DEFACTO Morphing Database

Gracy Jason Joseph
Data and Knowledge engineering
Otto-von-Guericke University
Magdeburg, Germany
gracy.joseph@st.ovgu.de

Shweta Pandey
Data and Knowledge engineering
Otto-von-Guericke University
Magdeburg, Germany
shweta.pandey@st.ovgu.de

Abstract- There have been various approaches to the generation of doctored images through different techniques. In this paper, we have created an analysis and classification of the morphed images which are created by morphing two genuine images together. Images are morphed by some forgeries categories namely- Copy & Move, Splicing, Object Removal, and Morphing. We have used Morphing as a forgery technique for experimenting with the images.

In this Morphing technique, there are 3 methods that are mainly used, such as combined, splicing and complete. With that, we have calculated and analyzed the scores fetched through Biometric analysis and for the quality check we have made use of Forensic analysis.

Keywords- Forgeries, combined, splicing, complete, alpha, morphed, evaluation, Biometric, Dermalog matcher, Forensic, Detector, ANANAS web service, Error rate, Accuracy.

1. INTRODUCTION

Morphing is a technique that gives a transition to one image from another. Hence it is a fluid visual transformation of one to another image. [14]

It's been used in many Hollywood films or in gaming purposes in video or computers and in creating animated movies. However, the morphing technique is not limited for the above purposes, instead, it is a powerful technique that intensifies multiple multimedia projects such as presentations, e-book illustrations, etc. [14] Morphing technique has been rapidly increasing and it has become very easy for anyone to create a doctored image and use it for any purpose. It is increasing crimes, fake identity, impersonation et al. The morphed images can be used by an imposter to sham the identity; hence it is very crucial to detect these types of scammers. [7]

Morphing is a procedure between two genuine images, wherein first, the specification of the correspondence point is initiated which is also called 'feature specification'.[14] Then the feature primitive of both the images is processed, which is mostly manual work. After which each primitive denotes the image landmark and then corresponding features of the mapping function are reckoned. Basically morphing process is a three-phase technique that mainly consists of feature specification, warp generation, and transition control**.** [14]

Numerous techniques have been made available for the detection of the doctored or morphed images and various categories have been used to analyze them. While morphing, many parameters are being taken into consideration, such as setting up alpha parameter, setting up methods to blend the images. We have used a 'DEFACTO dataset'[8] to study a wide range of forgeries. It is publicly available and it is being used widely to compare the results. [8]

This dataset has been used for experimentation purposes meant to study doctored images and to analyze them. We have created many morphed images for each of the methods by using ANANAS web services.

2. RELATED WORK

In recent years, we have come across numerous datasets that are publicly existing. As mentioned by a researcher that the first available dataset was the Columbia Gray Dataset [8] which contained only one forgery approach that is splicing and the images were created in grayscale. Post which a newer version Color Dataset was released**.** [8]

After that, there are many other datasets that were released which even consisted of tampered images with better realistic tampering. CASIA V1.0 and V2.0 [6][8] were the ones which contain only splicing and cove-move approach with post-processing methods and they are still widely used because it comprises of many forged images. There were datasets that were released that had image and video manipulations and it is not possible still to access their entire versions.

It is being stated that the first dataset with real-world tampered images was the Wild Web which had more than 10000 images. [8] There are datasets that are released with the author collecting images from active Reddit thread Photoshop Battles which is called the PS-Battle Dataset, [6][8] in which the people try to create a better manipulation of the photo. In this, the people have collected 90000 doctored images and 10000 genuine images, and this particular dataset is meant to give a dataset with the long-lasting benchmark. [8]

With the above datasets being existing, we have used a novel dataset DEFACTO [8] which were publicly available and which had images of the celebrities with their genuine as well as morphed images of it. That we have used as a comparison subject with the morphs created by ANANAS web service by running some Python scripts. Also, when it comes to morphing attack detection methods, there are various no-reference schemes. There are also some approaches that gave the best performance in detecting the morphing attacks and for security assessment scenarios, a morphed image is a threat if more than one sample is verified positively against it. [11] [12]

3. DATASET OVERVIEW

The images that we have collected is from the DEFACTO dataset which is publicly available for experimentation purpose. This database consists of 200 genuine images and their morphed images.

3.1 Forgeries technique

There are four major categories of forgeries namely copy-move, splicing, object-removal, and morphing. [8]

Copy-move is the forgery technique that consists of the duplication of a particular element inside the image. Splicing is a method wherein one part or portion of an image is copied and added to the other image. [8] In the object-removal technique, a part of a whole object is removed from an image by the use of an algorithm. While in morphing, warping, and blending are the two processes that are used in two images together. In each of the above techniques, there is post-processing which is applied such as rotation, scaling, contrasting et al. [8] An image composite can be most likely a formation of the basic operations. Hence the detection of these categories would be a lot more different.

3.2 Method used in our experiment

In our case, we used the method of morphing. In the morphing technique when 2 images are being transformed, there are multiple intermediate images that are created. The first intermediate image would be similar to the first image and the last intermediate image would be similar to the second image. The one which is in the center is the 50:50 of both the faces. [14] We have the images which would be taken from the DEFACTO database and then it would be morphed by using ANANAS web service.

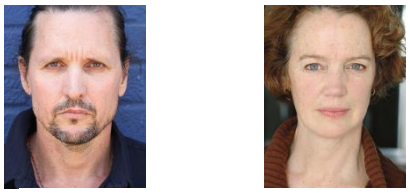
After which we can start our experiment by analyzing these images through 'Biometric and Forensic analysis'. [9][12][7]

There are some forgeries that could be easily detected by human eyes. But we wanted to create some datasets which would be challenging for us to detect.

4. EXPERIMENT

4.1. Morphing Technique

We considered two original images from the DEFACTO database. [8] This database has front-facing images of artists with a very neutral expression as the base to create face morphing forgery. It has a total of 200 images. It also has images that are morphed already. We need to create their morphs and would be analyzing them.



Face A Face B

Fig 1: Genuine images from DEFACTO database

The face morphing process can be seen on the images, taking 2 original images and using the parameter landmarks to extract a set of facial landmarks and with that according to the image the first roughly alignment begins with the eyes, then a weighted factor/parameter alpha $\alpha \in [0,1]$ [8] is taken into consideration of the two images and then with the landmarks computation, the two images are blended with a precise alignment. Post, the two images are α (alpha) blended with a factor, there is a local scaling performed for a better skin tone. The alpha factor is needed to decide which particular face's structure should be more visible. For a warping procedure, blending each image can be weighed by the alpha factor. [13]

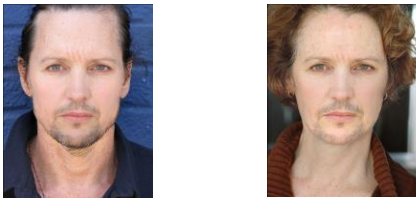
In our case, we have taken only alpha factor into the experiment as it gave us the proper morphed picture. And as there are methods into morphing we have taken each method with the parameters.

A. Combined Method

The images are first morphed using combined method with alpha parameter as 0.5 which means that 50% facial features of face A and 50% facial features of face B have to be taken.

With this, the resulting image should be a 50:50 ratio of both. The alpha factor 0.5 is the average point that we have taken. The images are blended according to the parameters set. The color balances, facial structure is then produced accordingly.

Below is the morphed image through the 'Combined method'. [10]

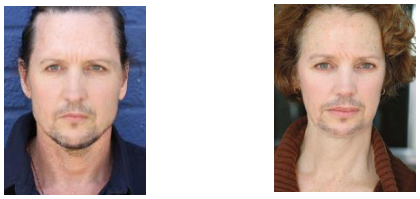


Morph A Morph B

Fig 2: Morphed images created through ANANAS web service

B. Splicing method

The images are then experimented with the second method of morphing called splicing. We have taken the alpha parameter as 0.5 to check how the final result looks like. The images created by splicing won't be similar to the combined method.



Morph A Morph B

Fig 3: Morphed images created through ANANAS web service

The images give the result from warping and blending of the facial regions with the subsequent adding into one of the original backgrounds. It makes the images clear of ghosting artifacts. [9]

C. Complete method

Complete [8] is the third method of morphing which doesn't give a perfectly blended image. But it shows us the color association, the points where the two images would be blended and how.

The result is not a perfectly doctored image but it gives the middle transition of both the images. It gives the 50:50 image of both.

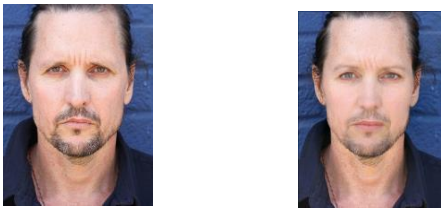
Warped and blended image

Fig 4: Transitioned image created through ANANAS web service

4.2 Biometric analysis

Biometric quality is one of the three-folds in which a morphed image should be successfully verified by any AFR-system (automated face recognition) against its own original image. [10] The biometric quality and a particular AFR system show the quality in terms of error rates in a system. The greater the error rates, there would be a higher chance of the biometric quality of the doctored image.

In our case, we took the method Dermalog face recognition in 'Biometric analysis'. [10] It is the technique that combines the biometrics of facial structure. [11] Specially with the combination with ID cards, it is able to properly verify that the person with the information matches the given identification documents. This particular technique operates reliably in various conditions such as age, gender, size, contrast, light, color, etc. It is also used as a part of border control in most of the countries, hence it is quite reliable and gives a good verification rate. [10] We have used this method in order to get the verification score of the self-created morphed image, to know if it gives a proper evaluation result. For a better score, we have taken the threshold score of more than 80.



Face A Morph A (combined)

Fig 5: The scores from the first original image and morphed image from the combined method

In the above experiment, we have used a matcher and we got a score of 99 and the evaluation was a Match.
That means the Morphed image has captured the facial features of the genuine image.
Post this, the matcher technique [10] is used on the original face B and second morphed image by the combined method.

Face B Morph A (alpha:0.5) Morph A (alpha:0.7)

Fig 6: The scores from the second original image and morphed images from the combined method

In the above experiment, we have used a matcher and we got a score of 59.01 and the evaluation was a NoMatch.
That means the percentage of facial features captured by the Morphed image from the genuine image B is less.

We then altered the alpha value from 0.5 to 0.7 and then we got the score as 92.64 and the evaluation as a match, which means now the morphed image has captured the features with the maximum percentage.
Figure 7 is an example of morphed images, as to how the images were created from the ANANAS web service, and with different morphing methods.
The original images were taken by the 'DEFACTO dataset'. [10] We have used three methods of morphing which are complete, splicing and combined. Complete [8] is a blended image of two original images. It is a 50:50 of both of them.
A morphed image created through a combined method is created by the parameters by taking the alpha factor as 0.5. In this, if we see, the first morphed image gives a visual classification with 1st genuine image as the background and the second morphed image with 2nd genuine image as the background. When you see visually of the morphed images of the combined method the transition, texture and face structure changes.
And if we visually see the morphed image of the splicing method then the skin color, the elongated face changes.

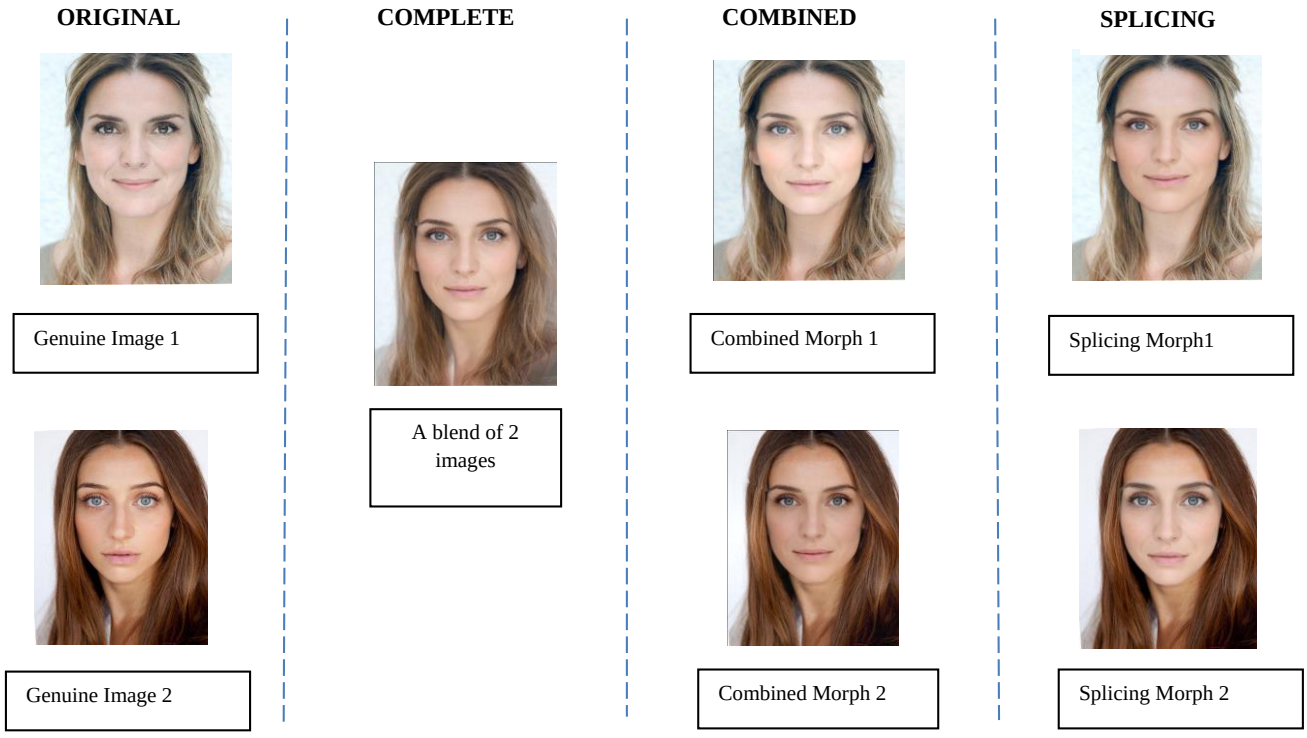
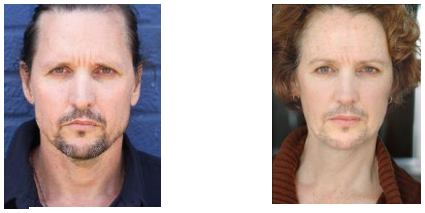


Fig 7: Example of morphed/doctored images: complete, combined and splicing approaches; original images are from DEFACTO database

In this case, we are experimenting in making such a morphed image which cannot be detected. There are times when after taking the alpha factor 0.5 the image doesn't turn out to be a good doctored image, which means the score is less than the threshold value. It means it's not a better morphed image of the original and has not captured that percent of facial features of the original image.
Here we have played with the alpha factor value, [11][10] till we get a perfectly morphed image. 'Dermalog face recognition' method is used to compare the original to the other morphed image. Through Matcher we got the scores of both the images. The threshold score which we used for analyzing was 80 and the evaluation result as a match. If incase the score was below 80, we made it sure we morph the image in such a way that the score reaches the threshold value which matches as the evaluated result.

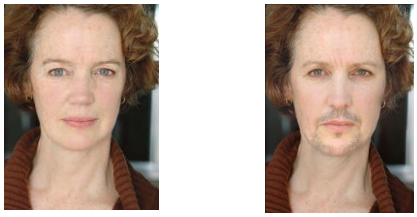
We have done this, in order to get a perfectly morphed image so that we can experiment it to find the analysis if it really gives us a better result in finding a doctored image. The score that we got after comparing the original image with its morphed image was 59.01 which was less than our threshold value of 80. We then changed the alpha factor to 0.6 then 0.7 till we get a perfectly morphed image with the evaluated score greater than the threshold value.

Now we have performed the same technique for the 2 genuine images with the second morphed image created by the combined method.



Face A Morph B (combined)

Fig 8: The first original image and morphed images created from the combined method



Face B Morph B (combined)

Fig 9: The second original image and morphed images created from the combined method

The second morphed image was compared with both the genuine images and the scores were above the threshold value 0.80, so we didn't alter any factors.

The same way we have performed the alteration of samples which aren't higher than the threshold value. And it is done on the morphed images made through splicing and combined.

4.3 Forensic analysis

Forensic analysis means that the morphed images which do not have the forensic traces of illegitimate procedures of image editing. [10]

Forensic is similar to biometric analysis as here the forensic quality is linked to the error rates of the morph detection procedure. There are two samples, positive and negative. [10] In which the positive sample is the morphed image and negative sample is the original image. The higher the rate of the error, the higher the forensic quality of the doctored image. The simpler term of forensic analysis means a detailed investigation by means of detecting. [10]

For e.g., in case of the passport checking, the analysis of the printed image is carried out and scanned in the application procedure for the further detailed investigation or off-line authenticity. [11]

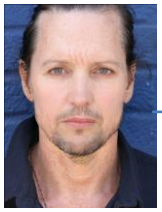
There are issues/ attacks wherein the attackers try to divert various forensic procedures and dissemble their counter-forensic procedures for which methods are created to reveal such atrocious attacks. [11][12][13]

In our experiment, we have used detector methods like Keypoints, GoogleNet-based, Benford, Degradation, VHH 19-based (naïve and multi-class) and High-Dim LBP. [10] Each of the detector methods was used in different methods of morphing. For e.g., for a morphed image created through the combined method, we have used these detector methods as a forensic analysis to detect a score, which gives us the rate as to how a particular morph is detected and evaluated.

We have taken the 3 methods Keypoints, Degradation and Benford OvGU and we have decided a threshold score of 0.5. That means if a morphed image has been detected with a score of more than 0.5, then that particular image is Morphed and if it's not then that image is not morphed. Basically, a morphed image when it is created is so much precise, that it

is not possible to detect it by simpler means. It has to be detected to prevent forensic attacks.

Hence we have come up with an experiment, in which the morphed image which we have created is been detected by all the three methods of detectors and then based on the threshold value, the evaluation has been made. In simpler terms, it helps to identify if a given image is morphed or no. There were so many detector methods, hence, we used three of them and then made it in a tabular format so that we can classify them in a proper way and know which detector can be the best.



Detector Methods	Score	Evaluation
Keypoints	0.56	Morph
Benford	0.99	Morph
Degradation	0.53	Morph

Morph A (combined)

Table 1: The detector methods are being analyzed with their respective scores for a morphed image created by the combined method

Detector Methods	Score	Evaluation
Keypoints	0.25	NoMorph
Benford	0.99	Morph
Degradation	0.50	Morph

Morph A (splicing)

Table 2: The detector methods are being analyzed with their respective scores for a morphed image created by splicing method

In Table 1, all the scores were more than 0.5 thresholds, are being evaluated as morphed and in table 2 it shows, the scores less than 0.5 are evaluated as NoMorph, which means even though the image could be a morphed image then detector cannot detect them as morphed.

Like this, we have calculated the scores and evaluation of all the morphed images created by morphing techniques and categorized them.

For the above, the tabular format is only for a single image, but for the bulk morphed images that we have created, to be precise, we have found the error rate and the accuracy of the detectors with respect to the original images from the DEFACTO so that we can get a clear picture of which method gives an average best score from the rest. **(see** Table 3,4,5)

5. ANALYSIS AND VISUAL CLASSIFICATION

As we were able to morph the original images from the 'DEFACTO' dataset [10] using the ANANAS web service (AWS). And by means of combined and splicing method, we were able to create them, the morphs which are created by us using AWS is being visually compared with the already morphed images of DEFACTO database.

This subjective visual comparison shows us the way the morphs are been generated and by what means they are indifferent to each other. The analysis of the entire dataset is being carried out by comparing it visually by checking their eyesight, their transitions, facial structure which includes prolonged face, oval face, etc., or the skin texture, jawline, double chin et al. If there are no visual differences, there could be a severe loss of quality loss of the images because of its compression. The morphs may also contain shadow artifacts. There would be some morphed images that are elongated or oval in shape than the other, or else the skin color would differ.

There are some examples shown below post the visual comparison.
The comparison of the morphed image created by the combined and splicing method through AWS with the morphed image from the DEFACTO database. This is only to analyze which morphed image is precise or can be used for the task.

Morph B (combined) ANANAS web service

Morphed Image (DEFACTO Dataset)

Fig 10: Morph B is the morphed image created by ANANAS web service using the combined method of morphing, whereas the other image is the morphed image from the DEFACTO dataset

Morph B (splicing)
ANANAS web service

Morphed Image
(DEFACTO Dataset)

Fig 11: Morph B is the morphed image created by ANANAS web service using splicing method of morphing, whereas the other image is the morphed image from the DEFACTO dataset

These comparisons are being 'visually' analyzed [6] for both the morphs created by combined and splicing and both the images are created morphed using different methods.

(a) Morph B (combined)
ANANAS web service

(b) Morphed Image
(DEFACTO Dataset)

Fig 12: Differences between the two morphed images from ANANAS web service and DEFACTO dataset (eye)

(a) Morph B (combined)
ANANAS web service

(b) Morphed Image
(DEFACTO Dataset)

Fig 13: Differences between the two morphed images from ANANAS web service and DEFACTO dataset (lips)

After visually comparing both the morphed images created by the combined method to the morphed image by DEFACTO, here are some of the differences:

- The difference in Eye color and ghostly shadow of pupil
- More wrinkles are seen on the image created by ANANAS web service
- More shadow artifacts are seen on the image created by ANANAS web service
- Lip is smaller and less filling in Morph B than the morphed image by DEFACTO dataset
- From figure 10, it is seen that the face is round than the other
- From figure 11, it is seen that the Morph B image is compressed and shorter than the other
- Both the images in figure 10 and 11, have different textures and skin complexions

6. ERROR RATE AND ACCURACY OF THE DETECTORS

Error rates are calculated as it is crucial to measure them for the various forensic techniques as the probative value of it is inseparably connected to the rates at which the testers make errors. [10]

We have taken around 412 genuine images from the 'DEFACTO' dataset [8] and morphed them using ANANAS web services. We used combined, splicing and complete

methods [10][9] of morphing on these images. We selected the combined morphed images and evaluated the efficiency of three detectors methods used such as OvGu Keypoints, OvGu Benford and OvGu degradation. OvGu Keypoints gave 376 images as morphed and 36 images as non-morphed. Similarly, Benford and Degradation gave 406 and 362 images as morphed; and 6 and 50 images as non-morphed. The detectors gave a score between 0 and 1. The threshold score is 0.5; any image having a score less than 0.5 is evaluated as morphed and images having a score of 0.5 or more than 0.5 are evaluated as not morphed. Although, all the images were already morphed the detectors still detected some images as non-morphed. Here we can clearly see some discrepancies in the performance of the detectors. When we applied these detectors on images that were morphed using the splicing method, we got similar results. We analyzed these results and found that detector OvGu Benford has performed better than the other two detectors. The efficiency of detector OvGu degradation is poor.

Our task is to evaluate the performance of the detectors on Genuine images from the DEFACTO dataset as well. We have taken more than 200 genuine images from the 'DEFACTO' dataset and using the three detectors we got the result as shown in the Tables 3,4,5. We have used the same threshold score of 0.5 as we have used before. We can see the performance of the OvGU Benford detector is the best and OvGu degradation has performed poorly. Out of 200 genuine images, OvGu Degradation has evaluated 199 images as morphed and 1 image as not morphed.

To compare the performance of the three detectors on morphed images (created using images from DEFACTO dataset) and genuine images (from DEFACTO dataset), we have created a confusion matrix (Fig 14) as shown in the Tables 3,4,5. While creating the confusion matrix, we have taken two parameters in consideration such as 'reality' and 'outcome'. The first parameter 'reality' defines the status of the image that is if the image is genuine or morphed. The second parameter states the result or outcome we get after evaluating that particular image using different detector methods. Using these parameters, we get the values True Positive, False Negative, False Positive and True Negative. [10][1][3] And using these values, we have calculated the performance of detectors.

6.1 Error rate

Error rate [4][10] is defined as the fraction of images where the prediction is wrong; which means where the images are incorrectly classified as morphed or not morphed. The best error rate is 0 and the worst error rate is 1.

To calculate the error rate, we need True Positive, False Negative, False Positive and True Negative. [10][1][3] Below is the formula for error rate. The basic structure of a confusion matrix is shown in Fig 14.

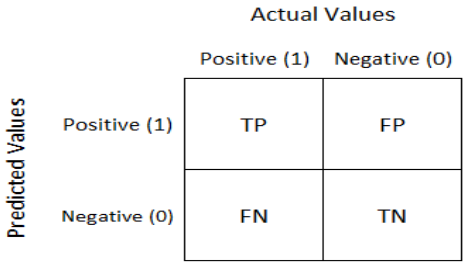


Fig 14: Basic structure of a confusion matrix

Image Source:
https://www.google.com/search?q=confusion+matrix+images

Error Rate = fraction of incorrectly classified images as morphed or not morphed / total number of images

$$= (FP + FN)/ (TP + TN + FN + FP)$$

...[4][10]

whereas,

- TP = True Positive is the correct positive prediction in the dataset
- FP = False Positive is the incorrect positive prediction in the dataset
- TN = True Negative is the correct negative prediction in the dataset
- FN = False Negative is the incorrect negative prediction in the dataset

Keypoints	Morphed	Not Morphed
Morphed	376	118
Not morphed	36	82

Table 3: Confusion matrix for Keypoints

The above table is the confusion matrix for the detector OvGu Keypoints. We see that 376 morphed images are correctly predicted as morphed by the detector which makes it TP (True Positive) in this case. 118 morphed images have been incorrectly predicted as not morphed, hence it is FP (False Positive). The detector has predicted 36 genuine images as morphed which makes it FN (False Negative). 82 genuine images have been correctly predicted as not morphed by the detector; making it TN (True Negative). By

using the values of TP, FP, FN, and TN, we have calculated the error rate of this detector and got the error rate value like 0.25 or 25%.

Similarly, we calculated the error rates of the other two detectors with the help of below shown confusion matrixes of these detectors. We have found the error rates of Benford and degradation at 0.009 or 9% and 0.4 or 40% respectively. These values show that OvGu Benford has the least error rate and OvGu Degradation has the highest error rate.

Benford	**Morphed**	**Not Morphed**
Morphed	406	0
Not morphed	6	200

Table 4: Confusion matrix Benford

Degradation	**Morphed**	**Not Morphed**
Morphed	362	199
Not morphed	50	1

Table 5: Confusion matrix for Degradation

Accuracy of Detectors	**Keypoints**	**Benford**	**Degradation**
Calculation	458/612	606/612	363/612
ACC	**74%**	**99%**	**59%**

Table 6: Percentile of Accuracy of Detectors

Error Rate of detectors is represented in the below figure.

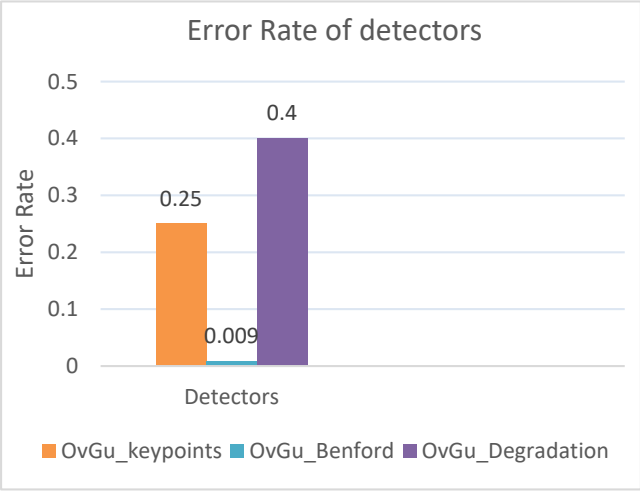


Fig 15. Error Rate of detectors

6.2 Accuracy

We have also calculated the accuracy of all the detectors using the confusion matrix. Accuracy [1] can be described as the fraction of images where the prediction is right that means when the images are correctly classified as morphed or not morphed.

The main goal of finding the accuracy of the detectors was to find the correctness and precision of the results shown by the detectors on the morphed as well as genuine images.

Accuracy = fraction of correctly classified images as morphed or not morphed / total number of images

$$= (TP + TN) / (TP + TN + FN + FP)$$

...[1]

With the help of the confusion matrix, we have calculated the accuracy of the three detectors such as 0.74, 0.99 and 0.59 for Keypoints, Benford, and Degradation respectively. We see that detector OvGu Benford has the highest accuracy whereas detector OvGu Degradation is the least accurate. On the basis of 'Error rate' [9] and 'Accuracy' [1] of all the three detectors, we deduced that the performance and efficiency of OvGu Benford are the best and OvGu Degradation has performed poorly and is least efficient.

The Error rate of Detectors	**Keypoints**	**Benford**	**Degradation**
Calculation	**154/612**	**6/612**	**249/612**
ERR	**0.25**	**0.009**	**0.4**

Table 7: Percentile of the Error rate of Detectors

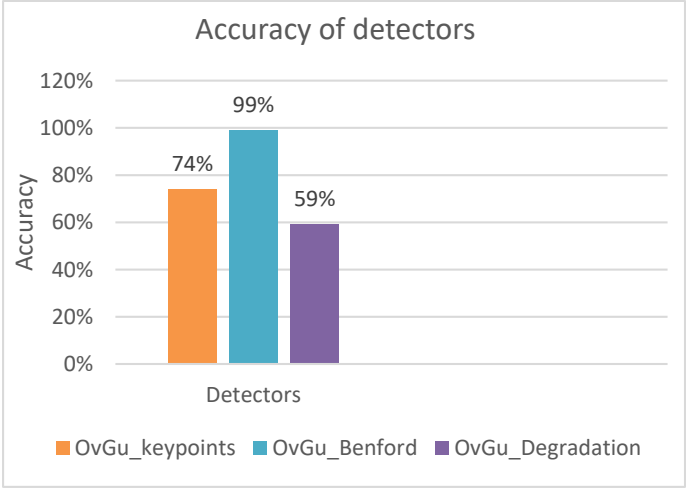


Fig 16: Accuracy of Detectors

6.3 Experiment and outcomes

To inspect the performance of the detectors we have performed the following experiments:

1. Within dataset performance [4]
2. Cross dataset performance [4]

Within dataset performance:

We have compared the performance of detectors within the 'DEFACTO' dataset. We examined the images which were already morphed using the three detectors. We found that the detector Benford has given the highest performance and degradation has performed the least.

Cross-dataset performance:

We also checked the performance of detectors on the morphed image created using ANANAS Web Service and compared that to their performance on morphed images in the 'DEFACTO' dataset. We have shown this performance graphically in the next section.

6.4 Graphical Representation and comparison

We have graphically represented (using histogram, see fig 17) and compared the performance of detectors for morphed images we created using ANANAS web services (AWS) and the images which were already morphed in the DEFACTO dataset. This is to visually compare the datasets so that we would know which

To create the graph, we have taken the average of all the detector scores of the morphed images created by ANANAS web service and morphed images from the 'DEFACTO' dataset.

The threshold score to classify an image as morphed is 0.5 or above. We see that OvGu Degradation has the best performance out of the three detectors and OvGu Benford has least performed. In fact, in the case of morphed images from the 'DEFACTO' dataset, OvGu Benford has not classified any image as morphed resulting in the least performance and efficiency. OvGu Keypoints has also performed well for both types of morphed images.

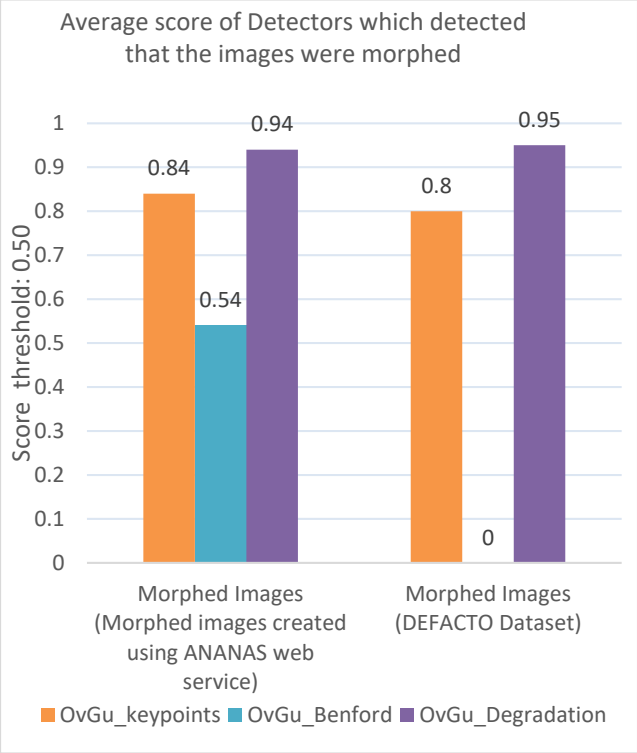


Fig 17. Comparative Analysis of Detectors for morphed images created by AWS and morphed images from the DEFACO dataset

7. CURRENT APPROACH

Various methods have been proposed and used to prevent the forgery of images in 'Digital Image Forensics'.[11] Currently, two main approaches are being used for this such as active approaches and passive approaches. [7] [5]

Active 'forensic' [7] approaches comprise of designing many kinds of watermarks or fingerprints of the image content and inserting them into the digital image. [7] In the initial stage the 'watermarks' [7] on the image is extracted and examined to see if the image has been doctored; if so then where the doctored location is. [7] Although this approach is not very popular because it is not necessary that all images are required to be watermarked. Hence, passive approaches are being widely used as they are independent of any prior information about an image. [7] This approach analyzes a particular clue or pattern in 'image forgeries' [7] that ensures during the creation or modification of the forged image. There are various types of traces defined to classify

an image as tampered/doctored or genuine. Some of these traces are Traces left in image acquisition, Traces left in image storage, Traces left in image editing, etc. [7][11]

Several traces are left by the lenses and sensor when an image is captured is tampered. Using the deviations produced by the camera lens, we can relate an image to a specific device or examine whether an image has tampered. Some parameters of the image like 'radial distortion' [7] also change which makes it easier to detect if the image has been morphed or not and helps in differentiating it with a genuine image.

Traces left in image storage consist of traces which the image acquires while getting compressed once or more than once in JPEG or other formats. The 'JPEG compression' [7] leads to some bias in the images such as quantization error, rounding error, etc. [7] The third type of trace left in images is the one left while editing the image. In copy-paste tampering, an image contains lighting and shadow traces which can be detected using various detecting algorithms [].

Another type of image detection is the 'blind space-splicing detection' [7] in which traces such as brightness and contrast adjustments, scaling and rotation are left after the tampering of images which is treated as a two-class classification problem and is solved with detection method based on DCT coefficient model. [7]

To detect the tampered or doctored images some 'block-based' [2] techniques are also being used. The block-based methods comprise of a feature extraction step to each block [ca1]. These block-based detection methods use robust features such as SVD (Singular Value Decomposition), PCA (Principal Component Analysis); these features are combined and then matched to detect the forgeries [2] in doctored images. Correlation characteristics between segments are also used to detect duplicated regions in a tampered image. [2]

8. FUTURE WORK

As the development of advanced artificial intelligence techniques is rising, a better solution for digital image forensics is proposed by deep learning[ca]. 'Deep learning' [7] approaches such as CNN (Convolutional Neural Network) and deep residual network can be used to revolutionize the field of digital image forensics. [7]
In CNN and deep residual network, features do not need to be manually designed; they are automatically learned. [7] Deep learning-based methods with adequate training datasets, can provide a far-reaching and more understandable explanation for certain forgery processes, compared with contemporary approaches. [7]

CNN is implemented in detecting tampered images by using 'median filtering'. [3] By considering the properties of median filtering, we can detect a forgery in small-size and 'JPEG compressed' [7] images.

The trace in a tampered image left by median filtering is too weak to be detected by the traditional CNN approach. [3] Hence, in contrast to the traditional CNN [ca] models, the first layer of the CNN 'framework' [3] is a filter layer which takes an image as the input and gives 'median filtering residual' [3] of the image as the output. [3] Then the convolutional layers and pooling layers are alternated to acquire the hierarchical representations and then calculate various features for further classification. This method will show significant progress in performance specifically in case of cut-and-paste forgery detection. [3] This approach is also successful in attaining 'better detection accuracy' [3] compared to other approaches.

Forgery in JPEG images consists of recompression of the image. [4] Another approach to detecting tampering in JPEG compressed images is by examining 'inconsistency in quantization table' [7], evaluating 'abnormality of compression artifacts' [4] and checking the 'periodicity characteristics' [4] of JPEG images. Basically, the quantization table is estimated by calculating quantization error minimization [4]. With the quantization table, we can easily detect the block inconsistency and pinpoint the tampered block. Periodicity of artifacts changes when an image is forged; it can be detected by implementing mathematical formulation in the quantization table and also combining periodic features of the image in spatial as well as frequency domains. This method is effective in detecting 'recompression' [4] with either aligned or misaligned boundaries of the image. [4]

As of now, deep learning approaches have not shown much performance gain in digital image forensics.[7] This is due to the fact that currently the structures are learned from image content related systems which results in the dilapidation of performance in several digital image forensics applications. However, deep-learning-based methods are reliable and promising, they are still not mature in digital image forensics. [7]

9. CONCLUSION

We have taken around 200 genuine images from the 'DEFACTO' dataset and morphed these images using ANANAS Web Service. We used three types of morphing techniques such as combined, splicing and complete. [8]
After morphing of the images, we have also performed 'biometric' analysis and 'forensic' analysis [11] on these morphed images. For biometric analysis in our experiment, we have used a Dermalog matcher in ANANAS Web Service. Using this matcher, we have matched each morphed image to its genuine image. The Biometric analysis basically examines biometric features (such as skin, hair, gender, age, etc.) in the images. We have taken a threshold score of 0.8 in the matcher to evaluate if the image is matched to the genuine counterpart. The matcher shows an image to be matching with the genuine one if the score is 0.8 or above. We have also visually examined the morphed images and compared them to their genuine counterparts. We found there are differences in facial features; i.e. variations in skin color and texture, hair color and texture, change of color in eye pupils. etc. However, when we varied the blending parameter alpha from 0.5 to 0.7, 0.6 and so on to see if there is any variation in the result shown by the matcher, we found that with changing the blending parameter the result and evaluation of the matcher vary. The morphed image matched with the genuine image once we increased the value of alpha.

To perform the forensic analysis on the morphed images we generated as well as the image which was already morphed in the DEFACTO dataset we relied on detector methods provided by the ANANAS Web Service. Forensic analysis is basically detecting the 'forensic qualities' like tell-tale traces, used anti-forensics left in an image after it is tampered or forged. [10] We have mainly used three detectors named OvGu Keypoints, OvGu Benford and OvGu Degradation. These detector methods detect the forensic qualities in a morphed image. The detector threshold score is set to 0.5 and an image is evaluated as morphed if it gets a score of 0.5 or above. We have also calculated the error rate and accuracy of the detectors using confusion matrices created using the morphed and genuine images from the dataset as well as the morphed images we created. From the scores, we have deduced that the detector OvGu degradation has the highest error rate and the least accuracy. On the other hand, OvGU Benford has the highest accuracy and least error rate.

In the end, we have presented a synopsis of all our observations and we can conclude that the detector Benford is most efficient and degradation is least efficient.

ACKNOWLEDGMENT

We would like to thank Prof. Andrey Makrushin and Prof. Dr.-Ing. Jana Dittmann of Department of Computer Science Research Group Multimedia and Security Institute of Technical and Business Information Systems, Otto-von-Guericke-University of Magdeburg for their constant guidance, support, and constructive feedback.

AUTHORS

The section from the beginning, that is 'ABSTRACT' till section 5 that is 'ANALYSIS AND VISUAL CLASSIFICATION' is written by Gracy Joseph.

The section which starts from 6 that is 'ERROR RATE AND ACCURACY OF THE DETECTORS' and ends with 'ACKNOWLEDGEMENT' is written by Shweta Pandey.

REFERENCES

[1] Abdullah Alharbi, Wajdi Alhakami, Sami Bourouis, Fatma Najar, and Nizar Bouguila. 2019. Applied Computing and Informatics Inpainting forgery detection using hybrid generative/discriminative approach based on bounded generalized Gaussian mixture model. *Applied Computing and Informatics* xxxx (2019), 1–8. DOI:https://doi.org/10.1016/j.aci.2019.12.001

[2] Abdullah Alharbi, Wajdi Alhakami, Sami Bourouis, Fatma Najar, and Nizar Bouguila. 2020. Inpainting forgery detection using hybrid generative/discriminative approach based on bounded generalized Gaussian mixture model. *Applied Computing and Informatics* xxxx (2020), 1–8. DOI:https://doi.org/10.1016/j.aci.2019.12.001

[3] Jiansheng Chen, Xiangui Kang, Ye Liu, and Z Jane Wang. 2015. Median Filtering Forensics Based on Convolutional Neural Networks. *IEEE Signal Processing Letters* 22, 11 (2015), 1849–1853. DOI:https://doi.org/10.1109/LSP.2015.2438008

[4] Yi-lei Chen and Chiou-ting Hsu. 2011. Detecting Recompression of JPEG Images via Periodicity Analysis of Compression Artifacts for Tampering Detection. *IEEE Transactions on Information Forensics and Security* 6, 2 (2011), 396–406. DOI:https://doi.org/10.1109/TIFS.2011.2106121

[5] Hany Farid. 2009. Image Forgery Detection [. March (2009), 16–25.

[6] Silvan Heller, Luca Rossetto, and Heiko Schuldt. 2018. The PS-Battles Dataset - an Image Collection for Image Manipulation Detection. (April 2018), 3–7. Retrieved from http://arxiv.org/abs/1804.04866

[7] Xiang Lin, Jian Hua Li, Shi Lin Wang, Alan Wee Chung Liew, Feng Cheng, and Xiao Sa Huang. 2018. Recent Advances in Passive Digital Image Security Forensics: A Brief Review. *Engineering* 4, 1 (2018), 29–39. DOI:https://doi.org/10.1016/j.eng.2018.02.008

[8] Gaël Mahfoudi, Badr Tajini, Florent Retraint, Frédéric Morain-Nicolier, Jean Luc Dugelay, and Marc Pic. 2019. Defacto: Image and face manipulation dataset. *European Signal Processing Conference* 2019-Septe, (2019). DOI:https://doi.org/10.23919/EUSIPCO.2019.8903181

[9] Andrey Makrushin and Andreas Wolf. 2018. An overview of recent advances in assessing and mitigating the face morphing attack. *European Signal Processing Conference* 2018-Septe, (2018), 1017–1021. DOI:https://doi.org/10.23919/EUSIPCO.2018.8553599

[10] Tom Neubert, Andrey Makrushin, Mario Hildebrandt, Christian Kraetzer,

and Jana Dittmann. 2018. Extended StirTrace benchmarking of biometric and forensic qualities of morphed face images. *IET Biometrics* 7, 4 (2018), 325–332. DOI:https://doi.org/10.1049/iet-bmt.2017.0147

[11] Ulrich Scherhag, Luca Debiasi, Christian Rathgeb, Christoph Busch, and Andreas Uhl. 2019. Detection of Face Morphing Attacks Based on PRNU Analysis. *IEEE Transactions on Biometrics, Behavior, and Identity Science* 1, 4 (October 2019), 302–317. DOI:https://doi.org/10.1109/TBIOM.2019.2942395

[12] Ulrich Scherhag, Andreas Nautsch, Christian Rathgeb, Marta Gomez-Barrero, Raymond N.J. Veldhuis, Luuk Spreeuwers, Maikel Schils, Davide Maltoni, Patrick Grother, Sebastien Marcel, Ralph Breithaupt, Raghavendra Ramachandra, and Christoph Busch. 2017. Biometric Systems under Morphing Attacks: Assessment of Morphing Techniques and Vulnerability Reporting. *Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI)* 1 (2017), 1–11. DOI:https://doi.org/10.23919/BIOSIG.2017.8053499

[13] Ulrich Scherhag, Christian Rathgeb, Johannes Merkle, Ralph Breithaupt, and Christoph Busch. 2019. Face Recognition Systems under Morphing Attacks: A Survey. *IEEE Access* 7, (2019), 23012–23026. DOI:https://doi.org/10.1109/ACCESS.2019.2899367

[14] Bhushan Zope and Soniya B. Zope. 2017. A Survey of Morphing Techniques. *International Journal of Advanced engineering, Management and Science* 3, 2 (2017), 81–87. DOI:https://doi.org/10.24001/ijaems.3.2.15