Experiments on Deep Face Recognition using Partial Faces

Ali Elmahmudi

Centre for Visual Computing, Faculty of Engineering and Informatics
University of Bradford
Bradford, UK
a.a.m.elmahmudi@bradford.ac.uk

Hassan Ugail

Centre for Visual Computing, Faculty of Engineering and Informatics
University of Bradford
Bradford, UK
h.ugail@bradford.ac.uk

Abstract—Face recognition is a very current subject of great interest in the area of visual computing. In the past, numerous face recognition and authentication approaches have been proposed, though the great majority of them use full frontal faces both for training machine learning algorithms and for measuring the recognition rates. In this paper, we discuss some novel experiments to test the performance of machine learning, especially the performance of deep learning, using partial faces as training and recognition cues. Thus, this study sharply differs from the common approaches of using the full face for recognition tasks. In particular, we study the rate of recognition subject to the various parts of the face such as the eyes, mouth, nose and the forehead. In this study, we use a convolutional neural network based architecture along with the pre-trained VGG-Face model to extract features for training. We then use two classifiers namely the cosine similarity and the linear support vector machine to test the recognition rates. We ran our experiments on the Brazilian FEI dataset consisting of 200 subjects. Our results show that the cheek of the face has the lowest recognition rate with 15% while the (top, bottom and right) half and the 3/4 of the face have near 100% recognition rates.

Keywords— Face Recognition, Partial Face, Deep Learning, Cosine Similarity, Support Vector Machine.

I. INTRODUCTION

Face recognition is an inherently powerful ability that humans possess. In this respect, there are special areas in the human brain dedicated to face processing and face recognition [1]. Though this may be the case, there are natural limitations on the human brain when it comes to face recognition, especially when tackling the problem of identifying individuals from large datasets containing exceedingly unfamiliar faces. On the other hand, more recently, machine based algorithms are proving to be far better placed in successfully addressing this problem.

Face recognition using partial facial data is a particularly challenging problem. This is evident even from human based face stimuli experiments. For example, apart from particularly familiar faces, humans find it difficult to discriminate similar faces from individual parts of the face such as the eyes, mouth and the nose [6].

The work of Murphy et al., [7], for example, based on facial stimuli shows the mechanism of human face perception. Their work, along with that of others, shows for humans faces are difficult to perceive when turned upside down. Moreover, in their experiments, they tried to test the ability for a participant to classify faces presented in whole and region by region using a dynamic aperture which moved incrementally through the facial picture. The main idea in this work was to understand the limits of human ability for face perception and recognition. In their work, they tested this idea in four ways, namely for identity, gender, age and expression under four conditions, which are, upright whole face, inverted whole face, upright aperture and inverted aperture. The results presented by the observers were put into categories of identity, gender, age and expressions. Their results indicated that the detrimental effects of an inverted whole face were no less in the aperture conditions of showing partial face to the participants.

Similarly, Andre and Nummenmaa [8] studied face recognition on the partial face subject to the presence of facial expressions. In one of their experiments, they tested the face recognition rates for the common six expressions - happiness, anger, sadness, disgust and fear. In the case of the partial face, they partitioned the face into two regions, one containing the eyes and the other containing the mouth. A considerable result of their work is that humans have poor recognition rates when it comes to the situation of the eye only and mouth only. On the other hand, they noted that the expression of smile produces slightly better recognition rates.

Many previous studies note that, when it comes to human face recognition, familiarity appears to be a key recognition factor. The rate of familiarity of course changes when the target face image is partial, occluded, with expressions and aged [7, 8].

On the other hand, machine learning algorithms can utilise the power of computations to use copious amounts of input data for training and use numerical analysis in order to produce outputs which can challenge the power human face recognition. Thus, machine learning helps a computer to build models from examples of input data with a view to making a more accurate decision. This is a distinct advantage that machine learning algorithms appear to have over human face perception and recognition. Thus, it is also plausible to state that machine learning algorithms can potentially provide better recognition rates on partial faces or, in the worst case, may aid humans to

perform better at face recognition, especially in challenging cases where very limited or partial facial data are presented.

II. RELATED WORK

When it comes to machine based face recognition, traditional algorithms, such as Principal Component Analysis [2, 3] and Linear Discriminant Analysis [4] have been widely utilised. However, more recently the use of the machine learning approaches such as Convolution Neural Networks (CNNs) [5] have been consistently proving to be accurate, efficient and thus very useful.

Though face recognition through machine learning algorithms is a well-addressed problem, and it has to date yielded promising results, the problem of face recognition has been predominantly restricted to the utilisation of the full face. In fact, to date, little work appears to have been done on the use of partial face for face recognition.

As far as machine based partial face processing and recognition are concerned the literature on this topic appear to be sparse. The earliest work on partial face we are aware of is that of Savvides et al., [10] in which they tested selected facial regions to establish quantifiers with discriminative ability. Based on grey scale images they applied the method of kernel correlation filters to reduce image dimensionality and for feature extraction [11]. They then utilised Support Vector Machine (SVM) to discriminate between features. Three main face regions were tested by their approach, namely the eye, nose and mouth. Results from their experiments suggest that the eye region has a higher verification rate compared to the mouth and nose regions.

Similarly, Cai et al., [24] proposed a facial variation modelling system for sparse representation for face recognition. This can help single sample face recognition algorithms to build facial variation bases to separate neutral, frontal faces from different facial views. Their experiments show that major enhancements can be performed in single image face recognition problems.

Another piece of work which is notable in this area is by Li et al., [12] which considered that the human face recognition problem in frontal views with varying illumination, disguise and occlusion. They presented a new method for face recognition which extracts a dynamic subspace from images and obtain the distinctive parts in each subject. A characteristic of discriminative components was represented by those parts in order to give a recognition protocol to classify face images by using k-nearest neighbour algorithm (K-NN) [13]. They applied their method to public databases such as ORL, Extended Yale B and the results illustrated that the recognition rates could be improved using partial face cues.

Further, Peng et al., [14], introduced a technique called Locality-Constrained Collaborative Representation (LCCR) to enhance discrimination of representative images. The LCCR was applied to different databases with five distance measures. In the case of partial faces, they used three facial features, i.e. right eye, nose and mouth with chin by masking the original images. The results indicate that the right eye mouth and chin

have high recognition rates – for example, by using LCCR and Cityblock distance measures [15].

Apart from some isolated pockets of work mentioned above, to our knowledge, to date, there has not been a rigorous study based on machine learning to test the recognition rates for various parts of the face. The prime aim of this work hence is to try and close that gap. Thus, in this work, we have proposed an efficient machine learning framework to look deeply into the effects of partial face for recognition. In particular, we try to understand the specific recognition rates for various parts of the face. To do this, we have implemented our framework using the state of the convolutional neural network (CNN) architecture for facial feature extraction, and we have utilised both Cosine Similarity (CS) and SVM for classification and testing.

The rest of the paper is organised as follows. In Section II, we explain the CNN architecture we have utilised along with a brief description of the VGG-Face model and the CS as well as the SVM based classification. In Section III, we discuss the face recognition experiments we have carried out using partial face data, and we report some of the interesting results we obtained following those experiments. In Section IV, we reflect on the results and finally, in Section V, we conclude this paper.

III. PROPOSED METHODOLOGY

One of a most popular examples of machine learning in the recent times has been those based on deep learning, otherwise known as Convolutional Neural Networks (CNNs), the use of which has been literary explosive in the area of visual computing. In fact, the exploitation of CNNs for face processing and face recognition is noteworthy here. CNNs are supervised machine learning techniques that can extract "deep" knowledge from a dataset through rigorous example based training - in a sense this approach mimics how human brain undertake the process of learning. CNNs have been successfully applied to feature extraction, face recognition, classification, and segmentation, to name some. As noted here, the explosion on the use of CNNs in recent times is due to their ability to learn complex features using nonlinear multi-layered architectures [17]. Though the origin of CNN goes back to the early 1990s, the predominant scepticism for using CNN has been based on the assumption that feature extraction using gradient descent will always overfit. The main argument for this has been that gradient based optimisation methods are prone to get stuck in local minima. However, in recent times, these assumptions have been overturned due to the promising results CNNs have produced across many domains of research. Thus, today, state-of-the-art deep learned models, based on CNN architectures are being used in almost all visual computing related domains. Examples include image perception [18], recognition [19], classification [20] [25], and information retrieval [21].

Generally, there are three ways of deploying CNNs. They are training a network from scratch, fine-tuning an existing model, or using off the shelf CNN features. The latter two approaches are referred to as transfer learning [22]. It is important to highlight that training CNN from scratch requires an enormous amount of data, which is often a huge and challenging task [23]. On the other hand, fine-tuning involves transferring the weights of the first n layers learned from a base network to a target

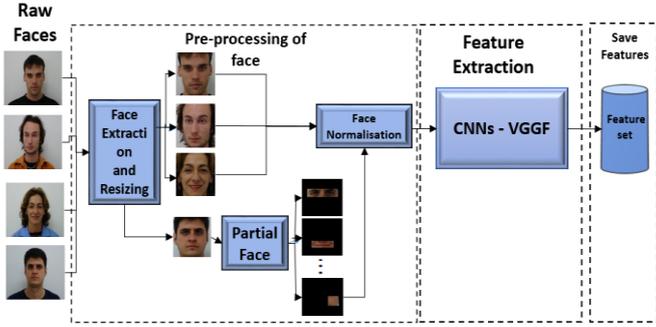network. The target network can then be trained using a new dataset.

Figure 1: Overview of our CNN VGGF based feature extraction approach for face recognition using partial face data.

For face perception work, using CNN, there are several pre-trained models which can readily be utilised for feature extraction, e.g. VGGF, FGG16, VGG19, OverFeat [9]. In our case, for feature extraction, we have utilised the VGGF pre-trained model which we discuss below. Thus, the methodology we adopt here uses the pre-trained VGGF model for feature extraction which is followed by CS [26] and linear SVM for classification. Figure 1 illustrates an overview of our feature extraction steps.

A. The VGG-Face Model

As mentioned above, there are several pre-trained models for CNN and one of the most popular and widely used in face recognition is the VGGF model - developed by Oxford Visual Geometry Group [9]. The model was trained on a huge dataset containing 2.6M face images of more than 2.6K individuals. The architecture of VGGF comprises 38 layers starting from the input layer up to the output layer. The input should be a colour image with a size of 224 by 224, and as the pre-processing step an average is normally computed from the input image.

In general, the VGGF contains thirteen convolutional layers, each layer having a special set of hybrid parameters. Each group of convolutional layer contains 5 maxpooling layers and there are also 15 rectified linear units (ReLUs). After these layers, there are three fully connected layers namely FC6, FC7 and FC8. The first two have 4096 channels, while FC8 has 2622 channels that are used to classify the 2622 identities. The last layer is the classifier which is a softmax layer to classify an image to which the individual face class belongs to.

B. Feature Extraction using the VGGF Model

Given an input image, X_0, it can be represented as a tensor $X_0 \in R^{HWD}$, where H is the image height, W is the width and D represents the colour channels. A pre-trained layer L of the CNN can be expressed as a series of functions, $g_L = f_1 \rightarrow f_2 \rightarrow \cdots \rightarrow f_L$.

Let $X_1, X_1, \dots, X_n$ be the outputs of each layer in the network. Then, the output of the ith intermediate layer can be computed from the function f_i and the learned weights w_i are via $X_i = f_i(X_{i-1} : w_i)$.

As we know that CNNs learn features through the training stage and use such features to classify images. Each convolutional (conv) layer learns different features. For example, one layer may learn about entities such as edges and colours of an image while further complex features may be learnt in the deeper layers. A result of conv layer involves numerous 2D arrays which are called channels. In VGGF, there are 37 layers, 13 of them are convolutions and the remaining layers are mixed between ReLU, pooling, fully connected and the last layer is the softmax. For example, in the conv1_1 layer, there are 64 filters with size 3x3 which are applied to an input image.

In order to decide the best layer within the VGGF model to utilise for facial feature extractions, we usually carry out a number trial and error experiments. In this particular case, we tested the layers 34 through to 37. In our experiments we tried other layers, but the best result came from layer 34. It noteworthy this layer is the fully connected layer and placed at the end of a neural network which means the extracted features represents the whole face.

The features from this layer are the results that arise from the fully connected layer FC6 after applying 'ReLU6', which gives a vector of 4096 dimensions. The suggestion that layer 34 was optimal was inferred by undertaking a number of face recognition tests where we used the full frontal face for both training and testing thereby an obtaining the rate of 100% recognition. The whole process of training and testing through feature extraction is further described in Algorithms 1.

Algorithm 1: Feature Extraction – from the Face dataset

Input: Training set M, with m classes
n_j = number of images in a given class
for $i = 1\ to\ m$ **do**
    for $j = 1\ to\ n_j$ **do**
        $im \rightarrow read\ an\ image$
        $im \rightarrow resize(im)$
        $im \rightarrow normalize(im)$
        $im\ features \rightarrow ExtractFeatures(CNNs(im))$
    endfor
endfor

C. Feature Classification

A classification in a supervised machine learning is a function that assigns new observational items to which a set of target categories or classes belong to. In other words, the objective of classification is to build a brief model of the distribution of class labels in terms of predicted features [28]. There are several techniques for this classification [28] - decision trees [29], K-NN [30], SVM [27] are good examples.

In this work, all extracted features in both the training and testing phases are used for the purpose of classification. In our experiments, for the classification scenarios, we have utilised the CS [26] and SVM classifiers [27]. There are two reasons for this choice. Firstly, we tested other classifiers and the best results were by using CS and SVM. Secondly, through our experiments and analysis, we found out that these two classifiers have an ability to separate data more accurately.

1. Cosine Similarity

A cosine similarity is a measure between two non-zero vectors. It uses the inner product space to measure the cosine of the angle between those two vectors [26]. The Euclidean dot product formula as in Equation 1 can be used to compute the cosine similarity such that,

$$a \cdot b = ||a|| \, ||b|| \cos \theta, \tag{1}$$

where a and b are two vectors and θ is an angle between them.

By using the magnitude or length, which is same as the Euclidean norm or the Euclidean length of vector $x = [x_1, x_2, x_3, \ldots, x_n]$ as in Equation 2, the similarity S is computed using the formulation given in Equation 3.

$$\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2 + \cdots + x_n^2}, \tag{2}$$

$$S = \cos \theta = \frac{A \cdot B}{\| A \| \, \| B \|} \tag{3}$$

$$= \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}},$$

where A and B are two vectors.

For classification, in all our experiments we compute the CS to find the minimum "distance" between the test image $test_{im}$ and training images $training_{im}^n$ by using Equations 4 and 5.

$$M_{CS} = min \left(dist(test_{im}, training_{im}^n) \right), \tag{4}$$

where im is an image number and n is a total images in the training set and,

$$CS(test_{im}, training_{im}^n) = \frac{\sum_{j=1}^{m} training_{jm}^i \, test_{im}}{\sqrt{\sum_{j=1}^{m} training_{jm}^{i2}} \sqrt{\sum_{j=1}^{m} test_{im}^2}}, \tag{5}$$

where m is a length of vector.

2. SVM Classifier

SVM is a supervised machine learning algorithm which can be used for both binary classification and multi classification problems. An SVM focuses on identifying the "margin" via a hyperplane to separate the data into classes. Maximising the margin reduces the upper bound on the expected generalisation error by creating the largest possible distance between the separating hyperplanes [27]. There are two types of SVMs, namely linear and a non-linear and in this work we use the linear SVM to separate the data by using a one-versus-one approach. We applied two different non-linear kernels (Gaussian and Polynomial of degree 2) and we found that the results were worse than the linear case. Hence our choice of linear kernel here.

Suppose we have training dataset (x_i, y_i), we can use the linear SVM as in Equation 6 such that,

$$\min_{w \in R^d} \frac{1}{2} \|w\|^2 + C \sum_i^N \max(0, 1 - y_i w^T x_i), \tag{6}$$

where w is a weight vector, N is a number of classes and C is trade off parameter between error and margin.

IV. EXPERIMENTS AND RESULTS

Here we present a comprehensive set of experiments conducted on face recognition using different parts of the face. To undertake this work, we have utilised face images from the FEI dataset [16] where images were manually cropped by removing the background.

The FEI dataset contains Brazilian faces of 200 students and staff with an equal number of males and females from FEI University. For each subject, there are 14 images bringing the total number of images in the dataset to 2800. The resolution of the images is 640 pixels by 480 pixels. All images are colour and taken against a white homogeneous background. The subjects are between 19 and 40 years old. The dataset contains images with variations in facial expressions as well as the pose. Figure 2 (a) shows sample images from the FEI face dataset.

In this work, we conducted two main sets of experiments – one which does not use the partial face as part of the training face data and the other in which partial faces have been utilised as part of the training data. In each case, we undertook 12 sub experiments involving the partial face on both classifiers. For training purposes 70% of the images were utilised which were also augmented through operations such as padding and flipping. The remaining 30% of the images were used for testing, in each case.

Figure 2: (a) Sample face data from the FEI dataset. (b) Parts of the face we have used for testing the recognition rates.

Thus, in our experiments, twelve test sets were generated thereby each test corresponding to one part of the face. The parts were eyes, nose, right cheek, mouth and the forehead. Also, faces were generated just with eyes and nose, bottom half face, the top half of the face, right half and three quarters of the face as well as the full face. Figure 2 (b) shows the parts of the face we have used for testing the recognition rates.

After extracting features from the VGGF model, the CS and the linear SVM classifiers were applied in order to investigate the rate of recognition for each facial part separately. The results of this experiment are summarised in Figure 3. As it can be inferred from the graph, the highest rate of recognition is achieved with the full face and the three-quarter faces with the recognition rate of 100% using both classifiers. However, the recognition rate starts dropping down slightly at the right half and the top half of the face respectively with SVM, but in the case of CS, the rate still holds at 100%. As we approach the bottom half of the face, the rate decreases further reaching to about 50% in the case of SVM and about 60% for CS. This decline continues until nearly 0.5% at the right cheek.
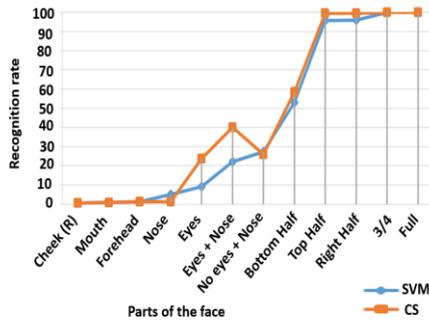
Figure 3: Face recognition rates using SVM and CS classifiers based on parts of the face without using individual facial parts of the face in training.
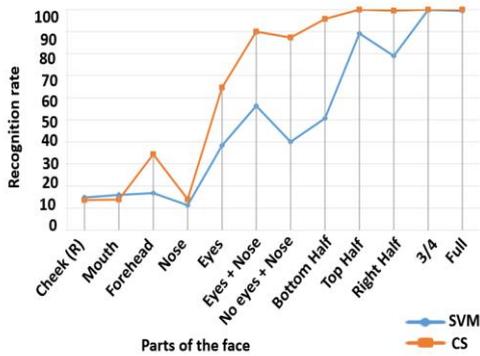

Figure 4. Face recognition rates using SVM and CS classifiers of based on parts of the face with the use individual facial parts of the face in training.

In the second set of experiments, to measure the rate recognition of face using parts of the face, we repeated the above procedure, but this time we added the individual parts of the face into training set too. As shown in Figure 4, it can be seen the recognition rates have significant improvements in this case. For instance, while the results from right cheek previously were nearly to 0%, it has moved up to 15% using both the classifiers. Also, in the case of the combined eyes and nose, it was 22% for SVM and 40% for CS previously, and in this case, has improved to about 57% for SVM and 90% for CS. However, we have noticed that not all recognition rates steadily increased in this particular case. In fact, in some cases, the results were slightly worse while using SVM. For example, a slight decrease in the recognition rate was observed at the bottom half of face, which was 53% and dropped down to 51%. In contrast, the CS has produced a significant improvement, for instance, the recognition rate for combined eyes and nose increased from 40% to 90%.

V. DISCUSSIONS

The results of the study we present here concerns CNN based face recognition using partial faces. We conducted two sets experiments using partial faces. In the first set of experiments, the training set did not include parts of the face. The highest recognition rate observed was on the 3/4 face where we found the recognition rate to be 100% by using SVM and in the case

of CS the right half, the top half as well as well as the 3/4 face returned the recognition rate of 100%. We have also observed that for combined eyes and nose, whether occluded or not, the percentage recognition was relatively low for SVM, but it was higher for CS. The worst recognition rates observed are for the smaller and perhaps less significant parts of the face such as the cheeks, the forehead and the mouth, for both classifiers which was about between 0-1.33%.

Contrary to the findings of the first set of experiments, in the second set, where the parts of face were added to the training set, the recognition rate did improve, especially in the smaller and less significant parts of the face, like the cheek from 0% to 15% and the mouth area improved from 1% to about 16%. Additionally, recognition using the nose area has slight improvement from 5% to 10%. Moreover, the rate on the forehead also has a significant increase from nearly 0.5% to around 18%, and the rate of recognition from the eyes' region has significantly increased from 10% to about 40%. Further, we notice that better recognition results overall were achieved by using the CS measure. This is particularly significant, for example, for eyes which was 23% and reached 65% and also the combined areas of eyes and nose show a marked improvement from about 40% to nearly to 90%.

By analysing the results of these experiments, we can make further observations about the accuracy of classifiers between CS and SVM in that, in general, the CS outperforms the SVM. This is also observed at the individual class level where we looked at the matching images picked by the classifiers, as discussed below by way of some examples.

In the first set of experiments, in which we did not train the classifier with parts of the faces, the image no.13 for subject no.1 matched with full face image and same subject by using the CS measure as in Figure 5 (a).

In the second set of experiments, where parts of the face were added to the training set, we can observe greater matching performance using CS especially for the smaller and less significant parts of the face such as part of the cheeks. For example, Figure 5 (b) shows the image no 11 being correctly matched to image no 254 for subject 11.
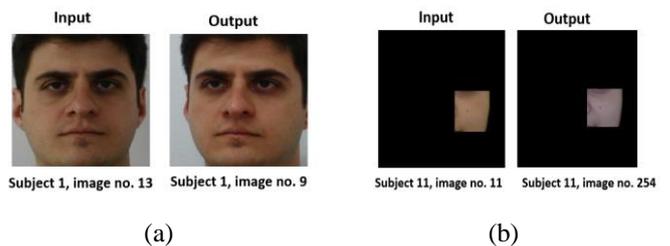

Figure 5: (a) An example of correct class level matching using CS. (b) The result of correct matching suing CS, for parts of the cheek

VI. CONCLUSIONS

In this paper, the use of deep learning approach for face recognition using partial face data is studied. Based on the popular CNN and using the VGGF model for extracting features from faces we utilised two different classifiers namely the cosine similarity and the linear support vector machine for

classification. We conducted our partial face recognition experiments on a publicly available database, namely the FEI database. To test the face recognition rates using partial face data, we conducted two sets of experiments. In the first set of experiments, we measured the partial face recognition rates without adding the partial faces to the training data. In the second case, we measured the recognition rates after adding the partial faces to the training data. The results show marked improved when partial data is added to the training set especially the case of smaller or less significant parts of the face, such as the forehead, parts of the mouth and parts of the cheek.

Our extensive experimental results validate that the cheek, nose, forehead and mouth have low recognition rates. On the other hand for top half of the face, right or left half and for 3/4 of the face the recognition rates reach 100%. In addition, the Cosine similarity measure greatly improves the performance of the classification when compared to the SVM.

There are many avenues through which this present work can be further taken forward. It is noteworthy, in this work, we have utilised the pre-trained VGG Face model. It might be worth testing our framework using other CNN architectures and models. We also feel our experimental setup can be further improved in terms of bringing more consistency in selecting facial features for recognition. Additionally, the results we have obtained in this work must also be challenged by running the experiments on much more comprehensive datasets.

ACKNOWLEDGEMENTS

REFERENCES

[1] S. M. Andersen, C. A. Carlson, M. A. Carlson, and S. D. Gronlund, "Individual differences predict eyewitness identification performance," Personality and Individual Differences, vol. 60, pp. 36-40, 2014.

[2] A. M. Aguilera, M. Escabias, and M. J. Valderrama, "Using principal components for estimating logistic regression with high-dimensional multicollinear data," Computational Statistics & Data Analysis, vol. 50, pp. 1905-1924, 2006.

[3] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on, 1991, pp. 586-591.

[4] S. Suganya and D. Menaka, "Performance Evaluation of Face Recognition Algorithms," International Journal on Recent and Innovation Trends in Computing and Communication, 2(1), pp. 135-140, 2014.

[5] A. Smola and S. V. Vishwanathan, "Introduction to machine learning. The Pitt Building," Trumpington Street, Cambridge, United Kingdom: the press syndicate of the University of Cambridge, 2008.

[6] J. W. Tanaka, M. D. Kaiser, S. Hagen, and L. J. Pierce, "Losing face: impaired discrimination of featural and configural information in the mouth region of an inverted face," Attention, Perception, & Psychophysics, vol. 76, pp. 1000-1014, 2014.

[7] J. Murphy and R. Cook, "Revealing the mechanisms of human face perception using dynamic apertures," Cognition, vol. 169, pp. 25-35, 2017.

[8] M. G. Calvo, A. Fernández-Martín, and L. Nummenmaa, "Facial expression recognition in peripheral versus central vision: Role of the eyes and the mouth," Psychological research, vol. 78, pp. 180-

[9] 195, 2014.

[9] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in BMVC, p. 6, 2015.

[10] M. Savvides, R. Abiantun, J. Heo, S. Park, C. Xie, and B. Vijayakumar, "Partial & holistic face recognition on frgc-ii data using support vector machine," in Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on, 2006, pp. 48-48.

[11] M. Savvides, B. V. Kumar, and P. Khosla, "Face verification using correlation filters," 3rd IEEE Automatic Identification Advanced Technologies, pp. 56-61, 2002.

[12] H. Li and C. Y. Suen, "Robust face recognition based on dynamic rank representation," Pattern Recognition, vol. 60, pp. 13-24, 2016.

[13] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," The American Statistician, vol. 46, pp. 175-185, 1992.

[14] X. Peng, L. Zhang, Z. Yi, and K. K. Tan, "Learning locality-constrained collaborative representation for robust face recognition," Pattern Recognition, vol. 47, pp. 2794-2806, 2014.

[15] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," City, vol. 1, pp. 300-307, 2007.

[16] D. C. E. Thomaz, FEI Face Database, The Artificial Intelligence Laboratory of FEI in So Bernardo do Campo, So Paulo, Brazil.

[17] Y. Bengio, I. J. Goodfellow, and A. Courville, "Deep learning," Nature, vol. 521, pp. 436-444, 2015.

[18] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al., "Imagenet large scale visual recognition challenge," International Journal of Computer Vision, vol. 115, pp. 211-252, 2015.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.

[20] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, p. 3, 2017.

[21] Y. Zhong, R. Arandjelovi, and A Zisserman, "Faces in places: Compound query retrieval," In BMVC-27th British Machine Vision Conference, 2016.

[22] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pp. 1717-1724, 2014.

[23] A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," in Proceedings of the 23rd ACM international conference on Multimedia, pp. 689-692, 2015.

[24] J. Cai, J. Chen, and X. Liang, "Single-sample face recognition based on intra-class differences in a variation model," Sensors, vol. 15, pp. 1071-1087, 2015.

[25] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in Computer vision and pattern recognition (CVPR), 2012 IEEE conference on, pp. 3642-3649, 2012.

[26] G. Sidorov, A. Gelbukh, H. Gómez-Adorno, and D. Pinto, "Soft similarity and soft cosine measure: Similarity of features in vector space model," Computación y Sistemas, vol. 18, pp. 491-504, 2014.

[27] S. Amarappa and S. V. Sathyanarayana, "Data classification using Support vector Machine (SVM), a simplified approach", Internationl Journal of Electronics and Computer Science Engineering. ISSN-2277-1956, pp. 435-445, 2014.

[28] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," Emerging artificial intelligence applications in computer engineering, vol. 160, pp. 3-24, 2007.

[29] S. K. Murthy, "Automatic construction of decision trees from data: A multi-disciplinary survey," Data mining and knowledge discovery, vol. 2, pp. 345-389, 1998.

[30] P. Cunningham and S. J. Delany, "k-Nearest neighbour classifiers," Multiple Classifier Systems, vol. 34, pp. 1-17, 2007.