

# Cascading multiple LDA classifiers for facial recognition

M. N. van Dijk (s1999478)  
m.n.vandijk@student.utwente.nl  
Enschede, June 26th 2020

**Abstract**—A classical approach to face recognition uses dimensionality reduction techniques to describe faces. A face can be projected onto a feature space that spans the significant variations among known face images. To design a classifier, a normal Gaussian distribution within this feature space is assumed. However, it was found assuming this kind of distribution is not optimal for facial recognition. In this thesis research will be done on subspace learning to deal with the small amount of data which lies outside the Gaussian distribution, but still has to be recognized by a classifier. To do this, a new way of face classification is proposed using classical facial recognition methods to cascade multiple classifiers. By cascading multiple classifiers, subsets initially not recognized by the first classifier can be classified. This thesis will investigate whether cascading multiple LDA classifiers can be beneficial for facial recognition and what the effect of several parameters is on the performance of this classification system. This is done by looking at both Authentics-Imposter Distribution curves, as well as ROC curves. Because the results show no improvements, it was concluded the sample clusters outside of the assumed distribution have to be modeled in a more accurate way.

## I. INTRODUCTION

Facial recognition is becoming increasingly popular over the past few years. It is being used by a large variety of applications used in the security, health care and marketing industry. In the past research has been done on facial recognition using algorithmic methods, for example using dimensionality reduction techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to create classifiers. These classifiers assume a Gaussian distribution of the samples within the dataset, which is usually not the case for datasets of faces.

For the past couple of years artificial intelligence centric methods for facial recognition have become increasingly popular as it yields better results than the classical face recognition methods. A disadvantage of using these methods to do face recognition is that it is not as efficient; it requires a copious amount of training data and an abundance of computational power.

A problem for classifiers in classical facial recognition is that its classifiers assume a Gaussian distribution and the classifiers struggle to recognize samples lying outside of this distribution. These samples will be referred to as 'outliers'. If these outliers can be described with a different kind of model, classifiers can be made specifically for these samples. In this thesis, a new way of facial classification is proposed, which cascades multiple classifiers. These classifiers will be made

using the classical facial recognition methods. When cascading multiple classifiers the samples initially not recognized by the first classifier can be classified by later classifiers. A method will be chosen on how to find outliers and describe them with a different model. New classifiers will be made based on this model.

This thesis will investigate whether cascading multiple LDA classifiers can be beneficial for facial recognition and what the effect of several parameters is on the performance of this classification system. The remaining of this paper is organised as follows. In section II an overview will be given of related work. In section III the method used to answer the research question will be discussed. To assess the quality of the method, ROC curves and Authentics-Imposter Distribution curves are used. These will be shown and discussed in section IV Experiments and Results. The Discussion in section V will discuss these outcomes. Section VI will conclude this thesis and provide questions for further research.

## II. RELATED WORK

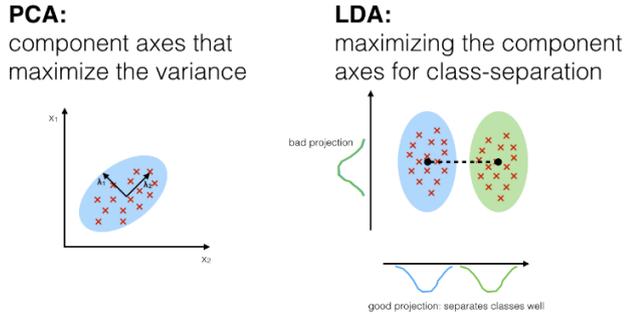
In the classical approach to facial recognition a face image is expressed as a multi-dimensional vector. These vectors can be projected onto a low dimensional feature space using dimensionality reduction techniques such as PCA and LDA. Because images of faces are similar in overall configuration, they can be described by a relatively low dimensional subspace, called the 'face space'. Image pairs belonging to the same person are generally called authentics. Image pairs belonging to the different people are generally called imposters. Pictures from the same person can be referred to as samples from the same class.

In 1987, L. Sirovich and M. Kirby used PCA to create 'eigenfaces' to characterize faces. [1] In 1991, M. Turk and A. Pentland introduced the use of the subspace spanned by eigenfaces (the face space) for recognition; classifying the face by comparing its position in face space to the positions of other faces. Assigning an image to the label of the closest point in the learning set is called nearest-neighbor classification. This assumes a Gaussian distribution of the multidimensional data. Turk and Pentland mention there is no reason to assume any particular distribution.

In 1996, Belhumeur points out that from a discrimination or classification standpoint, using PCA is not optimal. [2] This is because PCA maximizes the total scatter across all samples, including both the between-class scatter and the within-class

scatter. The between-class scatter is the scatter between different clusters of samples. The within-class scatter is the scatter within the samples belonging to the same class or person. This type of scatter is not needed for discrimination purposes and preferably as small as possible to prevent overlap. LDA, in contrast to PCA, aims to find the combination of features that best separates between classes. This is done by maximizing the component axes for class-separation (between-class scatter), while minimizing the within-class scatter. [3] An example for two-dimensional data can be seen in figure 1.

Fig. 1. PCA maximizes the overall scatter, while LDA maximizes the scatter between classes [4]



#### A. LDA log likelihood ratio

This research uses the expression for LDA log likelihood ratio one-to-one classifiers for biometric comparison between single reference and test samples, derived by L. Spreeuwers. [5] This classifier describes the distance between two images as a likelihood score of them being in the same class. In this research the number of PCA and LDA components will be used to change the performance of the classifier.

The number of PCA and LDA components are commonly chosen to be in the range of 50-200 and between 15-50 respectively.

#### B. Manifolds

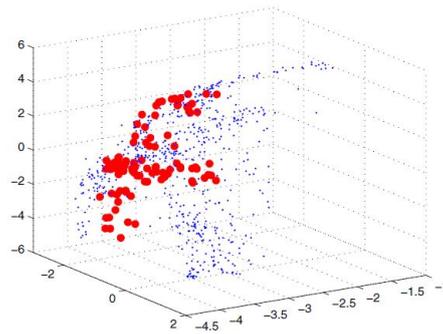
Factors such as illumination and face angle can change depending on the circumstances in which a picture is taken. These factors change a picture's place in face space. Pictures of faces taken under similar circumstances form clusters in face space. These clusters are called manifolds and can be modelled as multiple Gaussian densities using simple PCA/LDA classifiers. Research by A. Patel has been done on manifold learning using PCA to discover its principal axes from face space data. [6] In figure 2 an example of the manifold distribution of samples for facial recognition is shown. [7]

#### C. Classifier Combination methods

It has been observed that the accuracy of pattern classification methods can be improved by classifier fusion. Classifier combination schemes can be grouped based on the level at which they operate: feature level and score level.

In feature level combinations, features of each classifier are combined to form a joint feature vector. These feature

Fig. 2. Multiple samples of one face projected on the first three principal components of face space [7] The red dots indicate samples in the test set while the blue dots indicate samples in the training set



vectors are used to classify the set. The increased number of feature vectors will require a large training set and complex classification schemes. Score level combinations use outputs of the classifier for combination and require no knowledge of the internal structure of classifiers and their feature vectors. In this form of combination some information is lost, but it does have a lower complexity than feature level combinations. [8] [9]

Classifiers providing outputs in the form of likelihoods are commonly referred to as soft output classifiers. These classifiers can be combined using score level fusion methods such as Bayesian fusion methods, fuzzy integrals or the Demspter-Shaffer combination. [10]

P. Viola and M. Jones have proposed a boosted cascade of classifiers, in which the form of the detection process is that of a degenerate decision tree, called a cascade. A positive result from the first classifier triggers the evaluation of a second classifier, which can again trigger a third. The thresholds for each classifier are adjusted to minimize the number of false negatives. [11] Their cascade tries to reject as many negatives possible at the earliest stage possible. The cascade is designed to minimize the expected number of evaluated features.

### III. PROPOSED METHOD

Designing this multiple classifier system, multiple decisions were made. A straightforward approach was chosen in order to find outliers and design the next classifier.

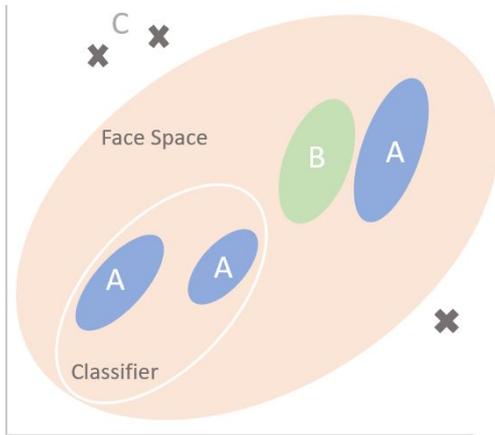
The first classifier was made using a training set. After designing the first classifier, the outliers from this training set have to be found. For this, it is important to see in what different ways these outliers can manifest themselves. This will be discussed in section III-A. The decisions made regarding extracting outliers to form new training sets will be explained in III-B. The structure and operation of the final classification system is discussed in section III-C.

#### A. Outlier manifestations

The implementation for outlier extraction used in this thesis finds pairs of samples that are from the same class, but are not close to each other in feature space. When two samples from the same class are not close together in face space, this can be

caused by multiple situations, which are illustrated in figure 3. One situation could be that one class covers two or more subspaces in face space. These subspaces could be both inside and outside the Gaussian distribution of face space assumed by the classifier. If the subspaces are farther apart from each other, samples from the same class but a different subspace score a low likelihood ratio. Another situation could be one class lying entirely outside of the Gaussian distribution of face space assumed by the classifier. This could mean samples in the same subspace still get a low likelihood score. Wrong samples caused by, for example, a bad picture will likely not be close to any other samples from the same class, or will even lie outside the actual face space. It is not possible to design a classifier for these wrong samples.

Fig. 3. Simplified 2D visualisation of how outliers can manifest in face space. The light blue circle shows the Gaussian distribution assumed by the classifier. A. shows one class covering multiple subspaces in face space, both inside and outside the assumed distribution B. shows another class entirely outside of the assumed distribution C. the crosses show pictures not belonging to any subspace



### B. Finding Outliers

In figure 4 the main approach is shown for the generation of two classifiers from one set of pictures. In this approach to finding outliers, only the likelihood score of the sample to the samples in the same class is taken into consideration. This is done because of easy implementation; no internal knowledge about the classifier was needed. The classifiers are made using a previously set number of PCA and LDA components and the training set. The LDA log likelihood of each individual picture to each other picture in the data set is determined using the found classifier. Using all scores for non-matching picture pairs and a False Accept Rate (FAR), a threshold for the classifier is determined. This threshold will be used for the implementation of the classifier in the final classification system, elaborated in section III-C. It will also be used to extract outliers. Each sample is compared to all other samples from the same class. In the case of 20 samples per class, 19 pairs are compared for each picture, resulting in 380 pairs. All scores for matching pairs are compared to the set threshold. If a pair scores below the threshold, it will be saved. After comparing all scores, all falsely rejected pairs

for the training set are saved. It was chosen to extract only the samples occurring more than a predetermined amount,  $C$  times to the list of outliers.  $C$  can be varied to change the performance of the classifier.

It was decided to design the outlier extracting process this way because it is a simple yet effective way to extract the samples having a bigger distance to the other samples of the same class. Other, more complex options for extracting these samples may have yielded better results. Some of these methods will be elaborated in discussion.

The found outlier samples will be used to train the next classifier. After this all above-mentioned steps are repeated. For the last classifier, the outliers do not have to be determined since no new training set has to be formed.

### C. Classification system

Classifiers were combined on a score level. This requires no knowledge of the internal structure of the classifiers, which makes the implementation simple. The found classifiers are cascaded in a similar way as done by P. Viola and M. Jones as described in related work. For a pair of samples, the first classifier generates a likelihood score. If this score is above the threshold, the score will be accepted and 'boosted'. This boosting is done by adding a multiple of a chosen constant  $K$  to the score. If the score is below the threshold, the score generated the second classifier will be used and compared by the second threshold, and so on. If the pair is not accepted by any of the classifiers, the score generated by the last classifier will be assigned to the pair. Since the last classifier is not boosted, this is usually a very low score. In general the pair will not be accepted, even when the threshold is chosen to be very low. This cascading system is shown in figure 5.

This classification system can be extended for more than two classifiers. The score added for each level will then be  $K(n-1)$ ,  $n$  being the level of the classifier starting at 1 for the last classifier.

### D. Parameters

To generate different results, different parameters can be chosen. The number of PCA and LDA components influence the performance of the classifier. A higher number of PCA components  $p$  will improve the description of the dataset. When too many PCA components are used, the dataset will be over-described making the description of the overall data less accurate. A higher number of LDA components  $l$  will provide better separation between classes.

A classifier dividing or boosting constant  $K$  also has to be chosen. If  $K$  is bigger, the first classifier will have a bigger influence on the outcome.

The maximum count  $C$  is used to train the classifiers and has an influence on the number of outliers which are made out of falsely rejected pairs. If an image occurs more than  $C$  times in the falsely rejected pairs, then it will be added to the set of outliers used to make the next classifier.

Next to this a False Accept Rate (FAR) is chosen to determine a threshold for each of the classifiers. The FAR refers to the ratio of the number of pairs falsely accepted to the total number of imposter pairs.

Fig. 4. Diagram showing the generation of two classifiers from one dataset for training

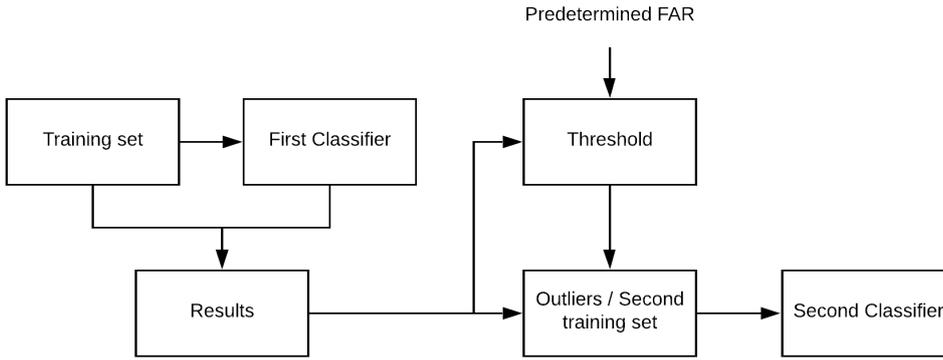
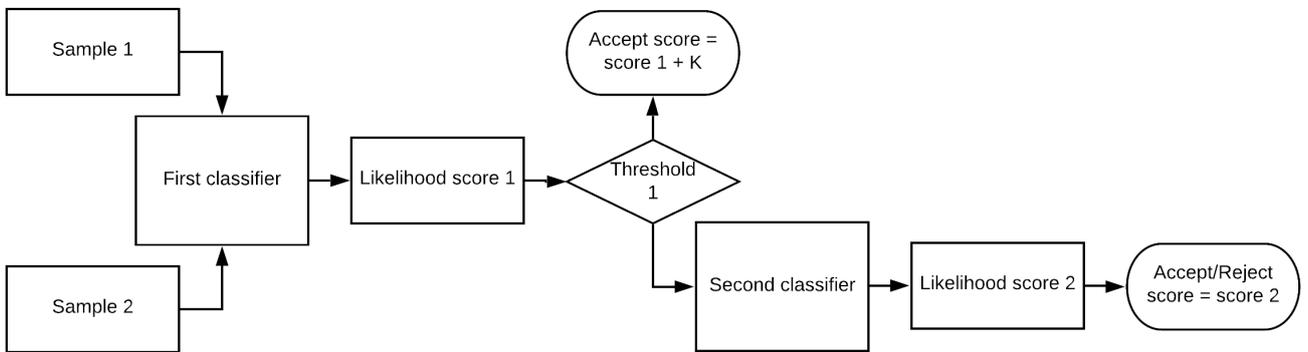


Fig. 5. Diagram showing the classification of test data using two classifiers



### E. Quality assessment

To compare and verify the performance of different classifier structures, two different methods will be used.

1) *Authentic-Imposter Distribution Curve*: The Authentic-Imposter Distribution Curve shows a normalized distribution curve for all authentic scores plotted next to a curve for all imposter scores. This can be used to compare the scores for imposter and authentic pairs. If the distribution curves have more overlap, it is harder to distinguish the imposter pairs from the authentic pairs by scores. If the distribution curves have no overlap an error rate equal to 0% can be achieved choosing an appropriate threshold. [12]

2) *ROC Curve*: The Receiver Operation Characteristic curve illustrates the performance probabilities generated by varying the threshold. This is done by calculating the True Match Rate (TMR), which is equal to the proportion of authentic pairs that have a score above the determined threshold. The threshold is varied by varying the False Accept Rate (FAR). If the curve is more convex to the left, the performance of the classifier improves. To express this quality numerically, the Equal Error Rate (EER) can be calculated. The EER is equal to FAR when  $FAR = 1 - TMR$ . This can be illustrated as the intersection between the ROC curve and a straight line from the upper left corner to the right bottom corner of the figure.

Next to the FAR, the False Reject Rate (FRR) will also be used to evaluate a classifiers performance. The FRR refers to the ratio of the number of pairs falsely rejected to the total number of authentic pairs. [12]

3) *Threshold*: In general a lower threshold yields higher detection rates and higher false positive rates. [11] The number of recognized faces of the facial recognition system increases as the threshold decreases. A nearly perfect recognition could be achieved for a very high threshold, at the cost of many images being rejected as unknown. The trade-off between rejection rate and recognition accuracy is different for different applications of facial recognition. It would be ideal to set the threshold low, in a way few known face images are rejected as unknown while still detecting the incorrect classifications. [13]

## IV. EXPERIMENTS AND RESULTS

### A. Software and dataset

MATLAB is a widely used computing environment used by educational and research organisations around the world. As code for the LDA log likelihood ratio one-to-one classifier has been made available for MATLAB, this software is used to build and test the classifier structure. The version used for this research is MATLAB R2020a. The dataset used is part of the

Face Recognition Grand Challenge (FRGC) [14] and contains 5580 pictures of 297 different people, 20 pictures per person.

The number of pictures used for training and testing have to be chosen. In this research the choice was made to use 139 classes for training and the remaining 140 classes for testing. If a larger training set was used the formed classifiers would be better, but at the expense of the test results. If a bigger test set was chosen the test results would be more accurate but the classifiers have to be formed on a smaller dataset. To even out this problem a distribution of 1:1 was chosen.

### B. Experiment

To research whether cascading classifiers can be beneficial for facial recognition the ROC curve for cascaded classifiers will be plotted and compared to the ROC curve for a single classifier.

Using different values of FAR,  $K$ ,  $C$ ,  $p$  and  $l$ , different results will be obtained. The following sections the effects of each of these parameters will be investigated.

The results for cascading classifiers will be discussed in section IV-C. Then, in section IV-D and IV-E the effect of different numbers of PCA and LDA components will be investigated. In section IV-F the effect of changing  $K$  will be shown. In section IV-G and IV-H the effects of different  $C$  training FAR will be shown.

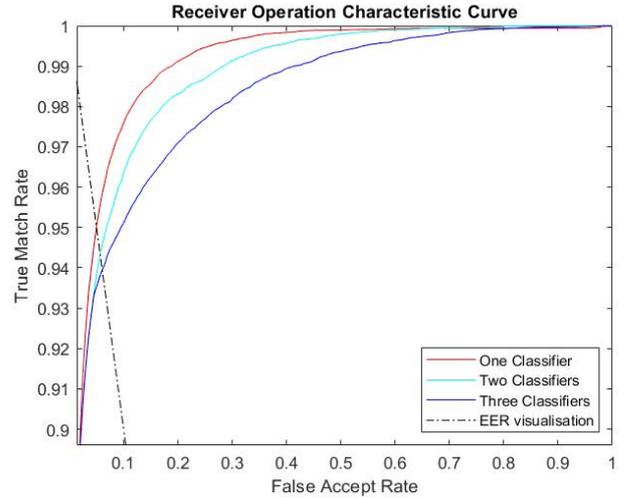
### C. Results for multiple classifiers

A conventional  $FAR = 1\%$  will be used to determine the threshold when training the classifiers.

For  $p = 70$  and  $l = 65$  exactly the same ROC curve was found for one classifier as for multiple classifiers, resulting in an EER of 3%. After that, minimal values  $p = 25$  and  $l = 10$  were chosen for the number of PCA and LDA components. This was done because for these minimal values the first classifier will not function as well, so more falsely rejected pairs are found which will produce outliers. To form a classifier, it is necessary to have sufficient data to base the classifier on. When the number of PCA and LDA components were chosen too high, few to none outliers were found, making it not possible to design a second classifier. To be able to see the effect of cascading classifiers in a small dataset, it is necessary to choose a lower number of PCA and LDA components.

To be able to see the effect of using multiple classifiers, the ROC curve for multiple classifiers was made. The results are shown in figure 6. As can be seen, the graph is less convex when more classifiers are used, indicating a worse performance. This could be because of the simplifications used to determine the outlier set, or because the distribution of the outliers is too complex to fit one classifier. The EER for one, two and three classifiers were found to be 5.0%, 5.5% and 6.0% respectively. For very high FAR, the TMR for multiple classifiers was found to be slightly higher than for one classifier. This indicates that the cascade of classifiers recognizes slightly more faces than one classifier when a very low threshold is chosen.

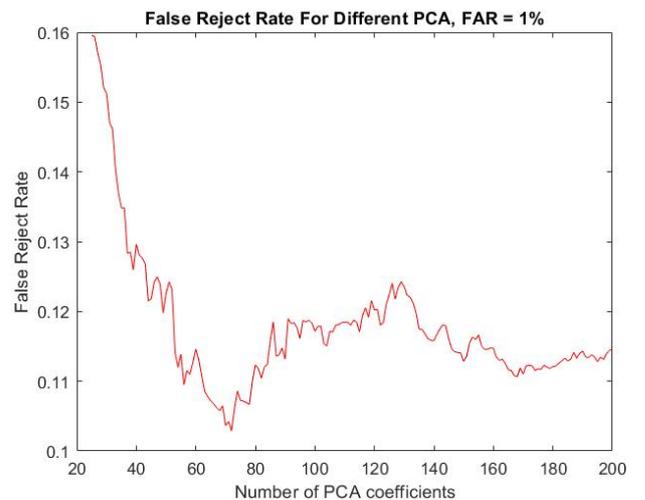
Fig. 6. False Non-Match Rate for different number of classifiers



### D. Number of PCA components

The FRR was calculated for different numbers of PCA components to find the optimal number of PCA components for a single classifier. This was done using the entire testset, containing 140 classes. The False Accept Rate was set to 1%. The result in figure 7 shows the optimal number of PCA components is 72.

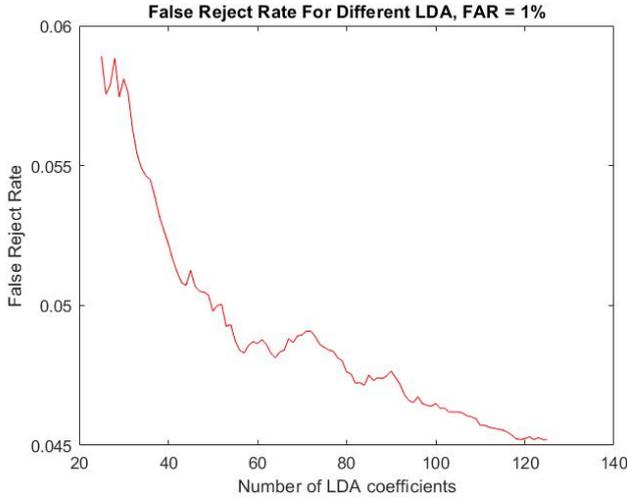
Fig. 7. False Reject Rate for different number of PCA components



### E. Number of LDA components

To find the number of LDA components that have to be used to form a well performing classifier, the False Reject Rate was calculated for different numbers of LDA components while the False Match Rate was set to 1% and the number of PCA components set to 150. As the number of LDA components has to be smaller than the number of PCA components, the results were plotted up to 120 components. The result in figure 8 shows the results improve as the number of LDA components increase.

Fig. 8. False Reject Rate for different number of LDA components



*F. The effect of K*

In this research it was chosen to determine a value for  $K$  using the authentic/imposter distribution curve. The value will be chosen so the matched pairs by the first classifier will always have a higher outcome than the matched pairs of the next classifier. This is done because the first classifier is based on the entire dataset and thus will be best for classifying most samples. It can be seen in the authentic/imposter distribution curve as zero overlap between the different classifier peaks.

For different  $K$ , the conditions  $p = 25$ ,  $l = 10$ ,  $FAR = 1\%$  and  $C = 3$  were used. Looking at the Authentics-Imposter Distribution Curve the effect of a different  $K$  can be seen in figure 9. Each classifier is seen in a peak in the scores, with the rightmost peak having the highest scores being the first classifier. If  $K$  is chosen to be smaller, the peaks move closer together. The chosen values of  $K$  did not influence the ROC curve nor the EER of the final classifier system, but it is expected that for smaller  $K$  the classifiers will overlap even more and this will effect the ROC curve. The value  $K = 20$  was chosen, as for this value the classifier values do not overlap, but the gap between the different classifiers is not too big. This is important as scores accepted by the second or third classifier which are less boosted or not boosted still have to be higher than the threshold of the overall system.

*G. The effect of C*

To choose  $C$ , the conditions  $p = 25$ ,  $l = 10$ ,  $FAR = 1\%$  and  $K = 20$  were used. The effect of changing  $C$  on the ROC curves for three classifiers is shown in figure 10. The EER values are given in table I. The ROC curve converges quicker for lower values of  $C$ . The EER also decreases for lower values of  $C$ , showing a better performance for these values. Lower values of  $C$  work better as this makes the training set for the next classifier bigger, so the second classifier can better describe the data.

Fig. 9. Authentics-Imposter Distribution Curve for three classifiers, top:  $K = 10$ , bottom:  $K = 20$

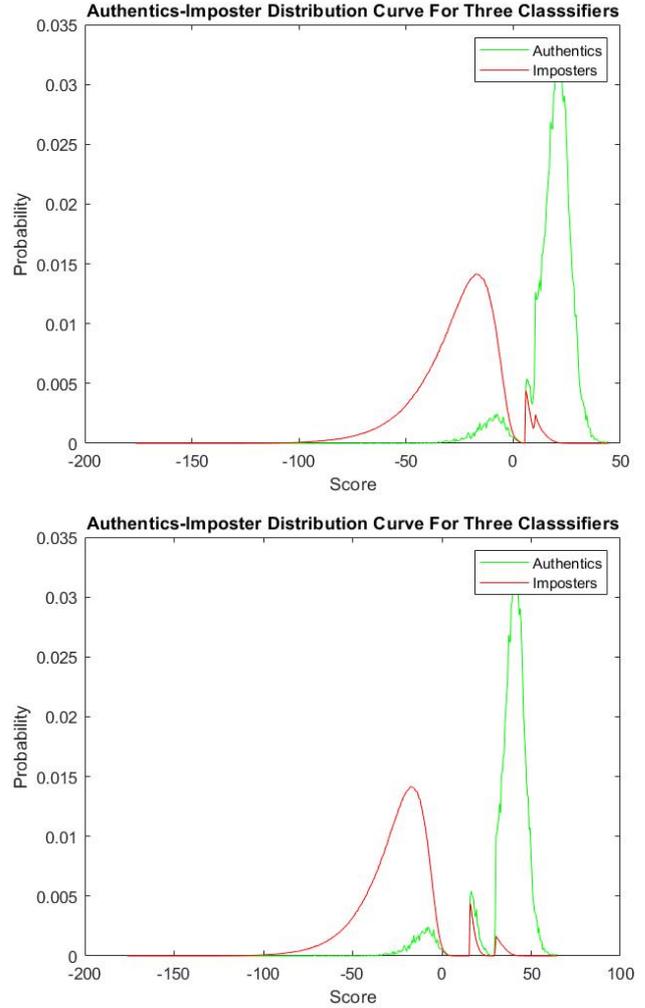


Fig. 10. Receiver Operation Characteristic curve for different values of  $C$  for three classifiers

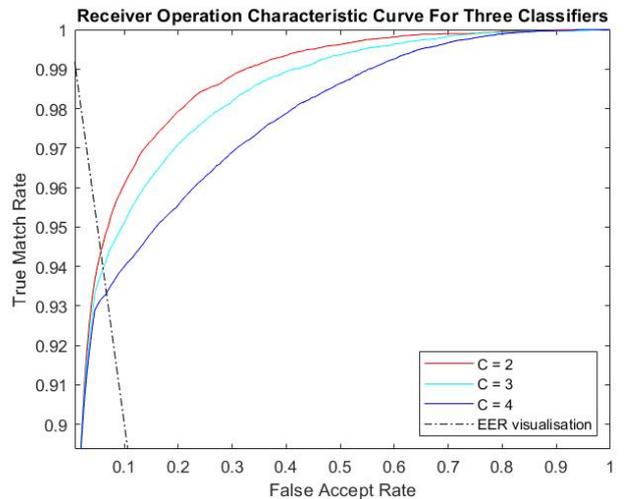


TABLE I

EER VALUES FOR DIFFERENT NUMBER OF CLASSIFIERS AND DIFFERENT C

	C = 2	C = 3	C = 4
2 classifiers	5.4%	5.7%	5.9%
3 classifiers	5.6%	6.1%	6.7%

TABLE II

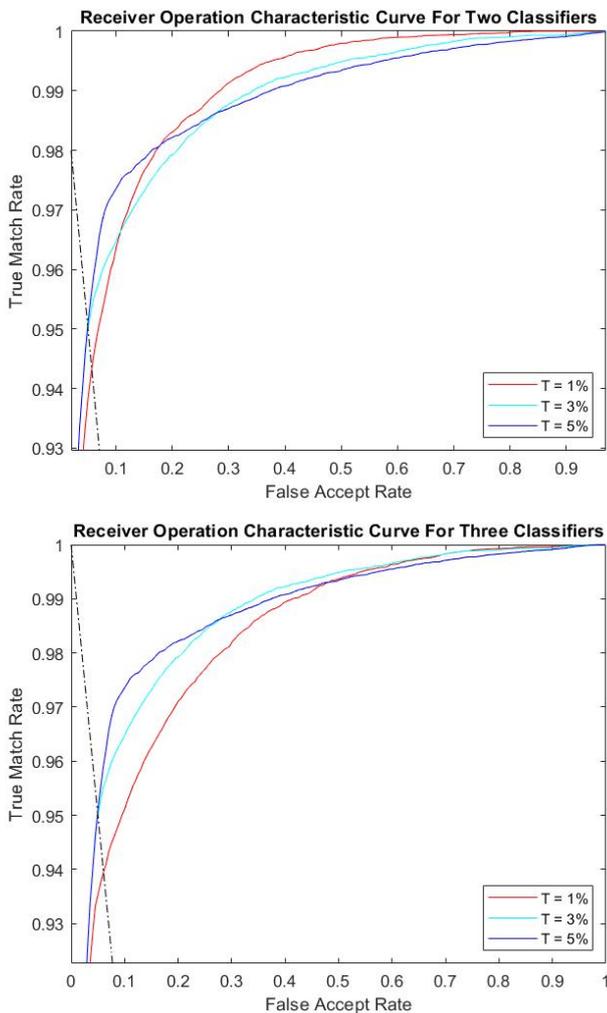
EER VALUES FOR DIFFERENT NUMBER OF CLASSIFIERS AND DIFFERENT FAR

	FAR = 1%	FAR = 3%	FAR = 5%
2 classifiers	5.6%	5.0%	4.9%
3 classifiers	6.0%	5.0%	4.9%

### H. Choosing FAR for threshold

For different training FARs, the conditions  $p = 25$ ,  $l = 10$ ,  $K = 20$  and  $C = 3$  were used. The results are shown in figure 11. It can be seen that for a low final FAR a higher training FAR works best, but for higher final FAR a lower training FAR works better. Similar results can be found using three classifiers. Corresponding EER can be found in table II. Higher FAR score lower EER. This is because a higher training FAR results in a lower threshold for the classifier. This makes the first classifier more powerful so the results will resemble the single-classifier results more.

Fig. 11. False Non-Match Rate for different training FAR



## V. DISCUSSION

In this research a relatively simple approach was chosen to determine outliers. This is discussed in V-A. It was assumed

these outliers could be modelled using a simple classifier assuming a Gaussian distribution. This is discussed in V-B

### A. Methods for finding outliers

In this research false reject pairs were used to determine outliers. All samples occurring more than a certain amount of times in false reject pairs were labelled as outliers. However, this method can fail to extract all of the samples lying outside the Gaussian distribution while also including samples lying inside the assumed distribution. For example, if 5 out of 20 faces lie outside of the assumed distribution, this will result in at least 15 false non-match pairs for each face lying outside of the distribution. The faces lying inside the assumed distribution will occur in the array of non-match pairs multiple times, so they will also be assumed to lie outside of the Gaussian distribution. As a result, all pictures in this class will be extracted and used in the new training set to produce the next classifier. If this happens for too many classes, the entire training set will be labelled as an outliers and only one classifier can be formed. On the other hand, if the new training set is too small, the quality of the classifier formed will suffer.

Several other approaches can be used to extract the next training set. One approach would be using PCA to extract clusters of points oriented closely to one other. The clusters not falling in the assumed distribution can be used to design the next classifier. Building an extension to determine outlier clusters could be complex and requires knowledge about how the classifier operates.

Another approach would be looking at the mean of all pairs within the same class, instead of looking at the individual scores for each pair. If the mean is higher than the threshold, the sample is close to most of the other pictures and likely to be in the same distribution. If the mean is lower than the threshold, the sample is far away from most of the other pictures and likely to be outside of the distribution. This method can fail if too many samples lie outside of the distribution.

Instead of extraction based on index one could also base the extraction on class. In this case the classes having lower scores will have all their samples extracted into the next set.

### B. Forming classifiers for a non Gaussian distribution

As discussed in section III-B, there are multiple ways outliers can manifest themselves. For useful classification, outliers outside of any cluster or outside face space have to be filtered out of the data. In figure 2 it can be seen that the outliers will probably still be very widespread, not forming a Gaussian distribution. A solution could be to design multiple classifiers working different parts of the distribution. One has to take into account the different clusters of samples and

design classifiers specifically for these different clusters. It also has to be taken into account the outlier clusters will be assigned a different class. Multiple classes can then belong to the same person. A system could be designed to map the class assigned to the clusters back to the original class of the same person.

## VI. CONCLUSION AND FURTHER RESEARCH

This article investigates whether cascading multiple LDA classifiers can be beneficial for facial recognition and what the role of different parameters is for the performance of this cascade. A method of cascading classifiers based on outliers was proposed. The easily implementable approach of finding outliers using false reject pairs in a training set was chosen. When changing parameters, better results were found when the cascaded classifier resembled the single classifier more. For all different tried parameters, the number of false accepts increases more rapidly than the number of true matches as the number of classifiers increases. By adding a classifier, the EER increased approximately 0.5%. This indicates that the chosen fashion of cascading classifiers is not beneficial for facial recognition. The reason for the malfunctioning of the classifier system is that the training set used for training the classifiers contains multiple different clusters with different kind of distributions. These cannot be picked up by one or multiple classifiers using the simple outlier method.

Further research can be done on designing more robust ways to find and classify outliers in a dataset. Experiments can be done using this classification system in combination with multiple enrollment. The performance of the classification system can be optimized choosing a different number of PCA and LDA components on different levels. Furthermore, the number of classifiers can be increased. Experiments can be done using the triplet loss methods to optimize the results.

## REFERENCES

- [1] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *J. Opt. Soc. Am. A*, vol. 4, pp. 519–524, Mar 1987.
- [2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," in *European conference on computer vision*, pp. 43–58, Springer, 1996.
- [3] L. Spreeuwers, R. Veldhuis, S. Sultanali, and J. Diephuis, "Fixed far vote fusion of regional facial classifiers," in *BIOSIG 2014: Proceedings of the 13th International Conference of the Biometrics Special Interest Group* (C. Busch and A. Brömme, eds.), (Germany), pp. 1–4, Gesellschaft für Informatik, 9 2014.
- [4] A. Kapri, "Pca vs lda vs t-sne," 2020. [Online; accessed June 23, 2020].
- [5] L. Spreeuwers, "Derivation of lda log likelihood ratio one-to-one classifier," *University of Twente students journal of biometrics and computer vision*, vol. 2014, no. 1, p. 5, 2014.
- [6] A. Patel and W. A. Smith, "Manifold-based constraints for operations in face space," *Pattern Recognition*, vol. 52, pp. 206 – 217, 2016.
- [7] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell, "Face recognition with image sets using manifold density divergence," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 581–588 vol. 1, 2005.
- [8] J. Kittler, "Combining classifiers: A theoretical framework," *Pattern analysis and Applications*, vol. 1, no. 1, pp. 18–27, 1998.
- [9] S. Tulyakov, S. Jaeger, V. Govindaraju, and D. Doermann, *Review of Classifier Combination Methods*, pp. 361–386. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.
- [10] D. Ruta and B. Gabrys, "An overview of classifier fusion methods," *Computing and Information Systems*, vol. 7, pp. 1–10, 01 2000.
- [11] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, pp. I–I, 2001.
- [12] P. R. K. Thomas A. Chmielewski, "Biometrics techniques: The fundamentals of evaluation," *J. Opt. Soc. Am. A*.
- [13] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991. PMID: 23964806.
- [14] P. J. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. K. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," vol. 1, pp. 947– 954 vol. 1, 07 2005.