takashimizu and Maiko Iiyoshi. "New parameter of roundness R: circularity corrected by aspect ratio". In: *Progress in Earth and Planetary Science* 3 (January 2016). DOI: 10.1186/s40645-015-0078-x.

[15] Nobuyuki Otsu. "A Threshold Selection Method from Gray-Level Histograms". In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (January 1979). Conference Name: IEEE Transactions on Systems, Man, and Cybernetics, pp. 62–66. DOI: 10.1109/TSMC.1979.4310076.

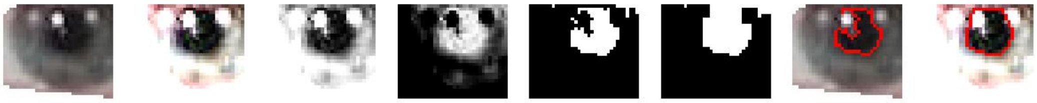
A Bonafide Contours



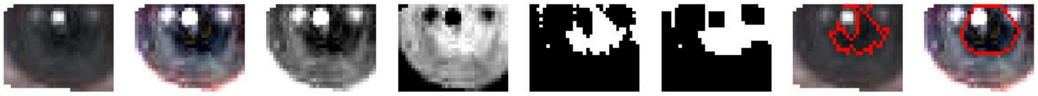
Roundness: 1.304
Convex hull roundness: 1.047
Ratio contour area to convex hull area: 0.915



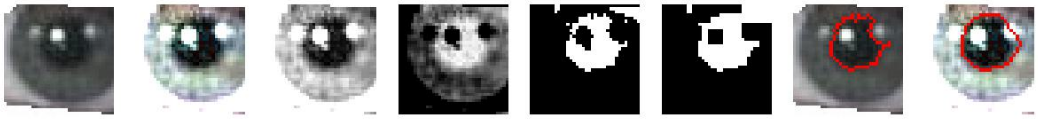
Roundness: 4.373
Convex hull roundness: 1.142
Ratio contour area to convex hull area: 0.615



Roundness: 2.139
Convex hull roundness: 1.122
Ratio contour area to convex hull area: 0.798

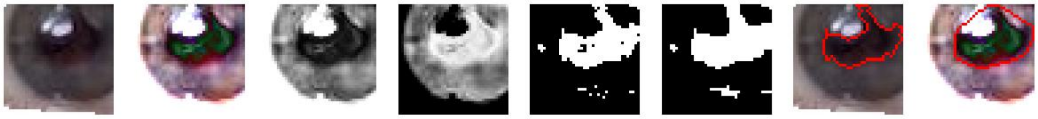


Roundness: 4.854
Convex hull roundness: 1.080
Ratio contour area to convex hull area: 0.561

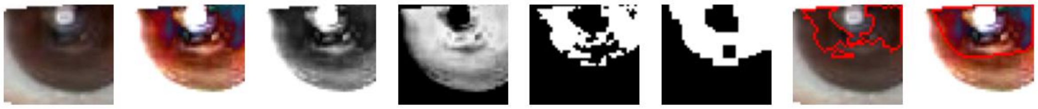


Roundness: 1.906
Convex hull roundness: 1.051
Ratio contour area to convex hull area: 0.822

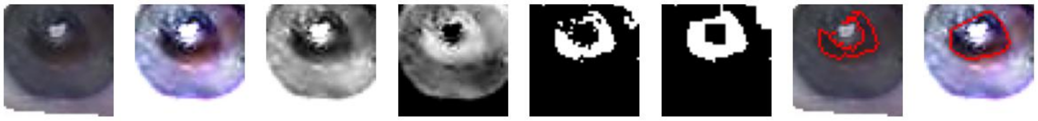
B Morphed Contours



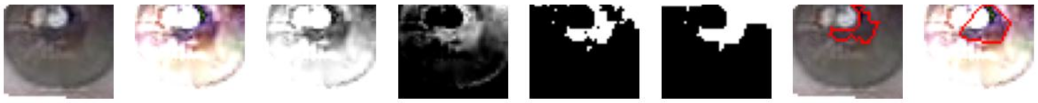
Roundness: 2.994
Convex hull roundness: 1.111
Ratio contour area to convex hull area: 0.664



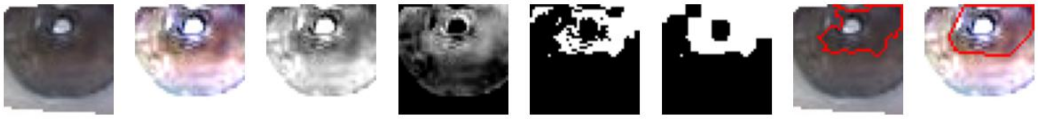
Roundness: 6.575
Convex hull roundness: 1.329
Ratio contour area to convex hull area: 0.577



Roundness: 5.477
Convex hull roundness: 1.084
Ratio contour area to convex hull area: 0.613



Roundness: 3.754
Convex hull roundness: 1.329
Ratio contour area to convex hull area: 0.546



Roundness: 2.349
Convex hull roundness: 1.225
Ratio contour area to convex hull area: 0.812