Forensic Facial Recognition Utilizing Cartoonization Segmentation

Wouter Suidgeest
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
w.f.suidgeest@student.utwente.nl

ABSTRACT

The topic of implementing automated facial recognition software has been a challenge for decades. Multiple very well working systems have been developed over the past years and are already in use. However, most of these systems lack transparency, causing them to be unusable in the field of forensic science. To assist in this field, the Facial Identification Scientific Working Group (FISWG) has described and documented a set of facial features suitable for facial recognition in forensic science, and how they need to be treated. However, these specific features are currently not sufficiently detectable by automated systems, as some features need to be segmented very precisely. In this study, a novel face recognition technology is designed, which utilizes cartoonization technology in the segmentation process, as the simplified shapes in cartoonized images could make the segmentation of FISWG features more precise. Figure 1 shows a visualisation of the individual steps of the proposed system. While the method seems promising from visual inspections of the individual steps alone, due to a number of currently unresolved problems in certain types of images, the cartoonization process does not yet improve the reliability of the system.

Keywords

Facial recognition, Cartoonization, FISWG, Feature detection, segmentation

1. INTRODUCTION

Automated facial recognition (FR) is a technology that has become more and more relevant in the past decades. Usages can now be found in everyday life, such as in airport identification systems, or to unlock your phone. Due to the large use of CCTV security camera's, automated FR could become a very important aspect of forensic science. To implement such applications in the challenging context of forensic science, and as a response to the many concerns in earlier applications [5, 22], the Facial Identification Scientific Working Group (FISWG) [8] has defined and documented a set of facial characteristic descriptors. These characteristic descriptors are sets and details of facial features from which a person can be recognized, such as the shape of their eyebrows, or the size of their face.

35th Twente Student Conference on IT Jul. 2nd, 2021, Enschede, The Netherlands.

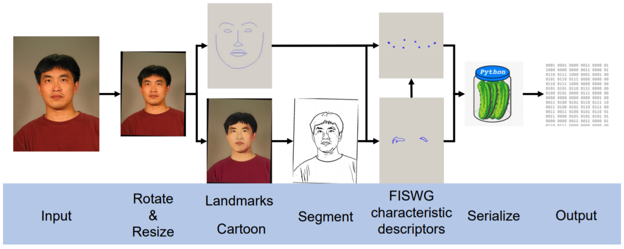


Figure 1: Visualisation of the new system.

FISWG characteristic descriptors have not completely been validated yet, but they have been proven to be an effective measurement to distinct a particular subject from a group of subjects [25]. However, automated FR using FISWG characteristic descriptors is still a challenge, especially in the context of segmenting and detecting the characteristic descriptors.

Cartoonization is a technology that tries to recreate an image in a cartoon style. This is a valuable technology in multiple fields of science, such as computer vision and computer graphics [20]. Typical cartoonization systems achieve this goal by searching for areas of similar colours, and by replacing these with similar shapes of the main colour, resulting in a cartoon-styled image. Some applications focus on cartoonizing facial features only, and the specific process of the necessary steps differ per application, but multiple technologies are proven to be successful. The resulting image displays a drawing of the initial image, with much coarser distinctions between different shapes.

1.1 Objective and goals

With the problem and its context established, this study sets the following goals:

- **Goal 1:** To implement a facial recognition system, solely based on FISWG characteristic descriptors in portrait pictures by using cartoonization to help in identifying and segmenting these descriptors.
- **Goal 2:** To test the system as described in **Goal 1** against an off-the-shelf facial recognition system, in various forensically relevant situations.

In order to achieve these goals, this research will try to answer the following two research questions (RQ), subdivided into sets of more testable sub-RQs:

- **RQ1:** To what extent can we use cartoonization technology to segment reliable FISWG features for facial recognition?
  - **RQ1.1:** Which cartoonization applications are most successful in segmenting FISWG features?

- **RQ1.2:** How can FISWG features be extracted from a cartoonization application as identified in **RQ1.1**, and used for a facial recognition system.

- **RQ2:** How well does a facial recognition system as identified in **RQ1.2** perform compared to an off-the-shelf system.
  - **RQ2.1:** How well does a facial recognition system as identified in **RQ1.2** perform compared to an off-the-shelf system, for normal, quality images such as in passports.
  - **RQ2.2:** How well does a facial recognition system as identified in **RQ1.2** perform compared to an off-the-shelf system, for quality images, but for challenging cases such as twins.
  - **RQ2.3:** How well does a facial recognition system as identified in **RQ1.2** perform compared to an off-the-shelf system, for images of poor quality, such as CCTV footage.

Expected is, that at the end of this research, a clear answer can be given to both RQs and all of their sub-RQs. Where, as compared to traditional systems, facial recognition using cartoonization has improved performance with respect to certain FISWG features.

2. RELATED WORKS

In this section, research will be explored in which will be described how a forensic facial recognition (FFR) application should work, and what requirements it should meet. Next, different studies and developments of FR applications will be discussed, and it is shown how these don't meet the previously established requirements, to display the research gap in FFR.

2.1 Requirements

Some issues with existing solutions have already been briefly mentioned, but in order to highlight them properly and to be able to create a new, proper, system a list of requirements for an FFR system needs to be determined. The requirements for a correct facial recognition system are mainly focused on ensuring the fairness and reliability of the system, and a lot of research already exists in this field. This paper will try to focus on the following requirements:

1. Requirements for morphological analysis (the method of comparing faces by facial features and components) as defined by FISWG in their Facial Comparison Overview and Methodology Guidelines [10].
   (a) Morphological analysis is based on the evaluation of the correspondence among facial features, components (e.g. nose, ear and their components) and their respective component characteristics. And discriminating characteristics such as scars or tattoos.
   (b) The morphological analysis process does not rely on the classification or categorization of features (e.g., round face, Roman nose).
   (c) The examination and decision-making process should be fully documented and include an independent technical review.
   (d) The method requires consistency.
2. Extra requirements for this research, including some requirements from Microsoft's facial recognition principles [15].
   (a) The system needs to be transparent.
   (b) The system needs to treat all people fairly
   (c) The system needs to be reliable.

2.2 Existing solutions

In this section, a selection of some of the current state-of-the-art facial recognition systems will be displayed, and an explanation will be given on why they are unfit for forensic use.

The first system we look at is the in 2014 developed GaussianFace [12], the first system to outperform humans in terms of facial recognition accuracy. This system achieved such performance by separating an image into many small vectors, and labeling these using local binary patterns. Afterwards, they let an improved version of the Kernel Fisher discriminant analysis determine important features out of these labels. That, in combination with a lot of diverse training data, resulted in such accuracy. Now, looking at the requirements for an FFR system, this system in no way meets requirements 1a and 1b, as the analysis is based on features determined by an algorithm, instead of on valid facial features.

Next, we look at some of the currently most popular facial recognition models; Google FaceNet [19], Facebook DeepFace [21], and OpenFace [2]. These systems have been developed by tech giants or top universities, and return state-of-the-art results. However, all of them achieve this by training a CNN, and letting this CNN determine a vector which represents the input face. And just like for GaussianFace, this doesn't ensure that requirements 1a and 1b are met, making the systems unfit for forensic use.

Next to these standard state-of-the-art systems, efforts have been made to develop facial recognition systems using biometric features [25]. But to date, simply no solution exists which correctly (by requirements 2c and 2b)) manages to identify and compare faces while also meeting requirements 1a, 1b, 1c, and 1d [1].

3. METHODOLOGY AND APPROACH

As an improvement over existing solutions, this research mainly focused on creating a moderately reliable system, which is very transparent, and solely uses biometrically justifiable features. Since the novelty of this research can be found in the additional step of cartoonization technology in segmentation, the rest of the system was kept fairly simple. This section will explain for each part of the system how it works, and which choices were made during the process.

3.1 Pre-processing

The first step in processing an image is making sure the image is fit for the rest of the process. The first necessary pre-processing steps that this system makes are rotating and resizing the input image. Currently, the images are rotated using hand-annotated pupil coordinates. Out of these coordinates, the angle of the rotation is calculated using the following formula:

$$angle = -\arctan((l\_eye_x - r\_eye_x)/(l\_eye_y - r\_eye_y)) \quad (1)$$

Rotating is necessary to calculate distance based facial features later on, as with a rotated face, the coordinates of the points in question will be shifted.

Next, the image is resized, such that $max(width, height) <= 720$. This is to improve the processing speed of the calculation-heavy components of the application, while still preserving

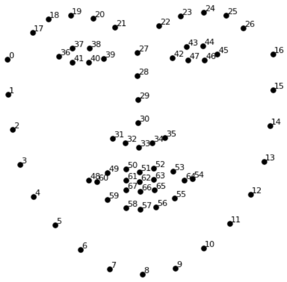


Figure 2: The 68 facial landmarks as detected by dlib.

System \ feature	RC	OS	CF	SR	PD
CartoonGAN [3]	Yes	Yes	No	Yes	No
White-Box [23]	Yes	Yes	No	Yes	No
APDrawingGAN [24]	No	No	Yes	Yes	Yes
Toon-Me [14]	Yes	No	Yes	No	Yes

Table 1: Overview of cartoonization systems by features: RC = Returns color, OS = Supports all OS, CF = Cartoonizes face only, SR = Supported by research, PD = Preserves small details.

the image quality. Next to that, having input images of equal size allows us to hard-code some size based parameters later in the system.

3.2 Landmark detection

After pre-processing the image, landmark detection should be done on the image in order to globally locate facial features. For this process, dlib's state-of-the-art face-recognition library was used [6]. This library can detect all faces in an image, and for each face, locate 68 facial landmarks as shown in Figure 2 using HOG and Linear SVM. Out of these landmarks, the outer landmarks of the eyes and eyebrows (17, 21, 22, 26, 36, 39, 42, and 45) are stored for later use.

3.3 Cartoonization

For the step of cartoonization, an existing technology needed to be chosen. In order to keep the rest of the system accurate and reliable, the cartoonization system needed to meet a couple of requirements:

- The system needs to make the outlines of shapes more clear while preserving the original global shape.
- The system needs to remove small, unnecessary details from the image.
- The system needs to be reasonably fast.
- The system needs to be free and open-source.
- The system needs to be easily implementable.

During this research, multiple options were considered, where all of which had different pros and cons. Table 1 lists the differences between the considered systems by a list of features partially explaining the predetermined requirements. Out of these options, the choice was made to use the White-Box cartoonization system which was developed for a CVPR 2020 paper [23]. This choice was made because it is one of the systems that support cartoonization of entire images, causing it to excel in preserving large shapes while neglecting smaller details. Because by visual inspection this system proved to be very well in preserving original colours and shapes. Because the system is easily implementable by the existing TensorFlow implementation. And because the system is one of the best performing cartoonization technologies overall.

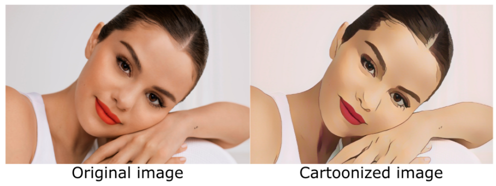


Figure 3: Example white-box cartoonization.

The White-Box cartoonization method works by training a Generative Adversarial Network (GAN) to separately identify three white-box representations from an image; "The surface representation that contains a smooth surface of cartoon images, the structure representation that refers to the sparse color-blocks and flatten global content in the celluloid style workflow, and the texture representation that reflects high frequency texture, contours, and details in cartoon images" (Wang et al., 2020 [23]), with these different layers, a cartoon is created. By using white-box technology, the system manages to provide much control and adjustability for the user. Offering the option for it to be used in many use-cases. In Figure 3 an example cartoon using the White-Box cartoonization method can be seen.

3.4 Segmentation

Over the past decades, especially since the introduction of convolutional neural networks (CNNs), many new advanced segmentation systems have been developed. The most notable ones being R-CNN, Fast R-CNN, Faster R-CNN, and Mask R-CNN, where each of these systems is an improvement, especially in processing time, of the previous system. However, in order to keep the system as transparent as possible, and because the additional step of cartoonization should also help in improving the segmentation process, the choice was made to mainly focus on more traditional algorithms.

Out of these more traditional segmentation algorithms, this research looked at the usability of watershed segmentation, binary threshold, and adaptive thresholding. For all of these methods, a Gaussian blur is added to the image, and then it is converted to a grayscale image. Using this grayscale image, the segmentation algorithm transforms this to a binary image containing only the 'darkest' parts of the image. Each of these algorithms decides which areas belong to the 'darkest' in a different way. The watershed algorithm does this by imagining the grayscale image as a landscape where the darker a pixel is, the higher its position is. Then, it selects and returns the lines running over the highest ridges. Binary threshold is a method where a certain value is chosen, and every spot that's darker than this value is selected. This value is often chosen using Otsu's method, which automatically tries to optimize this value by minimizing the weighted within-class variance. This results in a value which lies within the peak grey values of the image, resulting in a value which segments the largest differences. Adaptive threshold is the same as binary threshold using Otsu's method, but the threshold value is decided for smaller individual regions.

Eventually, two of these looked promising for segmenting the required features, binary threshold using Otsu's method, and adaptive thresholding. However, in order to

Figure 4: Output of adaptive thresholding segmentation for an image with and without cartoonization.

also be able to process faces that are further away, and in order to extract more possible facial features, adaptive thresholding with block size = 23 and C = 5 was used. In Figure 4, it can be seen what the result of this segmentation process is, and how the cartoonization makes the segmentation more detailed.

3.5 Outline detection

With the binary image resulting from the segmentation process, OpenCV's [16] FindContours() method was used to extract the outlines corresponding to the binary shapes. Out of these contours, the contours of the eyebrows are selected using the coordinates of the landmarks previously determined by dlib.

3.6 Feature extraction

As previously discussed, the newly designed system will purely use facial features in the form of FISWG characteristic descriptors. This is because they are strictly defined, can easily be calculated, and are based on biometrical research. Since this study focuses on the segmentation process, two eyebrow-related characteristic descriptors have been chosen. As previous research has shown that these are rich containers of information [11, 18], and they are clearly distinguishable parts of the face.

3.6.1 A-E measurements

The first set of characteristic descriptors are the A-E measurements of the eyebrows. As can be seen in Figure 5, this is a set of 5 distance measures over the edge points of the eyes and eyebrows. In text, they can be described as follows:

- A: Vertical distance between the right-most and left-most position of the eyebrow.
- B: Vertical distance between the left-most position of the eye, and the left-most position of the eyebrow.
- C: Vertical distance between the right-most position of the eye, and lowest position of the eyebrow.
- D: Horizontal distance between the left-most position of the eye and left-most position of the eyebrow.
- E: Horizontal distance between the right-most position of the eye and right-most position of the eyebrow.

The coordinates of the 5 points required to calculate these distances can easily be taken from the previously calculated landmarks and outlines. Furthermore, all of these distances are calculated relative to the eye size, here measured as the horizontal distance between the left-most and right-most positions of the eye.

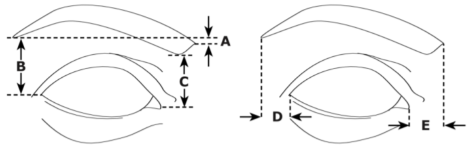


Figure 5: A-E characteristic descriptors of the eyebrow, as defined by FISWG [9].

Figure 6: Bézier curves drawn over the outlines of eyebrows.

3.6.2 Shapes of the eyebrows

The second type of characteristic descriptors this system uses is the shapes of the eyebrows. FISWG does not advise an explicit method to specify this shape, so multiple options could have been valid. However, it is important that the shape is specified using a shape descriptor, in order to be able to compare different eyebrows properly. In this system, we chose equidistant sampling of 128 points on the original shape and calculated the corresponding 100 points Bézier curves as shape descriptor. Bézier curves are curves that are encoded as multiple points, where the curve follows a smooth line from a point towards the next point. Bézier curves are also fairly easy to calculate and as a shape descriptor, invariant to translation and rotation [17]. In Figure 6, two eyebrow outlines are plotted in red, with the corresponding Bézier curves drawn in blue lines over them. And as can be seen, the overall shape is largely preserved.

3.7 Similarity scores

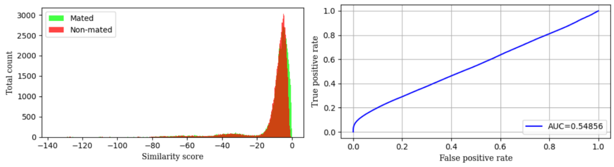
In order to eventually compare two images with each other, similarity scores between these images need to be calculated. The entire previously described process converted a facial image to one 10-dimensional vector (5 A-E measurements per eye), and one 200-dimensional vector (100 points Bézier curves per eyebrow), both of which are confirmed means to describe the given face. Since the sizes of and values in these vectors are so different, two similarity scores are calculated between two images. For both of these scores, the similarity was chosen as the inverse of the total difference between the two vectors. In other words, the similarity score of the A-E measurements (ASS) between two vectors AE_1 and AE_2 is calculated by the following formula: (where $d(x, y)$ is a function that calculates the absolute distance between the two given numbers x and y)

$$ASS = -\sum_{i=0}^{9} d(AE_1[i], AE_2[i]) \tag{2}$$

And the similarity score of the Bézier curves (BSS) between two vectors $Bezier_1$ and $Bezier_2$ is calculated by the following formula: (where $d(a, b)$ is a function that calculates the euclidean distance between the two given coordinates a and b)

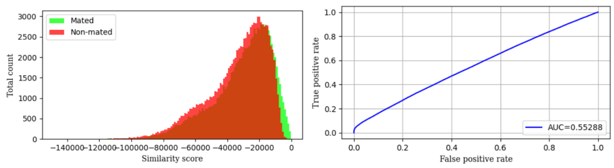
$$BSS = -\sum_{i=0}^{199} d(Bezier_1[i], Bezier_2[i]) \tag{3}$$

For both these similarity scores, the result always is a neg-



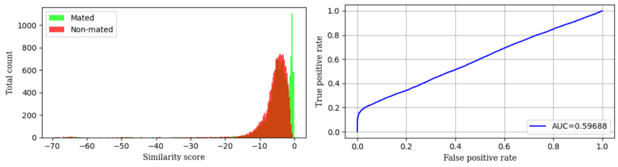
(a) Histogram (b) ROC curve of 7a

Figure 7: Histogram and ROC curve of similarity scores of A-E measurements for the system with cartoonization.



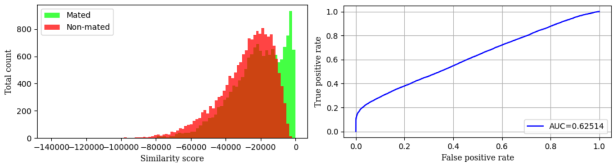
(a) Histogram (b) ROC curve of 8a

Figure 8: Histogram and ROC curve of similarity scores of eyebrow shapes for for the system with cartoonization.



(a) Histogram (b) ROC curve of 9a

Figure 9: Histogram and ROC curve of similarity scores of A-E measurements for the system without cartoonization.



(a) Histogram (b) ROC curve of 10a

Figure 10: Histogram and ROC curve of similarity scores of eyebrow shapes for the system without cartoonization.

ative number. This ensures that the more the two given vectors resemble, the higher the score is.

Eventually, to decide whether a calculated similarity score between two images should be rated as mated (both images show the same person), or non-mated (the two images show a different person), a threshold value needs to be chosen. If the similarity score is higher than this threshold value, the images are classified as mated, otherwise, they are classified as non-mated. This method is chosen as opposed to e.g. training a biometric classifier, as the method is very transparent, giving very explainable results. In order to decide such a threshold value, the similarity scores for a lot of mated and non-mated pairs need to be calculated using this system, and from the results, an optimal value can be chosen.

4. RESULTS

With the newly proposed architecture defined, a prototype could be developed and tested in order to get actual results. For these tests, images from the Fall 2003 collection of the FRGC dataset [7] were used. This dataset contains over 10000 quality images of 141 different subjects under different circumstances. Out of this set, every image was tested against every other image of the same subject, and against an equal amount of images of random other subjects.

The results of these tests are presented in two ways; by a histogram plotting the resulting similarity scores, and by the ROC curve with its area under the ROC curve (AUC) value of this histogram. Out of the histogram, the desired discrimination threshold value can be determined. The ROC curve visualizes the discriminative ability of the system, by plotting the true positive rate against the false positive rate for different discrimination threshold values. The AUC is also given, this value is equal to the probability that the model ranks a random positive example more highly than a random negative example [13], and can be used as a mean to compare two models. A model is perfect for AUC $= 1$, and completely random for AUC $= 0{,}5$.

The results from these tests are plotted in Figures 7 and 8. And as can be seen from Figures 7a and 8a, the histograms for mated and non-mated similarity scores nearly completely overlap. This is reflected in the ROC curves in Figures 7b and 8b, where both measurements reach an AUC of approximately 0,55, indicating that the model's results are close to random.

In order to see what the actual effect of cartoonization on this entire process is, the same tests were also run over a random subset of 3735 images, but without the process of cartoonization. Note that no other parameters in the system were changed for these tests.

The results from these tests are plotted in Figures 9 and 10. And as can be seen from Figures 9a and 10a, the histograms for mated and non-mated similarity scores are more separated than they were for the system with cartoonization in Figures 7a and 8a, however, they still overlap quite a lot. This is reflected in the ROC curves in Figures 9b and 10b, where the AUC-values of both measurements are 0,59 and 0,62, indicating that the model is better than the model with cartoonization, but still of very low quality.

With the entire system defined, a prototype developed, and a large-scale test performed, the system can be reviewed against the requirements as established in Section 3. In the following list, a short recap will be given of each requirement, while stating whether and how this system meets this requirement

1. FISWG requirements:
   (a) Morphological analysis is based on the evaluation of the correspondence among facial features and their respective component characteristics, by merely using FISWG documented characteristic descriptors.
   (b) The morphological analysis process does not use classification or categorization of features, by merely using FISWG documented characteristic descriptors.
   (c) The entire process is documented well, and an independent technical review could be formed out of the documentation.
   (d) The method is consistent, as every analysis is performed in exactly the same way.
2. Extra requirements:
   (a) The system is transparent, as the system mainly

resorts to traditional processing and evaluations techniques while being well documented.

(b) The system treats all people fairly, as every analysis is performed in exactly the same way and the model was trained on a large dataset of various types of faces.

(c) The system is not reliable, as can be seen in the results from the first tests, the system barely has any real discriminative ability.

5. CONCLUSION AND DISCUSSION

As can be seen from the results of the verification phase, this system, which is engineered for using cartoonization technology, scores significantly better without cartoonization than with. From this, we can conclude that this implementation does not yet achieve the preferred results, namely an improved segmentation process due to cartoonization. However, from a simple visual inspection of the individual steps, it can be seen that some parts do actually yield promising results. And because the system manages to meet all requirements for a correct forensic facial recognition application, except for decent reliability, the studied process could still be of value for future research. The rest of this section will, step by step, explain which parts of the system are still open to improvement, and in which way.

5.1 Cartoonization

Because of this study's tight planning, and the fact that it was preferable if the prototype could be run on a simple windows notebook, the choice of cartoonization technology could have been explored better. From the other options, also next to the ones listed in this paper, some seemed just as, if not more, promising for this purpose as White-Box cartoonization. However, after choosing the White-Box cartoonizer, there was no time to also test the prototype with other technologies.

5.2 Outline detection

Because of two major problems in the segmentation process, outline detection often isn't accurate. The first issue is caused by the White-Box cartoonization. This system often draws, previously non-existing, dark lines to indicate edges, including around the nose. As can be seen in Figure 11, if this happens, the nose and eyebrow are sometimes seen as one single object after segmenting. When later comparing this shape (nose + eyebrow) with a correct shape (eyebrow), the results do not tell much about the resemblance between the two given eyebrows. This is an issue for some quality, controlled images such as the one in Figure 11, but it happens almost constantly, and more intense, for uncontrolled images from a further distance. Most probably because some parameters were optimized for controlled images. Next to the fact that some outlines get drawn incorrectly during cartoonization and segmentation, outlines sometimes also still get chosen incorrectly. This can for example happen if the eyebrow landmarks detected by dlib match better with the top of the eye than with the eyebrow.

In order to see how large the influence of these outline detection errors is, a small extra test was performed on 21 images of 7 random subjects from the data. In this test, the subset was tested in four ways. First, by visual inspection, a subset of 7 images was filtered, which were all the images for which no large outline detection errors occurred. Next, these were tested with and without cartoonization in order to give a more complete overview of the effects. In Table 2 the results from these tests can be

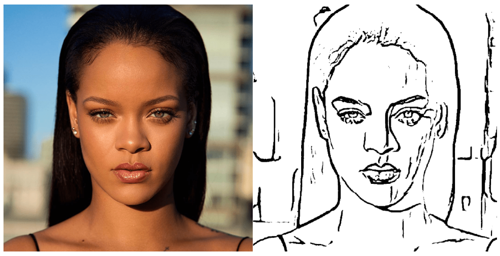


Figure 11: Example of incorrect eyebrow outline detection. (Initial image taken from Mike Marsland/Getty Images)

Conditions		A-E AUC	Shape AUC
Uncartoonized	Unfiltered	0.59	0.61
Uncartoonized	Filtered	0.66	0.77
Cartoonized	Unfiltered	0.63	0.57
Cartoonized	Filtered	0.94	0.77

Table 2: AUC values of the system on a small subset of the data, filtered on whether outline detection errors occurred.

found. And as can be seen here, all AUC values increase if these errors are filtered out, and especially the cartoonized results on A-E values seem very promising.

5.3 Feature extraction

As previously described, the shape descriptor chosen to represent the eyebrow shapes was the Bézier curve, as it is invariant to translation and rotation [17], and easily implementable. However, other research suggests that 2D Fourier shape descriptors yield more promising results, partly since they are also invariant to scaling next to translation and rotation [4]. While in this study there was no time to also explore the effects of using a Fourier shape descriptor for the eyebrows, there is the possibility that this could have improved the results. Especially since the results for shape comparison are consistently worse than that of the A-E measurement comparison.

5.4 Uncontrolled vs controlled images

The current prototype was made with a focus on 'controlled' images, which are the images from the FRGC dataset [7] taken in a studio setting. However, the eventual, previously shown, tests were run on the entire dataset, consisting of controlled and uncontrolled images. Which are images taken from a different distance and with different lighting. While the prototype was not engineered for this input, this evidently caused worse results. By running a simple analysis on the same datasets, but with the uncontrolled images filtered out, the AUC value of both, the A-E measurements and the eyebrow shapes are already increased by more than 10 percent.

5.5 Research questions

With the global conclusions drawn, and after a thorough discussion about this study's process and results, conclusions to the initial research questions can be drawn.

- **RQ1.1**: As was decided during the development of the new architecture in Section 5, cartoonization applications that accentuate facial features while preserving their original shape are useful for the segmentation of FISWG features. However, as was discussed before, it is important that this process is done by a system that does not add any new lines or contours to the image, as that effect can cause unexpected results.

- **RQ1.2**: As expected, all shapes are very clear after cartoonizing an image, and by using a standard landmark detection in combination with a traditional segmentation algorithm such as adaptive thresholding, the shapes can often be extracted correctly.
- **RQ1**: Cartoonization technology can be used as an extra step in a traditional facial recognition system, making the resulting features more precise. However, it is very important that cartoonization is done perfectly, as it can also make the results a lot less reliable if done incorrectly.
- **RQ2.1**, **RQ2.2**, **RQ2.3**, and **RQ2**: Eventually this study did not test the newly developed system against an off-the-shelf facial recognition system. Instead, it was tested against itself without the step of cartoonization, and only for one specific image quality. So from these tests only, no reliable conclusions to these research questions can be given. However, it was concluded that the facial recognition system as defined in this paper was not of very high quality, so an off-the-shelf system would most likely outperform the new system on most images.

6. FUTURE WORK

As can be read in the discussion, many parts of this research are open to improvements, and additional attention is required before any useful conclusions about cartoonization can be drawn. For future research, any of the problems discussed in the discussion section could be explored better. More concretely put, this would come down to first; exploring the effects of different cartoonization systems. Second; studying whether the outline detection can be refined in a way such that errors are prevented and detected. Third; to try different shape descriptors such as the 2D Fourier shape descriptor. A final point for further research would be to tweak the system such that it can handle images from different distances equally well. This can be achieved by e.g. making some currently hard-coded parameters variable based on the interpupillary distance (IPD), or by zooming and resizing the image based on the IPD.

7. REFERENCES

[1] T. Ali, R. Veldhuis, and L. Spreeuwers. *Forensic Face Recognition: A Survey*. Number TR-CTIT-10-40 in CTIT Technical Report Series. Centre for Telematics and Information Technology (CTIT), Netherlands, Dec. 2010.

[2] T. Baltrusaitis, P. Robinson, and L.-P. Morency. Openface: An open source facial behavior analysis toolkit. pages 1–10, 03 2016.

[3] Y. Chen, Y.-K. Lai, and Y.-J. Liu. Cartoongan: Generative adversarial networks for photo cartoonization. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9465–9474, 2018.

[4] S. Conseil, S. Bourennane, and L. Martin. Comparison of fourier descriptors and hu moments for hand posture recognition. In *2007 15th European Signal Processing Conference*, pages 1960–1964, 2007.

[5] N. R. Council. *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press, Washington, DC, 2009.

[6] dlib facial recognition. http://dlib.net/face_recognition.py.html. Accessed: 2021-06-20.

[7] Face recognition grand challenge. https://www.nist.gov/programs-projects/face-recognition-grand-challenge-frgc. Accessed: 2021-04-30.

[8] Facial identification scientific working group website. https://www.fiswg.org/index.htm. Accessed: 2021-04-29.

[9] Fiswg; facial image comparison feature list for morphological analysis. https://www.fiswg.org/FISWG_Morph_Analysis_Feature_List_v2.0_20180911.pdf. Accessed: 2021-06-25.

[10] Fiswg; facial comparison overview and methodology guidelines. https://www.fiswg.org/fiswg_facial_comparison_overview_and_methodology_guidelines_V1.0_20191025.pdf. Accessed: 2021-06-25.

[11] F. Juefei-Xu and M. Savvides. Can your eyebrows tell me who you are? In *2011 5th International Conference on Signal Processing and Communication Systems (ICSPCS)*, pages 1–8, 2011.

[12] C. Lu and X. Tang. Surpassing human-level face verification performance on lfw with gaussianface. 04 2014.

[13] Google; machine learning crash course. https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc. Accessed: 2021-06-23.

[14] V. Madhaven. Toon-me github. https://github.com/vijishmadhavan/Toon-Me, 2013.

[15] Microsoft; facial recognition: It's time for action. https://blogs.microsoft.com/on-the-issues/2018/12/06/facial-recognition-its-time-for-action/. Accessed: 2021-06-24.

[16] Open source computer vision website. https://docs.opencv.org/3.4/index.html. Accessed: 2021-06-20.

[17] N. M. Patrikalakis. *Shape Interrogation for Computer Aided Design and Manufacturings*. Springer, 2002.

[18] J. Sadr, I. Jarudi, and P. Sinha. The role of eyebrows in face recognition. *Perception*, 32:285–93, 02 2003.

[19] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2015.

[20] S. Sepehri Nejad and M. A. Balafar. Component-based cartoon face generation. *Electronics*, 5(4), 2016.

[21] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. 09 2014.

[22] R. Van Noorden. The ethical questions that haunt facial-recognition research. *Nature*, 587(7834):354–358, 2020.

[23] X. Wang and J. Yu. Learning to cartoonize using white-box cartoon representations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[24] R. Yi, Y.-J. Liu, Y.-K. Lai, and P. L. Rosin. APDrawingGAN: Generating artistic portrait drawings from face photos with hierarchical gans. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '19)*, pages 10743–10752, 2019.

[25] C. Zeinstra. *Forensic Face Recognition: From characteristic descriptors to strength of evidence*. PhD thesis, UT, Netherlands, Nov. 2017. CTIT Ph.D. thesis series no. 17-439.