Future Generation Computer Systems 99 (2019) 213–225

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Deep face recognition using imperfect facial data

Ali Elmahmudi, Hassan Ugail *

Centre for Visual Computing, Faculty of Engineering and Informatics, University of Bradford, Bradford BD7 1DP, UK

HIGHLIGHTS

- We show the performance of machine learning for face recognition using partial faces and other manipulations of the face such as rotation and zooming which we use as training and recognition cues.
- We use the state of the art convolutional neural network based architecture along with the pre-trained VGG-Face model through which we extract features for machine learning.
- Our results show that individual parts of the face such as the eyes, nose and the cheeks have low recognition rates though the rate of recognition quickly goes up when individual parts of the face in combined form are presented as probes.

ARTICLE INFO

Article history:
Received 14 December 2018
Received in revised form 1 March 2019
Accepted 10 April 2019
Available online 27 April 2019

Keywords:
Face recognition
Convolutional neural networks
Deep learning
Cosine similarity

ABSTRACT

Today, computer based face recognition is a mature and reliable mechanism which is being practically utilised for many access control scenarios. As such, face recognition or authentication is predominantly performed using 'perfect' data of full frontal facial images. Though that may be the case, in reality, there are numerous situations where full frontal faces may not be available — the imperfect face images that often come from CCTV cameras do demonstrate the case in point. Hence, the problem of computer based face recognition using partial facial data as probes is still largely an unexplored area of research. Given that humans and computers perform face recognition and authentication inherently differently, it must be interesting as well as intriguing to understand how a computer favours various parts of the face when presented to the challenges of face recognition. In this work, we explore the question that surrounds the idea of face recognition using partial facial data. We explore it by applying novel experiments to test the performance of machine learning using partial faces and other manipulations on face images such as rotation and zooming, which we use as training and recognition cues. In particular, we study the rate of recognition subject to the various parts of the face such as the eyes, mouth, nose and the cheek. We also study the effect of face recognition subject to facial rotation as well as the effect of recognition subject to zooming out of the facial images. Our experiments are based on using the state of the art convolutional neural network based architecture along with the pre-trained VGG-Face model through which we extract features for machine learning. We then use two classifiers namely the cosine similarity and the linear support vector machines to test the recognition rates. We ran our experiments on two publicly available datasets namely, the controlled Brazilian FEI and the uncontrolled LFW dataset. Our results show that individual parts of the face such as the eyes, nose and the cheeks have low recognition rates though the rate of recognition quickly goes up when individual parts of the face in combined form are presented as probes.

1. Introduction

Faces are the most painted pictures in the visual system within the life time of a human being. In this sense, it is not surprising that humans possesses remarkable ability to be able to recognise faces. Typically, it takes a glimpse of someone's face for us to remember that individual. Thus, it is not surprising that humans have a dedicated region in the brain for solely processing faces as well as for recognising them [1]. When it comes to face recognition by humans, it is thought that the brain remembers important details such as the shapes and colours of crucial features corresponding to the eyes, nose, forehead, cheeks and the mouth [2]. Moreover, the human brain can cope with significant variations in lighting, facial expressions as well as faces observed from afar. However, contrary to this, for a computer, in general, a variability of an appearance of a face has a direct effect on its capacity for recognition. For instance, the variations in

* Corresponding author.
E-mail addresses: a.a.m.elmahmudi@bradford.ac.uk (A. Elmahmudi), h.ugail@bradford.ac.uk (H. Ugail).

https://doi.org/10.1016/j.future.2019.04.025

Fig. 1. An example of how partial faces may be presented as input probe images for face recognition.

illumination, expression, pose and other added physical changes, i.e. glasses or beard, may have a huge impact on the recognition rates. Though this may be the case, due to the copious amount information a computer can cope with and along with the increased data processing power of machines, it is thought there are areas a machine algorithm, such as a convolutional neural network (CNN), can at least match or outperform when it comes for face recognition. Thus, the aim of this work is to explore the intriguing question of how a computer performs face recognition subject to imperfect facial information as recognition cues. More specifically, this work is geared to explore how various parts of the face perform on the task of face recognition and how faces observed from a distance – via the effect of zooming out, for example – as well as how the rotation of facial images perform on the task of face recognition.

In Fig. 1, we illustrate a typical example of how face recognition can be called for based on partial facial data as input probe images. Some of the recent approaches to classify and recognise a face are discussed in [3–5]. As the example illustrates, under no circumstances, the full face is available and only parts of the face such as the eyes, forehead, mouth, nose or the cheeks of the given subject are available as input probe data as [6–8].

The practical application of this work is borne out by the increasing need to undertake functionally automatic face recognition tasks in everyday environments. Like other biometric authentication tools such as the fingerprints, face recognition and face perception has become a very common practice [9,10]. To this end, reliable automated face processing [11,12] and recognition [13–16] tools which can utilise practical facial data, such as images that come from everyday CCTV cameras, are becoming paramount. In Fig. 2, we illustrate the overview of our face recognition framework using partial faces as probes.

2. Related work

As far as computer aided face recognition based on partial facial images are concerned, the literature surrounding this topic appear to be relatively sparse and not so consistent. Many algorithms have been introduced to solve face recognition problems, e.g. [17,18]. One of the earliest work on this topic we could identify was that of Savvides et al. [19]. In that study, they tested on various facial regions to establish quantifiers with discriminative ability. Based on grey scale images the method of kernel correlation filters was utilised to reduce image dimensionality and for feature extraction [20]. Following that, they utilised Support Vector Machines (SVM) to discriminate between various facial features. In their work, they test three main face regions, 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.

In a similar fashion, He et al. [21] introduced a technique called the Dynamic Feature Matching (DFM) for partial face recognition. Their study was based on a combination of fully convolutional networks (FCN) [22] with sparse representations. The purpose of FCN is to extract a feature map of images which has the capacity to cater for more discriminative features. The heart of their work is the utilisation of VGG-Face model [23] from which features were transferred to the FCN. This method appears to have produced good classification accuracy compared to other existing methods.

Furthermore, several robust face recognition methods have been suggested in order to address the challenges arising for face recognition due to face occlusion in different scenarios. In [24], Long et al. proposed Subclass Pooling for Classification (SCP) to solve the double occlusion problem by using limited data in a training set. They used a fuzzy max pooling method and average pooling schemes. Their results showed that a remarkable margin of performance can be achieved.

More recently, Lahasan et al. [25] proposed a framework named as the Optimized Symmetric Partial Facegraph (OSPE) for face recognition under different conditions. For example, occluded face, facial expression and variation of lighting are some of the cues they use in their experiments. Again, their experimental results have shown that some improvements in recognition rates can be achieved by introducing partial facial data.

Moreover, Duan et al. [26] introduced a technique called Topology Preserving Graph Matching (TPGM), in order to enhance

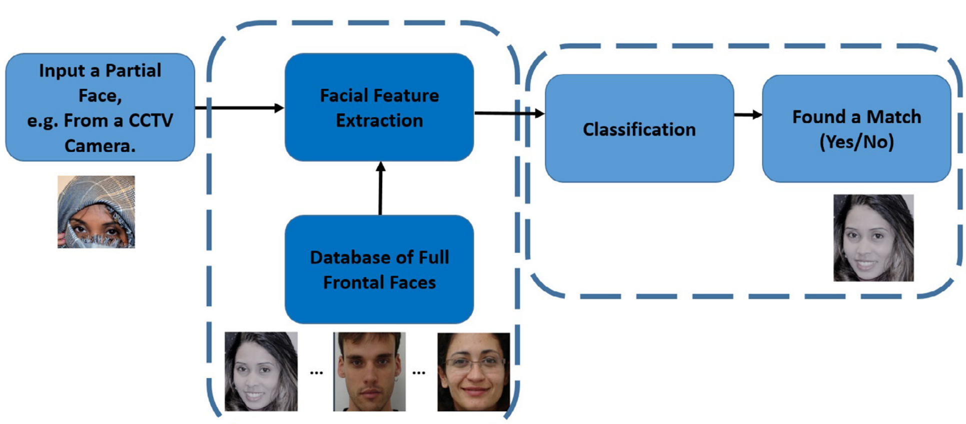

Fig. 2. An overview of our computational framework for face recognition using partial faces as probes.

Fig. 3. Humans possess inherent difficulty in accurately recognising inverted faces.

Fig. 4. Illustration of an example of a physical stimuli utilised to understand human face perception.

the recognition process in the case of using partial faces. Their method is based on building geometric graphs for probe faces and gallery faces. The TPGM method minimises a geometric and textural cost function. Outcomes of their experiments on four face databases demonstrated that their approach outperformed other state of the art methods at the time.

Similarly, Cai et al. [27] proposed a facial variation modelling system for sparse representation for face recognition. Based on a single sample face, they build facial variation bases to separate neutral, frontal faces from different facial views. Their experiments show that major enhancements can be performed for single image face recognition problems.

Another piece of work which is notable in this area is by Li et al. [28] 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 they obtain the distinctive parts for 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 the K-nearest neighbour algorithm (K-NN) [29]. They applied their method to public databases such as ORL and Extended Yale B. The results illustrated that the recognition rates could be improved using partial facial cues.

Furthermore, Peng et al. [30], 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 City Block distance measures [31].

Many mechanisms of human face perception in the literature have addressed how a human can perceive parts of the face including the inverted face [13–16]. The work of Murphy et al. [32], for example, based on facial stimuli shows the mechanism of human face perception. Their work, along with that of others, show, for humans, faces are difficult to perceive when turned upside down, as illustrated in Fig. 3. 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, as illustrated in Fig. 4. The main idea in this work was to understand the limits of human ability to 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 indicate 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 [33] 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.

However, when dealing with acute occlusions in a face, the performance of current methods declines remarkably. 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 with changes in the age of the subject [32,33].

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 of 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.

Based on the work carried on machine based face recognition, we note none of the studies has looked how machine learning favours in face recognition using partial faces in a consistent manner. Our aim in this study, therefore, is to close that gap. We study how different parts of the face favours in recognition. We also study how the rotation of the face as well as zooming out of the face at various levels fares recognition in a machine learning scenario. In our experiments, we use a CNN based architecture along with the pre-trained VGG-Face model to extract features. We then use two classifiers namely the cosine similarity (CS) and the linear SVM to test the recognition rates.

The rest of the paper is organised as follows. In Section 3, 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 4, we discuss the face recognition

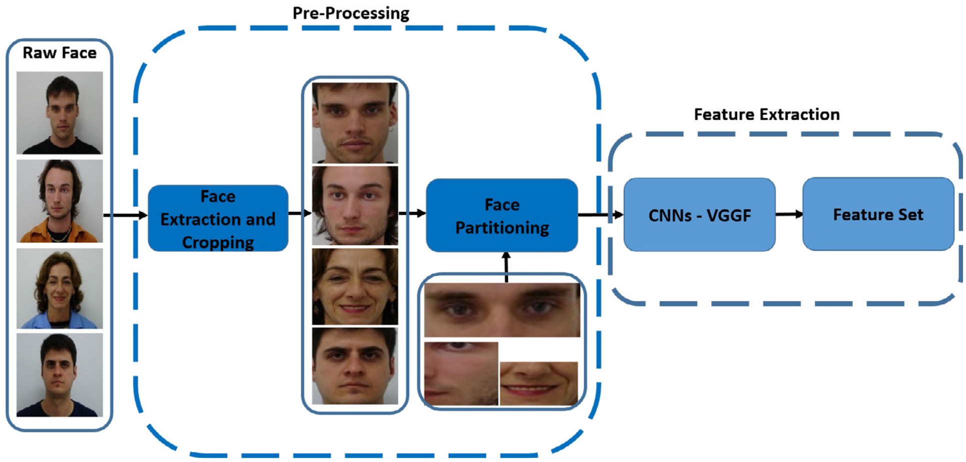

Fig. 5. Overview of our CNN VGGF based feature extraction approach.

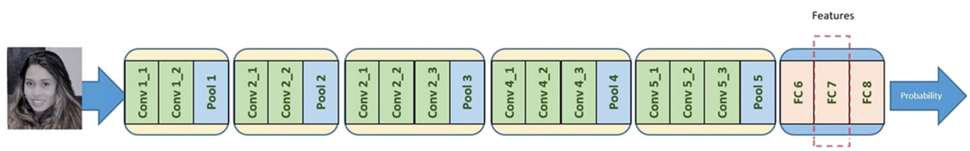

Fig. 6. Illustration of the convolutional layers in the VGGF model.

experiments we have carried out using partial face data, and we report some of the interesting results we obtained following those experiments. In Section 5, we reflect on the results and finally, in Section 6, we conclude this paper.

3. Proposed methodology

One of the 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. This machine learning approach mimics the human brain when 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 [34]. 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 notorious for getting stuck in the 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 [35], recognition [36], classification [37], and information retrieval [38].

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 [39]. It is important to highlight that training a CNN from scratch requires an enormous amount of data, which is often a huge and challenging task [40]. On the other hand, fine-tuning involves transferring the weights of the first few layers learned from a base network to a target network. The target network can then be trained using a new dataset.

For face perception work, using CNN, there are several pre-trained models which can readily be utilised for feature extraction, e.g. VGGF, VGG16, VGG19, OverFeat [23]. 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 [41] or linear SVM for classification. Fig. 5 illustrates an overview of our feature extraction steps.

3.1. 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 [23]. The model was trained on a huge dataset containing 2.6M face images of more than 2.6 K individuals. The architecture of VGGF comprises of 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 layers contains 5 maxpooling layers and there are also 15 rectified linear units (ReLUs). After these layers, there are three fully connected layers namely the FC6, FC7 and FC8. The first two have 4096 channels, while FC8 which has 2622 channels 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. We illustrate the architecture of this further in Fig. 6.

3.2. 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

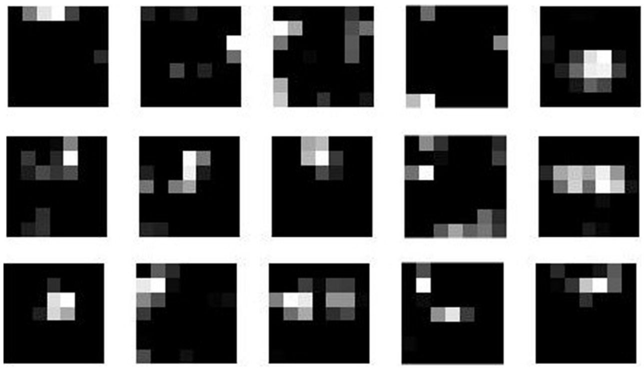

Fig. 7. A sample of the resulting features form the conv5_3 layer of the VGGF model.

represents the colour channels. A pre-trained layer L of the CNN can be expressed as a series of functions, $g_L = f_1 \rightarrow f_2 \rightarrow \ldots \rightarrow f_L$.

Let $X_1, X_2, \ldots, X_n$ 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 such that $X_i = f_i(X_{i-1} : w_i)$.

As we know, CNNs learn features through the training stage and then 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. For example, 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. However, after applying the conv5_3 layer to an input image, which has 512 filters with size 3x3, the features can be extracted for classification purposes. By examining the activations of that layer, one can obtain the main features as shown in Fig. 7, where a sample of the features is presented.

In order to decide the best layer within the VGGF model to utilise for facial feature extractions, one must usually carry out a number of 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 results came from layer 34. It is noteworthy that this layer is the fully connected layer and is placed at the end of a neural network which means the extracted features represents the whole face.

The features from the layer 34 are the results that arise from the fully connected layer FC7 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 obtaining the rate of recognition to be 100%. The whole process of training and testing through feature extraction is described further in Algorithms 1.

3.3. 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. There are several techniques for the classification — decision trees [42], K-NN [29], SVM [43] are good examples.

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 → read an image;
        im → resize(image);
        im → normalize(image);
        imfeatures → ExtractFeatures(CNNs(im));
    end
end
```

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 Cosine Similarity (CS) [41] and linear SVM classifiers [43]. 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.

The 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. The Euclidean dot product formula as in Eq. (1) can be used to compute the cosine similarity such that,

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

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

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

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

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

$$= \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}, \tag{4}$$

where, A and B are two vectors.

For classification, in all our experiments in this work, we compute the CS to find the minimum distance between the test image $test\_im$ and training images $training\_im^n$ by using Eqs. (5) and (6). The procedure for this classification is further illustrated in Fig. 8. Thus,

$$CS_{min} = min(dist(test\_im, training\_im^n)), \tag{5}$$

where n is a total number of images in the training set and,

$$CS_{dist}(test\_im, training\_im^n)$$
$$= \frac{\sum_{j=1}^{m} training\_im_j^i test\_im_j}{\sqrt{\sum_{j=1}^{n} training\_im_j^{i2}}\sqrt{\sum_{j=1}^{n} test\_im_j^2}}, \tag{6}$$

where m is a length of vector and $i = 1, \ldots, n$.

Similarly, SVM is a supervised machine learning algorithm which can be used for both binary classification and multi classification problems. The SVM focuses on identifying the margins via hyperplanes 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. It is clear that the SVM is geared

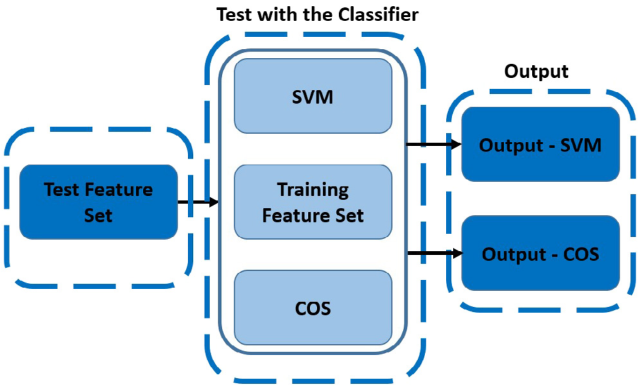

Fig. 8. Procedure adopted for the classification.

to solve binary classification problems. In all our experiments, we use the linear SVM to solve the multi-classes classification problem based on One-vs-One (OVO) approach [44]. This is also known as pairwise classification. The OVO decomposition constructs $n(n-1)/2$ binary classifiers for a given n number of classes. Then, for a final decision, the Error Correcting Codes (ECC) combination approach [45,46] decides how the various classifiers can be combined.

Consider that we have a training dataset (x_i, y_i), then we can use the linear SVM as in Eq. (7) such that,

$$\min_{w \in R^d} \frac{1}{2}\|w\|^2 + C\sum_i^N max(0, 1 - y_i w^T x_i), \tag{7}$$

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

As for the type of SVM (between linear and non-linear kernel), here we have chosen the linear SVM for the reasons as explained here. Prior to running the main experiments, we ran some trial experiments to test the accuracy of recognition results between linear SVM and non-linear SVM (with radial basis function kernels) to test their accuracy and efficiency. We found, in general, the linear SVM works well when small parts of the face are used as probes. For instance, in one trial experiment involving faces from 60 subjects, we found that for the right cheek, the linear SVM produced a recognition rate of 24.44% while the kernel SVM with radial basis functions only produced a rate of 2.77%. In addition to this, in general, the linear SVM is computationally more efficient in all the trial experiments we ran. Thus, we concluded that the kernel SVM with its marginal gains (only in larger parts of the face such as half or 3/4 face) does not lend overall additional advantages. Hence, we made use of the linear SVM throughout the rest of the experiments.

4. Experiments and results

Here, we present a comprehensive set of experiments we have conducted on face recognition using different parts of the face. To undertake this work, we have utilised face images from two popular face datasets, namely, the FEI [47] and LFW [48] dataset. All images in both databases were cropped to remove the background as much as possible using a cascade object detector in order to extract the face and the internal facial features [49]. However, for some images, with very complex backgrounds, as in the case of the LFW database, we cropped those faces manually. In this work, numerous settings of occlusion have been carried out in order to verify that our methodology can handle the normal and occluded face recognition tasks. For that purpose, we conducted two main sets of experiments — one which does not use the partial, rotated and zoomed face as part of the training face data and the other in which partial, rotated and zoomed faces have been utilised as part of the training data. In each case, we undertook 14 sub experiments involving the partial, rotated and zoomed out faces using both classifiers. For training purposes, 70% of the images per subject 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.

Fig. 9. Sample face data from the FEI dataset.

4.1. The FEI dataset

This database 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 in colour and are taken against a white homogeneous background. The subjects are between 19 and 40 years old. The dataset contain images with variations in facial expressions as well as the pose. Fig. 9 shows some sample images from the FEI face dataset.

4.2. Experiments on parts of the face using the FEI dataset

In our experiments, using part of the face using the FEI dataset, 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, the bottom half of the face, the top half of the face, right half and three quarters of the face as well as the full face. Fig. 10 shows the parts of the face we have used for testing the recognition rates.

After extracting features from the VGGF model, the CS without parts (CS-Wo) and the linear SVM (with 19 900 binary classifiers) without parts (SVM-Wo) were applied in order to investigate the rate of recognition for each facial part separately. The results of these experiments are summarised in Fig. 11. As it can be inferred from the graph, the highest rate of recognition is achieved with the full face and the three-quarter of the face with the recognition rate of 100% using both classifiers. However, the recognition rate starts to drop down slightly at the right half and the top half of the face respectively with SVM-Wo, but in the case of CS-Wo, 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

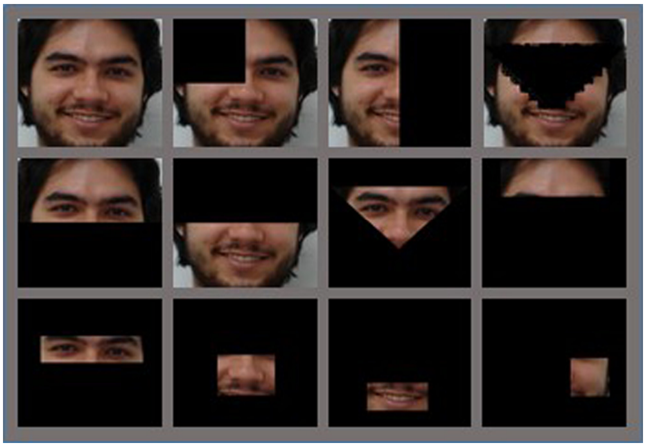

Fig. 10. Parts of the face we have used for testing the recognition rates on the FEI dataset.

of SVM-Wo and about 60% for CS-Wo. This decline continues until nearly 0.5% on the right cheek.

In contrast to that, in order to measure the rate of recognition 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 Fig. 11, 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-W and 40% for CS-W previously, and in this case, has improved to about 57% for SVM-W and 90% for CS-W. 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 for SVM-W. For example, a slight decrease in the recognition rate was observed for the bottom half of the face, which was 53% and dropped down to 51%. In contrast, the CS-W has produced a significant improvement, for instance, the recognition rate for combined eyes and nose increased from 40% to 90%.

4.3. Experiments on rotated faces using the FEI dataset

In this experiment, all the faces in test sets were rotated in eighteen degree increments, starting from 10° to 180°. In Fig. 12, we illustrate some sample rotations.

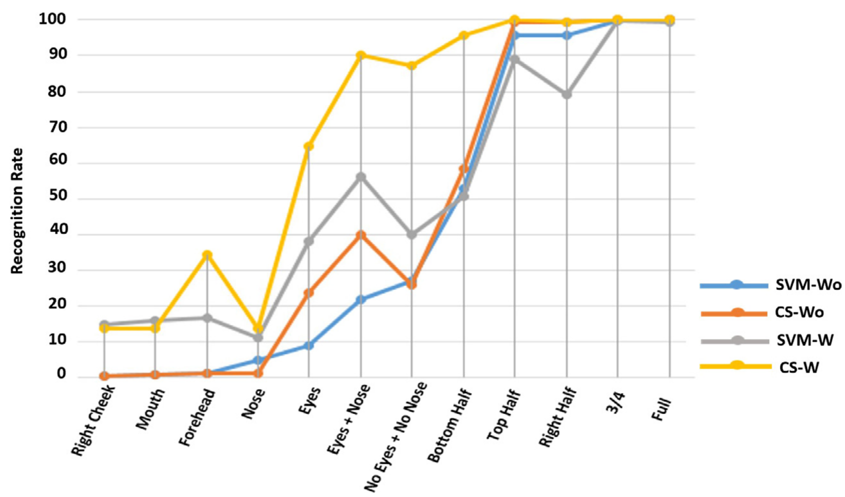

Fig. 11. Face recognition rates using SVM and CS classifiers based on parts of the face – without and with using individual facial parts of the face in training – on the FEI database.

Fig. 12. Illustration of face rotation (10° to 180°) on the FEI dataset.

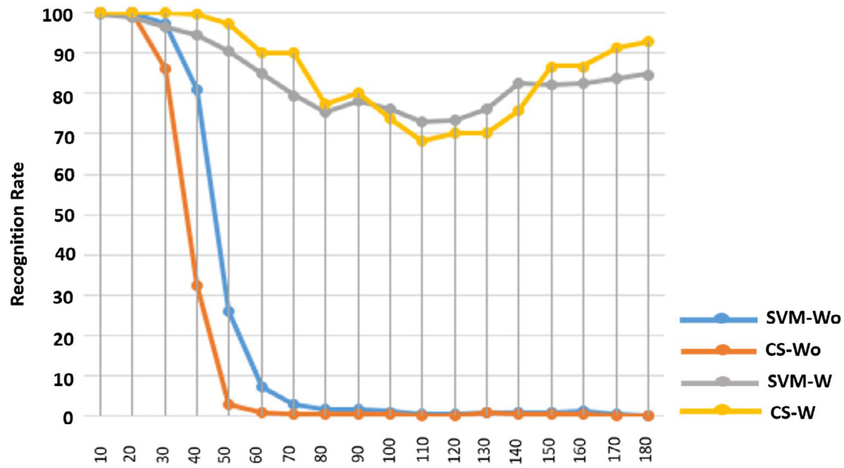

Fig. 13. Face recognition rates using SVMs and CS classifiers – based on face rotations without and with rotated faces in the training set – on the FEI dataset.

Fig. 14. An example of zooming out (10% to 90%) of faces on the FEI dataset.

All rotated images in each subset were passed to the VGGF model to extract features and we followed the same procedure in the experiments as above. Fig. 13 shows the recognition rates for the rotations by using the two classifiers. Note, the experiments are carried forward with adding and without adding rotated facial data into the training set.

In the case of without adding rotated images (SVM-Wo and CS-Wo) to the training sets, it is clear that the rotated faces at 10° to 20° respectively show the highest recognition rate which is approximately 100% in both classifiers as shown in Fig. 13. On the other hand, the recognition rate starts falling down partially at 30° to 40° and the rate of recognition achieved are between about 98% and 80% for SVM-Wo and CS-Wo. However, the worst case of recognition begins when the degree of rotation becomes high (50° to 180°) where the rate of recognition reaches almost 0% in some cases for both the classifiers.

In the second situation, when the rotated images are added to the training set, we can observe that the recognition has improved significantly among all the rotated faces for both classifiers (SVM-W and CS-W) as shown in Fig. 13. The CS-W recorded the highest recognition rate in most of the cases. For example, at 40° the previous rate was nearly 33% and it has enhanced to an impressive recognition rate of about 95% using CS-W. Using the SVM-W also it has gradually increased the recognition rate especially for higher degrees of rotation. For example, for 80° of rotation, without rotated data being added to the training dataset, the rate reaches from 2% to 76%. As the rotation increased, the rate of recognition became very low for SVM-W and CS-W while these rates have gone up dramatically from about 0% to between 82% and 84% for SVM-W and from nearly 0% to about 92% in CS-W, which again indicate that CS-W outperforms SVM-W.

4.4. Experiments on zoomed out faces using the FEI dataset

In this experiment, we zoomed out all the faces in the test sets from 10% to 90% in order to find out the effect of zooming on the rate of recognition. Here, we have nine groups of test images as shown in Fig. 14. Similar to what has been adopted for rotation experiments, the zoomed test faces were passed to the VGGF model in order to extract features and later passed them to the two classifiers (SVM and CS) for classification.

In the first part of this experiment, we evaluated the recognition rate in the case without adding zoomed out faces into the training set (SVM-Wo and CS-Wo). As we can see, in Fig. 15, the higher recognition rates were reached at 10% to 50% zooming levels by using SVM-Wo which about 100%. On the other hand, when the faced are zoomed out between 70% and 90% the recognition rates went down significantly ultimately reaching to approximately 0%. Contrary to that, in case of using CS-Wo, the recognition at zooming levels of between 10% and 50% produced lower recognition rates than SVM-Wo which is between 97% and 86%. Moreover, the recognition between 70% and 90% still is a remarkable improvement compared to the results for SVM-Wo.

As we added the zoomed out images into the training set, it became clear that there is a slight improvement in the recognition rates in the case of using SVM-W at zooming levels for 40% and 50% with 70%, 80% and 90% still steady, as shown in Fig. 15. The CS-W reached a higher rate of recognition at zoomed out levels from 10% to 60% where the recognition rate reaches 100%. Additionally, with 70% and 80% zooming out levels, we can notice a gradual increase in the recognition rate, for instance, from around 45% to about 52%. However, at the zooming out level of 90%, the recognition rates drop off to about 45% to 36%.

4.5. The LFW dataset

Labelled Faces in the Wild (LFW) [48] is a large dataset of face pictures designed for testing the capability of face recognition in simulating uncontrolled scenarios. All the images in the dataset have been collected from the Internet. The dataset has faces with large variations in expression, pose, age, illumination as well as resolution. LFW has 5749 subjects and a total of 13 000 images. The number of images per individual is not constant and about 4070 subjects have just one image. As shown in Fig. 16, the sample shows that the images have significant background clutter. For this reason, for our experiments on the LFW dataset, we have done some pre-processing to extract the face from the original image.

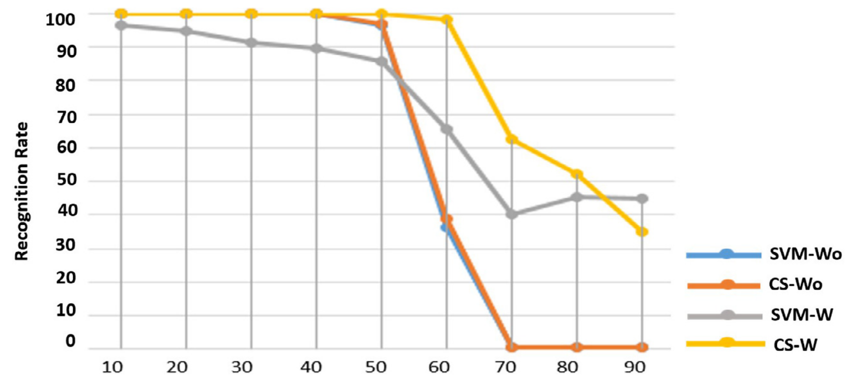

Fig. 15. The rate of recognition for zoomed out faces using SVM and CS classifiers on the FEI database.

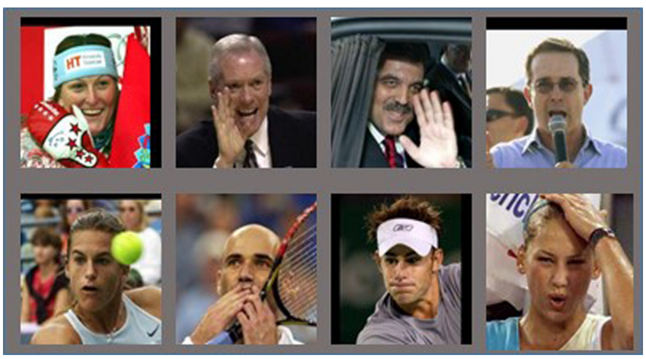

Fig. 16. Some sample face images from the LFW dataset.

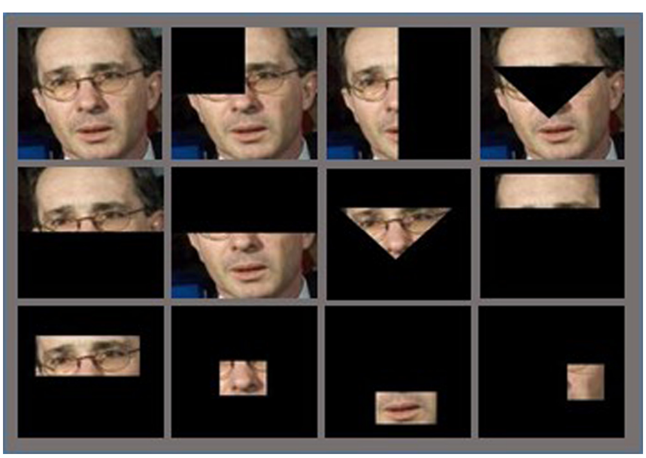

Fig. 17. A samples of facial parts from the LFW database.

In these experiments, using the LFW dataset, we followed the same procedures as with the FEI database, as above. Thus, two main sets of experiments were conducted — without and with parts for rotation and zooming out. In total we conducted, 14 sub experiments using both the SVM and CS classifiers.

4.6. Experiments on parts of the face on the LFW dataset

In the case of the partial face experiments, we followed the same procedures as we did for the FEI dataset to generate 12 datasets for our experiments. Fig. 17 illustrates some samples of the partial faces we utilised.

All the extracted features for partial faces from the VGGF model were passed to both the classifiers (SVM and CS), in both cases namely without parts for training (SVM-Wo and CS-Wo) and with parts for training (SVM-W and CS-W). In order to investigate the recognition rates for each facial part, we applied the classifiers separately. In the case of without adding parts into the training, it is clear that in a general CS-Wo outperforms the SVM-Wo for most of the parts of the face. By looking at Fig. 18, we can observe that the for the right cheek, mouth, forehead and the nose have there exist low levels of discrimination, with about 1% for both the classifiers. In contrast to that, the rate of recognition increases significantly for the parts containing the eyes, which reaches to about 40% using CS-Wo. As we increase the proportion of the face, the recognition rate improves significantly reaching nearly 100% for the 3/4 face and for the full face. Again, we note, the CS outperforms the SVM in all cases.

Similar to the previous experiments, we repeated all the tests on the facial parts and using the same classifiers but this time we added the facial parts to the training sets (SVM-W and CS-W). The results, as shown in Fig. 18, indicate there are marked improvements in the recognition rates by using SVM-W. For instance, the rate of recognition for the right cheek was about 1% and now reaches almost 10%. Regarding the mouth, forehead and the nose also we observe slight improvements in the recognition rates for both classifiers. As we increase the proportion of the face, the recognition rates significantly improve with CS-W which reaches to 70% instead of about 42% in the eyes, but this improvement did not occur with SVM-W where it is almost 7%. Furthermore, for the faces with occluded eyes and nose, bottom, top and right half, the rate of recognition enhanced significantly with CS-W, but it has a slight decrease for the 3/4 of face, from about 94% to around 93.5%.

4.7. Experiments on rotated face on the LFW dataset

For experiments, using rotated faces, similar to the experiments we ran on the FEI dataset, all faces in the LFW test sets were rotated in eighteen degree increments. After extracting features, we used the same methodology to find out the rate of recognition subject to facial rotation. Again, we trained the CNN both without rotated images and with them.

As shown in Fig. 19, the results show that in the first situation, without the rotated faces, the rate of recognition recorded is higher for rotated faces of 10° to 30° where the recognition rates were around the 98% mark for SVMs-Wo. Furthermore, for this case, using the CS-Wo, the results are even better. On the other hand, as the rotation increases, the proportion of recognition drops significantly for both the classifiers, which were approximately 55% for SVM-Wo and 58% for CS-Wo. We can also see that for rotations between 60° and 180°, the recognition rate falls to almost 0% for both the classifiers.

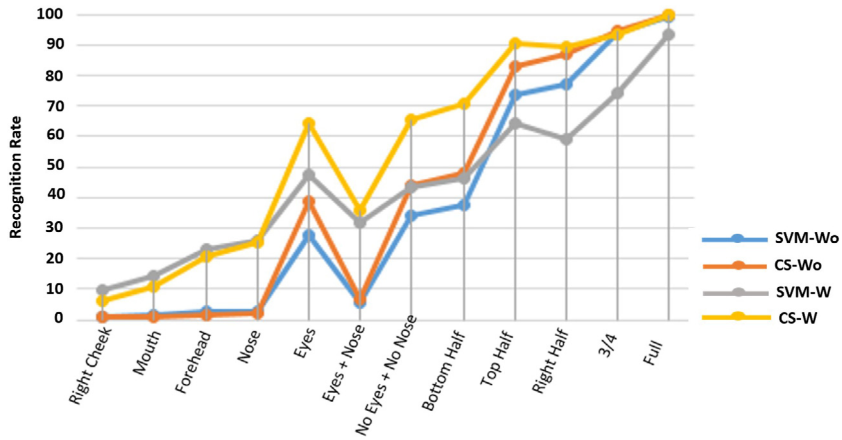

Fig. 18. Recognition rates based on parts of the face using both the classifiers, SVM and CS, without and with using individual facial parts of the face in the training, on the LFW dataset.

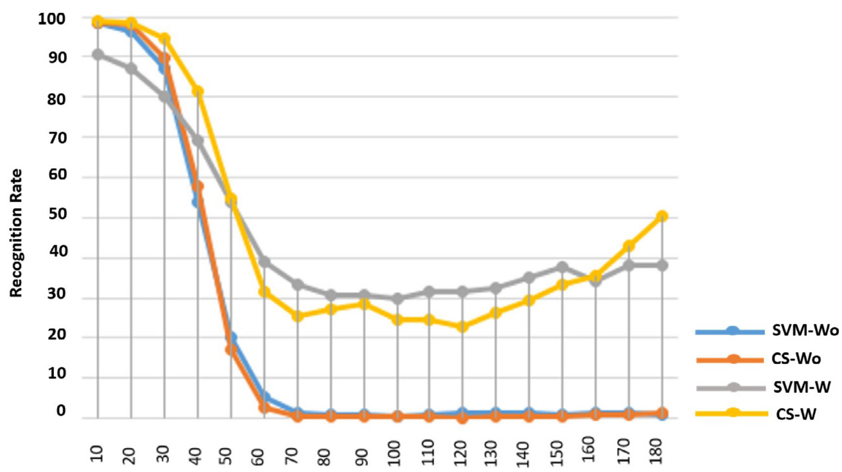

Fig. 19. Face recognition rates using SVM and CS classifiers based face rotations – without and with using individual rotated face as training data – on the LFW dataset.

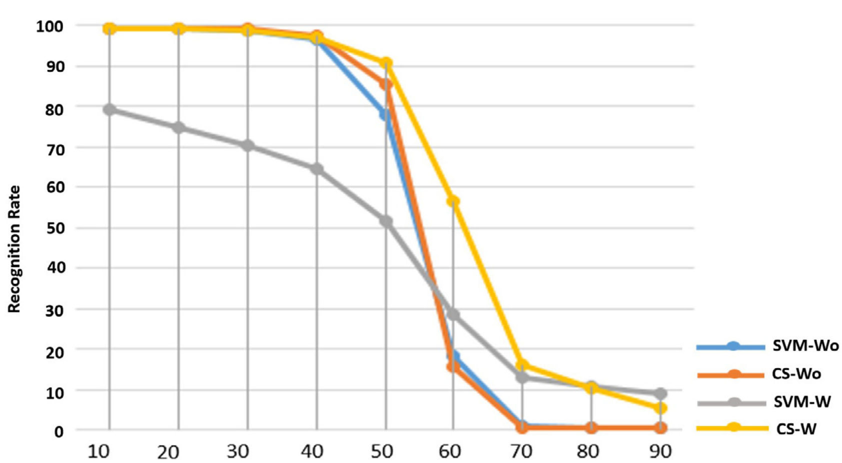

Fig. 20. The recognition rates for zoomed in images based on SVM and CS classifiers on the LFW database.

In the second case, once the rotated faces were added to the training set, the rate of recognition, in general, has improved for both classifiers, i.e., for SVM-W and CS-W. For example, the recognition rates at 150° which was initially at about 0.5% has improved to about 37%.

4.8. Experiments on zoomed out faces on the LFW dataset

Similarly, here in this experiment, again all the faces in the test sets were zoomed out from 10% to 90% in order to find out how the zooming out can affect the rate of recognition. In Fig. 20, we show the results of these experiments. There, we observe that

Fig. 21. The result of correct matching using CS, for parts of the mouth.

Fig. 22. The result of incorrect matching using the CS measure, for parts of the mouth.

Fig. 23. The result of correct matching using the CS measure, for the right cheek.

the rate of recognition is higher for zoomed out images for cases between 10% until 50% where the recognition rates range from 85% to 100% in the cases of SVM-Wo and CS-W. On the contrary, this rate has fallen sharply to almost 0% after 50% zoom out, both for SVM-Wo and CS-W. Again, once the zoomed faces are added to the training set, we note that the performance of the CS-W also improves further.

5. Discussions

From what we have presented for the partial facial experiments using the controlled FEI dataset, in the case of without adding the various parts into the training set, the highest recognition rate observed was for the 3/4 face, where we found the recognition rate to be 100% by using SVM-Wo. Also, in the case of CS-Wo the right half, the top half as well as the 3/4 face returned the recognition rate of 100%. However, the worst recognition rates observed are for the smaller and perhaps less significant parts of the face. By applying the same methodology to the uncontrolled LFW dataset, even bigger proportions of the face had slight decrease in the recognition rates compared with the FEI dataset which is between 76% to 99% for SVM-Wo and 83% to 99% for CS-W. Besides, according to the results obtained for smaller proportions of the face, the worst case observed is for the cheeks, mouth, forehead and nose but as far as recognition is concerned, the eyes appear to hold more information.

In the second part of the experiments, where we added the individual parts of the face to the training sets, the rate of recognition with the partial face improved dramatically, particularly for recognition using smaller proportions of the face. For example, the recognition rate for the right cheek improved from 0% to 15%

Table 1
The comparison of recognition results with He et al. [21] and ours on the LFW database.

Area	He et al. [21]	Ours
Upper	39.2%	90.2%
Down	7.8%	70.73%
Right	24.2%	88.9%
Left	27.6%	–

on the FEI dataset. We also note that the eyes still have the highest recognition rates among the other individual parts for both the FEI and LFW datasets. Whereas the combined eyes and nose recorded around 90% in the recognition rate in the controlled FEI dataset, opposite to that, in the case of the uncontrolled LFW dataset, this percentage dropped slightly. Furthermore, we note that better recognition results overall can be achieved by using the CS measure.

Thus, an important point to highlight here is that the CS measure, in general, appears to be a better classifier in this case, compared to both linear and non-linear SVMs. In the case of SVMs, they require complete re-training when new data are added which obviously has computational issues to deal with. However, in the case of the CS classifier this is not necessary. Though having said that, in the testing stage, the CS classifier can be more computationally intensive but given the greater degree of accuracy it provides it makes sense to utilise the CS classifier over SVMs in this case.

In order to compare our results with the state of the art techniques, we note that He et al. [21], presented work somewhat similar to what is presented in this paper, whereby they applied the recognition task to parts of the face by using data from LFW dataset. This is the closest work we have come across in the domain of partial face recognition by which we can make some form of direct comparison with our results. In Table 1, we show their results which we compare with ours. As one can clearly see, in this case, our results are significantly better.

From the results of the experiments we have undertaken, we can make further observations about the accuracy of classifiers between CS and SVM, in that, in general, the CS outperforms the SVM. In addition, we can observe at the individual class level by considering the matching images picked by the classifiers. For example, in Fig. 21 we present the images that were matched correctly for a subject by using CS-W and in Fig. 22, we show a case in which the CS classifier got confused.

In some cases, we also observed that even greater matching performance demonstrated by the CS classifier. For example, in Fig. 23 we show the images of cheeks being correctly matched by the CS classifier, though in this case, the classification may have resulted from the prominent mark on the cheek of the face of the individual, as shown in Fig. 23.

An interesting question that may come to one's mind may be on the generalisation capability of our proposed approach, for example, by eliminating one or two facial parts within the training sets. To investigate this, we ran some further experiments whereby we randomly left out two facial parts – i.e. the right cheek, no eyes + no nose – from the training data on the LFW database. We ran the experiments, without these two parts, as outlined previously. We utilised the linear SMV (L-SVM), kernel SVM (k-SVM) and CS for classification. The results of these experiments are summarised in Fig. 24. As we can see in Fig. 24, for the two removed parts, the rate of recognition falls drastically, for example, from about 25.55% to around 3.88% in the cheek for CS. In addition, for the part without eyes and nose, the recognition rate fell by almost half for all classifiers, and the k-SVM performed worst in this case. As far as recognition, in this experiment, for other parts are concerned, in general, the rates for SVMs fell slightly whereas the CS appears to be maintaining the recognition rates.

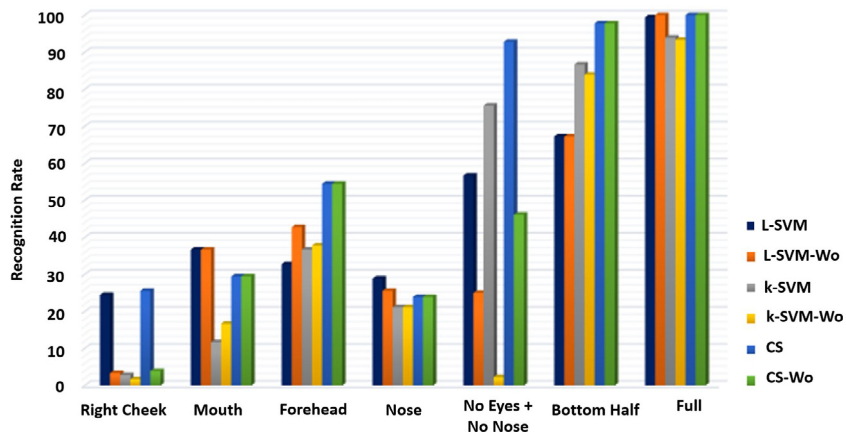

Fig. 24. The results of the recognition task when some parts of the face (cheeks and the part with no eyes and no nose) are removed from the training sets. Classifiers used here are linear SVM, kernel SVM and CS.

6. Conclusions

The ability for existing machine based face recognition algorithms to perform adequately in the cases of imperfect facial data – such as occluded faces, rotated faces or zoomed out faces – as cues remains a challenging task in the field of computer vision and visual computing. In this work, we have presented the results of some novel experiments we have undertaken to highlight these issues as well as to outline some potential solutions. To do this, we have utilised both controlled and uncontrolled public facial datasets through which we show how deep learning can be utilised for face recognition using imperfect facial cues. Thus, given some partial facial data, we show how feature extraction can be performed using popular CNNs such as the VGGF model. We show how classifiers based on popular SVMs as well as CS can be utilised to undertake facial recognition tasks.

In this paper, we have discussed a rather comprehensive set of experiments for face recognition using imperfect facial data. Our results show that as the proportion of the face gets smaller, regardless of the prominent nature of the facial features such as the eyes, nose or the mouth contained in it, the recognition rate appears to perform poorly. However, we note, even in the case of machine based face recognition, the eyes appear to carry more recognition cues compared to other individual facial features. Furthermore, when it comes to rotated faces, we note that it would be far better to avoid highly rotated faces, e.g. faces rotated between 110° and 120°, as they appear to be performing very poorly in recognition tasks, regardless of incorporating rotated faces as part of the training data. In the case of zoomed out faces, again it is advisable not to use highly zoomed out faces, e.g. faces zoomed out to 70% to 90% as probes. Finally, we note that the CS measure greatly improves the performance of the classification when compared to both the linear and kernel SVMs.

From an application point of view, we believe this work is still preliminary in that we have only utilised datasets which are somewhat controlled and far removed from practical scenarios. Therefore, it will be very useful to extend this work to assess its practical applicability in terms of extending our experiments where, for example, real CCTV footage of faces may be used as recognition cues.

Further, in this work, we have simply shown the capacity of CNNs, in particular, the use of VGGF for facial feature extraction and analysis of imperfect facial data. There exist a whole host of other methods and techniques that can be utilised to develop as well as to test facial recognition cues using imperfect facial data. One such mechanism would be the use of general adversarial networks (GANs) which are increasingly becoming popular as a tool for machine learning, which, for example, comparatively speaking would require a minimal dataset for training a neural network.

Acknowledgements

This work was supported in part by the European Union's Horizon 2020 Programme H2020-MSCA-RISE-2017, under the project PDE-GIR with grant number 778035.

Conflict of interest statement

None.

Declaration of conflicting interests

There's no conflict of interest to be declared.

References

[1] S.M. Andersen, C.A. Carlson, M.A. Carlson, S.D. Gronlund, Individual differences predict eyewitness identification performance, Pers. Individ. Differ. 60 (2014) 36–40, http://dx.doi.org/10.1016/j.paid.2013.12.011.

[2] J.W. Tanaka, M.D. Kaiser, S. Hagen, L.J. Pierce, Losing face: impaired discrimination of featural and configural information in the mouth region of an inverted face, Atten. Percept. Psychophys. 76 (4) (2014) 1000–1014, http://dx.doi.org/10.3758/s13414-014-0628-0.

[3] Z. Li, J. Liu, J. Tang, H. Lu, Robust structured subspace learning for data representation, IEEE Trans. Pattern Anal. Mach. Intell. 37 (10) (2015) 2085–2098, http://dx.doi.org/10.1109/TPAMI.2015.2400461.

[4] C. Ding, D. Tao, Pose-invariant face recognition with homography-based normalization, Pattern Recognit. 66 (2017) 144–152, http://dx.doi.org/10.1016/j.patcog.2016.11.024.

[5] C. Ding, D. Tao, Robust face recognition via multimodal deep face representation, IEEE Trans. Multimed. 17 (11) (2015) 2049–2058, http://dx.doi.org/10.1109/TMM.2015.2477042.

[6] T. Zhang, J. Li, W. Jia, J. Sun, H. Yang, Fast and robust occluded face detection in ATM surveillance, Pattern Recognit. Lett. 107 (2018) 33–40, Video Surveillance-oriented Biometrics. http://dx.doi.org/10.1016/j.patrec.2017.09.011.

[7] D. Marčetić, S. Ribarić, Deformable part-based robust face detection under occlusion by using face decomposition into face components, in: 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 2016, pp. 1365–1370. http://dx.doi.org/10.1109/MIPRO.2016.7522352.

[8] I. Choi, D. Kim, Facial fraud discrimination using detection and classification, in: G. Bebis, R. Boyle, B. Parvin, D. Koracin, R. Chung, R. Hammound, M. Hussain, T. Kar-Han, R. Crawfis, D. Thalmann, D. Kao, L. Avila (Eds.), Advances in Visual Computing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 199–208, http://dx.doi.org/10.1007/978-3-642-17277-9_21.

[9] M.A. Alsmirat, F. Al-Alem, M. Al-Ayyoub, Y. Jararweh, B. Gupta, Impact of digital fingerprint image quality on the fingerprint recognition accuracy, Multimedia Tools Appl. 78 (3) (2019) 3649–3688, http://dx.doi.org/10.1007/s11042-017-5537-5.

[10] S. Atawneh, A. Almomani, H. Al Bazar, P. Sumari, B. Gupta, Secure and imperceptible digital image steganographic algorithm based on diamond encoding in DWT domain, Multimedia Tools Appl. 76 (18) (2017) 18451–18472, http://dx.doi.org/10.1007/s11042-016-3930-0.
[11] A.M. Bukar, H. Ugail, Automatic age estimation from facial profile view, IET Comput. Vis. 11 (8) (2017) 650–655, http://dx.doi.org/10.1049/iet-cvi.2016.0486.
[12] H. Ugail, A. Al-Dahoud, Is gender encoded in the smile? A computational framework for the analysis of the smile driven dynamic face for gender recognition, Vis. Comput. 34 (9) (2018) 1243–1254, http://dx.doi.org/10.1007/s00371-018-1494-x.
[13] B. Olk, A.M. Garay-Vado, Attention to faces: Effects of face inversion, Vis. Res. 51 (14) (2011) 1659–1666, http://dx.doi.org/10.1016/j.visres.2011.05.007.
[14] J. Taubert, D. Apthorp, D. Aagten-Murphy, D. Alais, The role of holistic processing in face perception: Evidence from the face inversion effect, Vis. Res. 51 (11) (2011) 1273–1278, http://dx.doi.org/10.1016/j.visres.2011.04.002.
[15] J.J. Richler, M.L. Mack, T.J. Palmeri, I. Gauthier, Inverted faces are (eventually) processed holistically, Vis. Res. 51 (3) (2011) 333–342, http://dx.doi.org/10.1016/j.visres.2010.11.014.
[16] H. De, Adélaïde, R. Bruno, M. Daphne, Developmental changes in face recognition during childhood: Evidence from upright and inverted faces, Vis. Res. 27 (1) (2011) 17–27, http://dx.doi.org/10.1016/j.cogdev.2011.07.001.
[17] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Maui, HI, USA, USA, pp. 586–591. http://dx.doi.org/10.1109/CVPR.1991.139758.
[18] A.M. Aguilera, M. Escabias, M.J. Valderrama, Using principal components for estimating logistic regression with high-dimensional multicollinear data, Comput. Statist. Data Anal. 50 (8) (2006) 1905–1924, http://dx.doi.org/10.1016/j.csda.2005.03.011.
[19] M. Savvides, A. Ramzi, H. Jingu, S. Park, X. Chunyan, B.V.K. Vijayakumar, Partial and holistic face recognition on FRGC-II data using support vector machine kernel correlation feature analysis, in: Computer Vision and Pattern Recognition Workshop, New York, USA, 2006. http://dx.doi.org/10.1109/CVPRW.2006.153.
[20] A. Sleit, R. Abu-Hurra, W. AlMobaideen, Lower-quarter-based face verification using correlation filter, J. Imaging Sci. 59 (1) (2011) 41–48, http://dx.doi.org/10.1179/136821910X12863757400286.
[21] L. He, H. Li, Q. Zhang, Z. Sun, Dynamic feature learning for partial face recognition, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018. http://dx.doi.org/10.1109/CVPR.2018.00737.
[22] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015. http://dx.doi.org/10.1109/CVPR.2015.7298965.
[23] P. Omkar, V. Andrea, Z. Andrew, Deep face recognition, in: X. Xianghua, J. Mark, T. Gary (Eds.), Proceedings of the British Machine Vision Conference (BMVC), BMVA Press, 2015, pp. 41.1–41.12, http://dx.doi.org/10.5244/C.29.41.
[24] L. Yang, Z. Fan, S. Ling, H. Junwei, Face recognition with a small occluded training set using spatial and statistical pooling, Inform. Sci. 430–431 (2018) 634–644, http://dx.doi.org/10.1016/j.ins.2017.10.042.
[25] L. Badr, L. Syaheerah, V. Ibrahim, A.-B. Mohammed, S.-S. Rubn, Optimized symmetric partial facegraphs for face recognition in adverse conditions, Inform. Sci. 429 (2018) 194–214, http://dx.doi.org/10.1016/j.ins.2017.11.013.
[26] Y. Duan, L. Jiwen, F. Jianjiang, Z. Jie, Topology preserving structural matching for automatic partial face recognition, IEEE Trans. Inf. Forensics Secur. 13 (7) (2018) 1823–1837, http://dx.doi.org/10.1109/TIFS.2018.2804919.
[27] J. Cai, J. Chen, X. Liang, Single-sample face recognition based on intra-class differences in a variation model, Sensors 15 (1) (2015) 1071–1087, http://dx.doi.org/10.3390/s150101071.
[28] H. Li, Y. Ching, Robust face recognition based on dynamic rank representation, Pattern Recognit. 60 (2016) 13–24, http://dx.doi.org/10.1016/j.patcog.2016.05.014.
[29] Z. Zhang, Introduction to machine learning: k-nearest neighbors, Ann. Transl. Med. 4 (11) (2016) 218, http://dx.doi.org/10.21037/atm.2016.03.37.
[30] X. Peng, L. Zhang, Y. Zhang, K.K. Tan, Learning locality-constrained collaborative representation for robust face recognition, Pattern Recognit. 47 (9) (2014) 2794–2806, http://dx.doi.org/10.1016/j.patcog.2014.03.013.
[31] Y. Pan, J.L. Trahan, R. Vaidyanathan, A scalable and efficient algorithm for computing the City Block distance transform on reconfigurable meshes, Comput. J. 40 (7) (1997) 435–440, http://dx.doi.org/10.1093/comjnl/40.7.435.
[32] J. Murphy, R. Cook, Revealing the mechanisms of human face perception using dynamic apertures, Cognition 169 (2017) 25–35, http://dx.doi.org/10.1016/j.cognition.2017.08.001.
[33] M.G. Calvo, A. Fernández-Martín, L. Nummenmaa, Facial expression recognition in peripheral versus central vision: Role of the eyes and the mouth, Psychol. Res. 78 (2) (2014) 180–195, http://dx.doi.org/10.1007/s00426-013-0492-x.
[34] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444, http://dx.doi.org/10.1038/nature14539.
[35] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252, http://dx.doi.org/10.1007/s11263-015-0816-y.
[36] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, http://dx.doi.org/10.1109/CVPR.2016.90.
[37] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017. http://dx.doi.org/10.1109/CVPR.2017.243.
[38] Y. Zhong, R. Arandjelović, A. Zisserman, Faces in places: Compound query retrieval, in: BMVC - 27th British Machine Vision Conference, York, United Kingdom, 2016. http://dx.doi.org/10.5244/C.30.56.
[39] M. Oquab, L. Bottou, I. Laptev, J. Sivic, Learning and transferring mid-level image representations using convolutional neural networks, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 2014. http://dx.doi.org/10.1109/CVPR.2014.222.
[40] A. Vedaldi, K. Lenc, MatConvNet: COnvolutional neural networks for MATLAB, in: Proceedings of the 23rd ACM International Conference on Multimedia, ACM, New York, NY, USA, 2015, pp. 689–692, http://dx.doi.org/10.1145/2733373.2807412.
[41] G. Sidorov, A. Gelbukh, H. Gmez-Adorno, D. Pinto, Soft similarity and soft cosine measure: Similarity of features in vector space model, Comput. Sist. 18 (2014) 491–504, http://dx.doi.org/10.13053/CyS-18-3-2043.
[42] S.K. Murthy, Automatic construction of decision trees from data: A multi-disciplinary survey, Data Min. Knowl. Discov. 2 (4) (1998) 345–389, http://dx.doi.org/10.1023/A:1009744630224.
[43] X. Han, H. Ugail, I.J. Palmer, Gender classification based on 3D face geometry features using SVM, in: 2009 International Conference on CyberWorlds, Bradford, UK, 2009, pp. 114–118. http://dx.doi.org/10.1109/CW.2009.41.
[44] A.C. Lorena, A.C.P.L.F. de Carvalho, J.M.P. Gama, A review on the combination of binary classifiers in multiclass problems, Artif. Intell. Rev. 30 (1) (2009) 19, http://dx.doi.org/10.1007/s10462-009-9114-9.
[45] M.A. Bagheri, G.A. Montazer, S. Escalera, Error correcting output codes for multiclass classification: Application to two image vision problems, in: The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012), 2012, pp. 508–513. http://dx.doi.org/10.1109/AISP.2012.6313800.
[46] A. Passerini, M. Pontil, P. Frasconi, New results on error correcting output codes of kernel machines, IEEE Trans. Neural Netw. 15 (1) (2004) 45–54, http://dx.doi.org/10.1109/TNN.2003.820841.
[47] C.E. Thomaz, G.A. Giraldi, FEI Face database, 2010, https://fei.edu.br/~cet/facedatabase.html.
[48] G.B. Huang, M. Mattar, T. Berg, E. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments, in: Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Marseille, France. URL https://hal.inria.fr/inria-00321923.
[49] A. Al-dahoud, H. Ugail, A method for location based search for enhancing facial feature detection, in: P. Angelov, A. Gegov, C. Jayne, Q. Shen (Eds.), Advances in Computational Intelligence Systems, Springer, Cham, 2017, pp. 421–432, http://dx.doi.org/10.1007/978-3-319-46562-3_27.

Ali Elmahmudi is a doctoral researcher at the Centre for Visual Computing in the Faculty of Engineering and Informatics at the University of Bradford, UK. He received his Masters degree in medical image segmentation in 2007 from The Libyan Academy. His research interests include image and video processing, pattern recognition and machine learning.

Hassan Ugail is the director of the Centre for Visual Computing in the Faculty of Engineering and Informatics at the University of Bradford, UK. He has a first class B.Sc. Honours degree in Mathematics from King's College London and a Ph.D. in the field of geometric design from the School of Mathematics at the University of Leeds. Professor Ugail's research interests include computer based geometric and functional design, imaging and machine learning.