Face Presentation Attack Detection using Biologically-inspired Features

Aristeidis Tsitiridis, Cristina Conde, Isaac Martín De Diego and Enrique Cabello
Computer Science and Statistics, King Juan Carlos University, Calle Tulipán S/N, 28933, Mostoles, Madrid, Spain
tsitiridis@gmail.com, {cristina.conde, isaac.martin, enrique.cabello}@urjc.es

Keywords: Face Biometrics, Presentation Attack Detection, Anti-Spoofing, Biologically-inspired Verification, Biologically-inspired Biometrics.

Abstract: A person intentionally concealing or faking their identity from biometric security systems is known to perform a ‘presentation attack’. Efficient presentation attack detection poses a challenging problem in modern biometric security systems. Sophisticated presentation attacks may successfully spoof a person’s face and therefore, disrupt accurate biometric authentication in controlled areas. In this work, a presentation attack detection technique which processes biologically-inspired facial features is introduced. The main goal of the proposed method is to provide an alternative foundation for biometric detection systems. In addition, such a system can be used for future generation biometric systems capable of carrying out rapid facial perception tasks in complex and dynamic situations. The newly-developed model was tested against two different databases and classifiers. Presentation attack detection results have shown promise, exceeding 94% detection accuracy on average for the investigated databases. The proposed model can be enriched with future enhancements that can further improve its effectiveness and complexity in more diverse situations and sophisticated attacks in the real world.

1 INTRODUCTION

Present-day security systems exploit a variety of biological characteristics to identify individuals. There is an extensive range of security applications that utilise such characteristics to safeguard and restrict access or control. For example, non-intrusive biometric patterns extracted from the finger, palm, iris, voice, gait and their fusion in multimodal biometric systems, can provide an abundance of information about the identity of a person. However, none of these metrics are as informative, widely adopted or as publicly accepted as an individual’s face. Appearance perception and in particular facial, plays a vital role in survival and everyday social interactions. As a consequence, the human brain over millions of years has evolved to perform facial perception in an effortless, rapid and efficient manner (Ramon et al. 2011). Face perception is a challenging problem due to appearance variations from illumination, pose, facial expressions, aging, clothing accessories and temporal facial changes. It remains a widely explored and efficient method, applied in diverse environments without necessitating substantial participation and inconvenience from any individuals. Modern day applications making extensive use of facial perception tasks include mobile phone authentication, border or customs control, visual surveillance and human-computer interaction. However, the ever increasing complexity, power and processing speed has been pushing the biometric research community to explore new ways of optimising face biometric systems. Therefore, it should not come as a surprise that biology has recently become an increasingly valuable source of inspiration for reliable, power efficient and alternative methods (Meyers and Wolf 2008; Wang et al. 2013).

‘Face recognition’ is a generic term that describes a set of methods concerning face perception. More specifically, face verification applies various image processing methods in order to confirm an individual’s identity according to their travel documents and either accept or reject it. Face identification focuses on identifying a particular individual and the subject’s face is compared against other individuals from a pre-stored database. The present research work centres on face verification.

Every person’s face is continuously evolving in

Tsitiridis A., Conde C., MartÃn De Diego I. and Cabello E.
Face Presentation Attack Detection using Biologically-inspired Features.
DOI: 10.5220/0006124603600370
In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pages 360-370
ISBN: 978-989-758-225-7

complex ways over time and different facial features may unintentionally alter the appearance in a significant way e.g. scars, glasses, piercings, hairstyles, wrinkles, tattoos. More importantly for the purposes of this work, face biometric systems are susceptible to intentional presentation attacks. Impostors can acquire another person's high quality facial image print with small digital cameras. Such cameras help to discretely or secretively capture facial images from unsuspected individuals. Moreover, with the availability of face images from public or social media, it has become relatively easy to reproduce a person's frontal image. Basic presentation attacks usually are: a) printed face on a paper sheet. Sometimes a printed face is shown with eyes cropped out in order for a subject's eyes to blink underneath it. b) Digital face displayed on a screen from digital devices such as tablets, smartphones, and laptops. This kind of face presentation can be static or a video. In video attacks facial movements, eye blinking, mouth/lip movements or expressions are simulated. c) A 3D mask (paper, silicon, cast, rubber etc) specifically moulded for a targeted face.

In addition to the above, an impostor may also try spoofing an identity by using more sophisticated appearance alteration techniques or their combinations: 1) Glasses corrective or otherwise and/or contact lenses with possible colour change. 2) Hairstyle, change in colour, cut/trim, hair extensions etc. 3) Make-up or fake facial scars. 4) Real and/or fake facial hair. 5) Facial prosthetics and/or plastic surgery.

All of these elaborate impostor attacks are commonly known as 'Face presentation attacks or Spoofing Attacks' and give rise to another term 'Presentation Attack Detection (PAD)' which includes the detection of all intentional impostor attempts in face verification and identification with the use of algorithm specifically tailored for this problem. Accurate and fast PAD is a major concern in authentication systems across many platforms and applications. Finally, high PAD rates are extremely important for matters of personal and public security.

2 RELATED WORK

Presentation attacks in images are usually detected from motion, liveness, texture, quality and by spectral information from sensor-based approaches. Motion-based techniques are mostly employed with video sequences to detect motion anomalies between frames. Some representative methods apply Eulerian Video Motion Magnification (Wu et al. 2012), Optical Flow (Anjos et al. 2014), and non-rigid motion with face-background analyses fusion (Yan et al. 2012). Liveness-based approaches extract image features that investigate the liveness of a particular subject. Using this approach, algorithms scan liveness patterns of certain facial parts such as facial expressions, mouth or head movements, eye blinking, and facial vein maps (Pan et al. 2008; Chakraborty and Das 2014). Texture based methods investigate texture, structure and overall shape information of faces. Commonly used texture-based methods rely on Local Binary Patterns (Chingovska et al. 2012; Maatta et al. 2011; Kose et al. 2015), Difference of Gaussians (Zhang et al. 2012) and Fourier frequency analysis (Li et al. 2004). For quality characteristics, a notable image quality method in (Galbally et al. 2014) proposes 25 different image quality metrics as extracted between real and fake images in order to train classifiers which are used for detecting potential attacks. Sensor-based approaches involve the use and fusion of various sensors. A method that uses a light field camera sensor with 26 different focus measures with image descriptors and authors in (Raghavendra et al. 2015) have reported promising PAD scores. With the aid of infrared sensors authors in (Prokoski and Riedel 2002) analyse facial thermograms for rapid, and varied illumination environments. Similar thermography methods are presented in (Hermosilla et al. 2012; Seal et al. 2013).

In general, a biologically-inspired vision architecture consists of alternating hierarchical layers mimicking the various processing stages of the primary visual cortex (Hubel and Wiesel 1967). It is known that as visual stimuli travel up the cortical layers, visual information progressively exhibits a combination of selectivity and invariance to object translations such as size, position, rotation, depth etc. In the past, there have been many vision models and variants inspired from this layered approach such as the 'Neocognitron' (Fukushima et al. n.d.), 'Convolutional neural network' (LeCun et al. 1998), and 'Hierarchical model and X' (Riesenhuber and Poggio 2000). Over the years, these models produced remarkable results on a variety of different object perception tasks and today they are being recognised as an equal alternative to statistical approaches in computer vision. In face perception, biologically-inspired methodologies have been applied successfully for some years and have been proven reliable as well as accurate (Meyers and Wolf 2008; Rose 2006; Wang and

Chua 2005; Lyons et al. 1998; Slavkovic et al. 2013; Li et al. 2013; Wang et al. 2013; Perlibakas 2006; Pisharady and Martin 2012). There are many common characteristics between all of these algorithms and perhaps the most important aspect is the extensive use of Gabor filters i.e. texture-based features, as their building block operation. The main reasons for designing a biologically-inspired model would be its projected efficiency, parallelisation and speed in extremely demanding future biometric applications with uni-sensor and multi-sensor data. Contemporary state-of-the-art methods are efficient for current processes and sensors but would struggle to cope with an increase in available sensor information and sifting through each frame for example with sliding windows or pixel-by-pixel approaches would require an incredible amount of available resources in storage capacity, processing speed and power. Until today there hasn't been a complete biologically-inspired schema specifically developed to tackle sophisticated PADs in face verification systems. The main contributions of this research work were to:

a) Present a novel hierarchical biologically-inspired algorithm which behaves comparably to other state-of-the-art texture-based algorithms.
b) Introduce visual area V2 of the brain texture operations for face perception for the first time and integrate them with the existing biological-like approach.
c) Explore the model's applicability in face presentation attack detection using standard databases and classifiers.
d) Propose the basis for further research in this particular area.

2.1 Structure

There is undeniable evidence (D and D 1991; Van Kleef et al. 2010; Axelrod and Yovel 2012) concerning the layered hierarchical structure of the mammalian brain. The main advantage of this topology is the progressive creation of a view-invariant representation of objects with some important invariance properties such as size, position, rotation and illumination. Similarly, the structure of the model here follows an alternating layer configuration that pools bio-inspired features as extracted from face images (Figure 1).

The extracted features at higher dimensions can be either used with classifiers or directly with distance measures as will be demonstrated in the results section 5. Ultimately, the main purpose of this research is to detect face spoofing attempts and invariance properties such as size and position are important. Therefore on this particular pursuit, any additional invariance properties that would otherwise have been more meaningful for face recognition (Yokono and Poggio 2004; Pisharady and Martin 2012; Rolls 2012), add complexity or processing delay and are not explored.

2.2 Centre-surround Channels

Cone receptive fields in human retinae are tuned to different wavelengths of light. Bipolar retinal cells

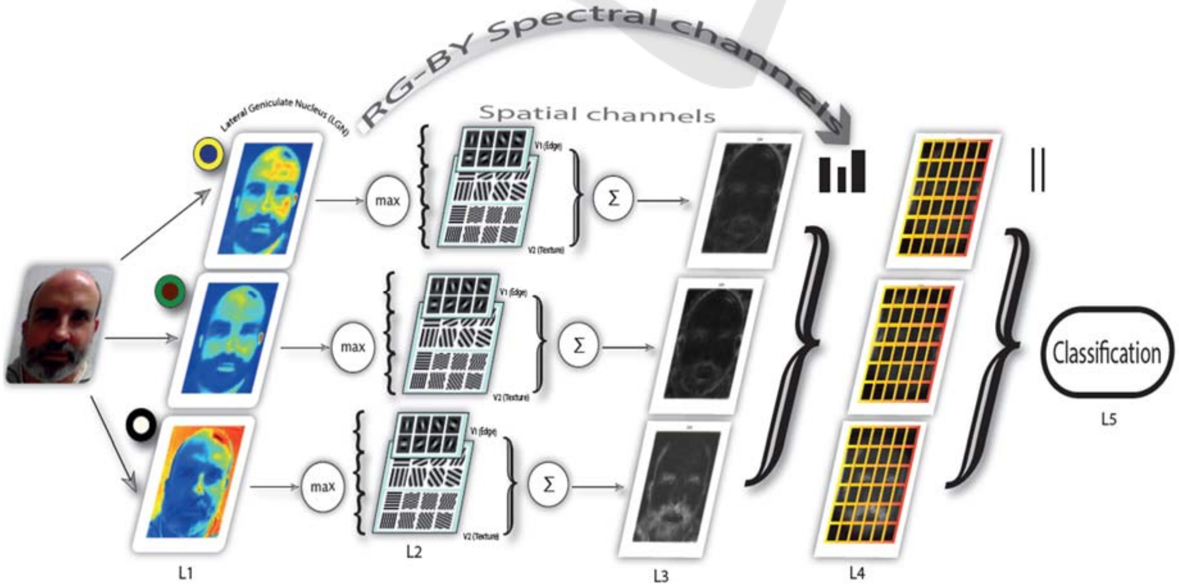

Figure 1: The proposed model structure. Several layers L1 to L5 progressively process spatial and spectral facial features.

bear the task of unifying incoming visual information from cones and rods (Engel et al. 1997). Furthermore, on-centre and off-centre bipolar cells operate in a centre-surround process between red-green and blue-yellow wavelengths. For example, on-centre Red-Green (RG) bipolar cells maximally respond when red hits the centre of their receptive field only and are inhibited when green is at their surrounding region. Vice versa, this operation is reversed for an off-centre RG bipolar cell where excitation only occurs when the detectable green wavelength is incident in the surrounding region. As shown in Figure 2, this is can be further applied for the blue-yellow and lightness channels.

The colour opponent space is defined by the following equations (Van De Sande et al. 2010):

$$O1 = (R-G)/\sqrt{2} \tag{1}$$

$$O2 = (R+G-2B)/\sqrt{6} \tag{2}$$

$$O3 = (R+G+B)/\sqrt{3} \tag{3}$$

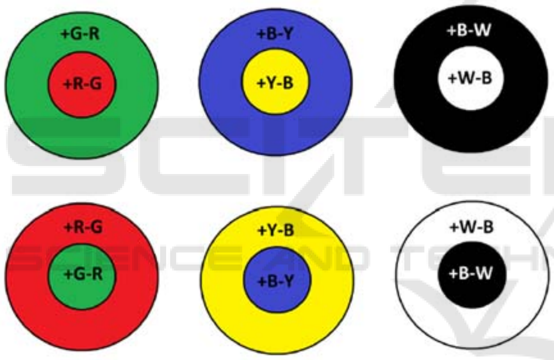

Figure 2: Examples of on-centre and off-centre receptive fields in bipolar cells for colour opponency channels. The plus sign refers to when the particular colour is on and the minus off.

2.3 Area V1 – Edge Detection

As visual signals travel to the primary visual cortex through the lateral geniculate nucleus, area V1 orientation selective simple cells process incoming information (Hubel and Wiesel 1967) from the retinae and perform basic edge detection operations for all subsequent visual tasks. They serve as the building block units of biological vision. It is already well established from literature that orientation selectivity in V1 simple cells can be precisely matched by Gabor filters (Marcelja 1980; Daugman 1985; Webster and De Valois 1985).

A Gabor filter is a linear filter which is defined as the product of a complex sinusoid with a 2D Gaussian envelope and for values in pixel coordinates *(x,y)*, it is expressed as:

$$G(x,y)=\exp\left(-\frac{X^2+\gamma^2Y^2}{2\sigma^2}\right)\cos\left(\frac{2\pi}{\lambda}X\right) \tag{4}$$

$$X = x\cos\theta – y\sin\theta \tag{5}$$

$$Y = -x\sin\theta+y\cos\theta \tag{6}$$

In equation (1), γ is the aspect ratio and in this work is set to 0.3. Parameter λ is known as the wavelength of the cosine factor and together with parameter σ the effective width, specify the spatial tuning accuracy of the Gabor filter. Ideally, to optimise the extraction of contour features from V1 units for a particular set of objects, some form of learning is necessary to isolate an optimum range of filters. However, this process adds complexity and it is time-consuming since it requires a huge number of samples, as experiments on convolutional neural networks have shown in literature. In order to avoid this step, Gabor filter parameters are hardcoded directly into our model following parameterisation sets that have been identified from past studies. Two different parameterisation settings have been considered (Serre and Riesenhuber 2004; Lei et al. 2007; Serrano et al. 2011). Our preliminary experiments have shown that the two particular Gabor filter parameterisation ranges, have no noticeable effect on PAD results. Thus, we chose the parameterisation values given (Serrano et al. 2011).

Additionally, it is known that V1 cell receptive field sizes vary considerably (McAdams and Reid 2005; Rust et al. 2005; Serre et al. 2007) to provide a range of thin to coarse spatial frequencies. Similarly, four different receptive field sizes were used here with pixel dimensions *3x3, 5x5, 7x7* and *9x9*. Coarser features are handled by area V2, explained in the next section.

2.4 Area V2 - Texture Features

The significance of textural information is often neglected or downplayed in past presentation attack detection and biologically-inspired vision studies. However, the role of cortical area V2 in shape and texture feature extraction is crucial and V2 cells share many of the edge properties found in V1. Nevertheless, V2 cell selectivity has broader receptive fields and is attuned to more complex features compared with V1 cells (Hegdé and Van Essen 2000; Schmid et al. 2014). In addition to broader spatial features, this layer processes textural information and is therefore capable of expressing the different nature of surfaces. This is a crucial advantage in face presentation attack detection where there is a wealth of information hidden within the texture of faces, facial features or face attacks. For example, texture of beards, skin, and glasses can prove a valuable feature against spoofing attacks mimicking their nature. Therefore, accurate representation of texture facial features or the lack of such features where these would be otherwise expected is an important indication of falsification.

V2 cells are effectively expressed by a sinusoidal grating cell operator though other shape characteristics also correspond well (Hegdé and Van Essen 2000). The grating cell operator has not only shown great biological plausibility with respect to actual V2 texture processes but has also proven superior to Gabor filters in texture tasks (Grigorescu et al. 2002). Its response is relatively weak to single bars but in contrast, it responds heavily to periodic patterns. The approach used here (Petkov and Kruizinga 1997) consists of two stages. In the first stage grating subunits generate the responses of on-centre and off-centre cells. In the following stage, grating cell responses from at a particular orientation and periodicity are added together.

A certain response G*r* of a grating subunit at position *(x, y)*, with orientation θ and periodicity λ is given by (Petkov and Kruizinga 1997):

$$\mathrm{Gr}(\mathrm{x},\mathrm{y})_{\theta,\lambda}=\begin{cases}1, \text{ if } \forall\ \mathrm{n},\ \mathrm{M}(x,y)_{\theta,\lambda,\mathrm{n}}\geq \rho\mathrm{M}(x,y)_{\theta,\lambda}\\ 0, \text{ if } \exists\ \mathrm{n},\ \mathrm{M}(x,y)_{\theta,\lambda,\mathrm{n}}< \rho\mathrm{M}(x,y)_{\theta,\lambda}\end{cases} \tag{7}$$

where $n \in \{-3 \ldots 2\}$, ρ is the threshold parameter between 0 and 1 (typically 0.9). For the current position *(x´, y´)* along *x* and *y* directions from starting point *(x ,y),* the maximum activities of *M* are calculated as followed (Petkov and Kruizinga 1997).:

$$\mathrm{M}(x,y)_{\theta,\lambda,\mathrm{n}} = max\begin{cases} s(x',y')_{\theta,\lambda,\varphi_n} \mid \\ n\frac{\lambda}{2}cos\theta \leq x'-x < (n+1)\frac{\lambda}{2}\cos\theta \\ n\frac{\lambda}{2}sin\theta \leq y'-y < (n+1)\frac{\lambda}{2}\sin\theta \end{cases} \tag{8}$$

$$\varphi_{\mathrm{n}}=\begin{cases}0, \mathrm{n}= -3,-1, 1\\ \pi, \mathrm{n} = -2, 0, 2\end{cases} \tag{9}$$

and

$$\mathrm{M}(x,y)_{\theta,\lambda,} = \max\left(\mathrm{M}(x,y)_{\theta,\lambda,\mathrm{n}}\right) \tag{10}$$

The responses at $M(x,\ y)_{\theta,\lambda,n}$ in equation 8, are simple cell responses with symmetric receptive fields along a line segment *3λ*. Essentially this means that there are three peak responses for each grating subunit at point *(x, y)* at a given orientation *θ*. This line segment is split in *λ/2* intervals. The particular position of each interval defines the response of on-centre and off-centre cells. In other words, a grating cell subunit is maximally activated when on-centre and off-centre cells of the same orientation and spatial frequency are activated at point *(x, y)*. In equation 9 φ_n is the phase offset and for values between 0 and π, it corresponds to symmetric centre-on and centre-off operations respectively.

In the second part of V2 grating cell design, a response *w* of grating cell centred on *(x, y)* along orientation θ and periodicity λ, is the weighted summation of grating subunits with orientations θ and $\theta+\pi$, as given below:

$$w(x,y)_{\lambda,\theta} = \int exp\left(-\frac{(x-x')^2+(y-y')^2}{2(\beta\sigma)^2}\right)\Big(\mathrm{Gr}(\mathrm{x}',\mathrm{y}')_{\theta,\lambda} + \mathrm{Gr}(\mathrm{x},\mathrm{y})_{\theta+\pi,\lambda}\Big)dx'dy', \theta \in [0,\pi) \tag{11}$$

Parameter β is the summation area size with a typical value of 5. For our experiments, the number of simple cells were empirically chosen at 3 and all other parameter values were set at default values according to (Petkov and Kruizinga 1997).

2.5 From Input Data to Presentation Attack Detection

Input Layer: The purpose of the input layer is to prepare image information by scaling down all input RGB images to a minimum of 300 pixels for the shortest edge. This particular image size was chosen as a good compromise between speed/time and computational cost.

Layer L1: This layer plays the role of the lateral geniculate nucleus and separates visual stimuli in the appropriate double-opponency channels (Figure) as given from section 2.2 while scaling all pixel values to the same range between 0 and 1.

Layer L2a: Gabor filter operations perform edge detection according to parameterisation values given in section 2.3. It is important to note that after obtaining filtered outputs from all Gabor filters (in total 192) for each double-opponency channel, a maximum operator is applied so that a particular maximum response of L2a vectors ($x_1 \ldots\ x_m$) in a neighbourhood *j* is given by:

$$\mathrm{r}=\mathrm{argmax}_{\mathrm{j}}(\mathrm{x}_{\mathrm{j}}) \tag{12}$$

The maximum operator is a well-known non-linear biological property exhibited by certain visual cells at low levels of visual cognition, pooling visualinputs from previous layers (Lampl et al. 2004; Riesenhuber and Poggio 1999) to greater receptive fields. This hierarchical process gradually projects meaningful visuospatial information to higher cortical layers in the mammalian brain (Figure).

Layer L2b: In this layer grating cell operations are performed according to the settings given in section 2.4. Subsequently, grating outputs are spatially summed with outputs from L2a, in order to form a single L2 output for each of the three double-opponency channels. Spatial summation is another cognition property of the visual cortex and like the maximum operator it is intended to combine several inputs into outputs for higher layers. Spatial summation is used in this layer in order to preserve the spatial integrity and sensitive texture information in faces (Figure 3).

Layer L3: The three double-opponency channels after spatial summation, contain both edge and texture features and along with the RG-BY spectral channels from L1 that contain the spectral differences of each image, are fed through histograms with a window size of 20 units and bin size of 10. These values were empirically selected from experimentation as ideal for the particular layer dimensions. These spatial histograms have been used before in the context of face recognition but with lower level features at L1(Zhang et al. 2005) .

Here they are employed at an intermediate level of feature processing and with various types biological-like features. It is further important to note here that since all these spatiospectral channels carry different types of visual information, they are never mixed together.

Layer L4: In this layer all image data from the previous layer are simply concatenated and sorted in a multidimensional vector for either the training or testing phase, without any further processing. Vector dimensions vary according to the size of the dataset and choice of parameters within the model. For example, if in the previous layer L3 different settings for the spatial histograms were to be used on different datasets (i.e. different number of images), the vector size would be different for each case.

Layer L5: Supervised classification takes place in this layer and any classifiers used can be trained with the extracted feature vector from L4. Training data are selected by following the 10-fold cross-validation technique. The supervised classifiers chosen for this work were a SVM with a linear kernel and KNN with Euclidean distance.

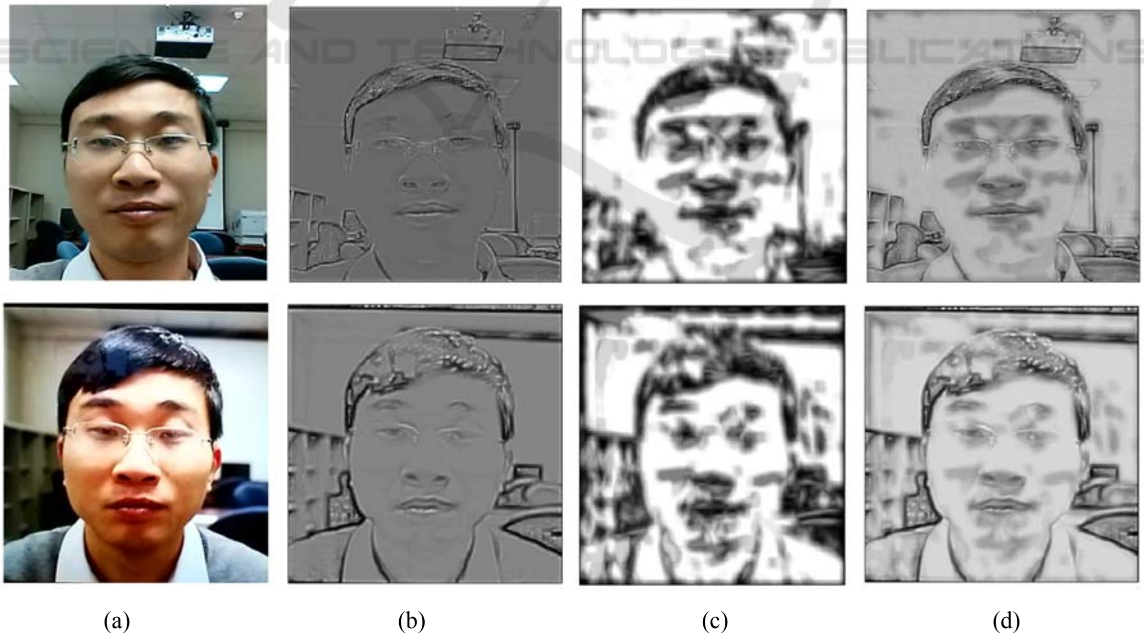

Figure 3: A genuine access attempt versus a photo-print attack. Top row shows the progressive process of a genuine photo attempt. Bottom row shows the printed photo attack. Column (a) shows the input layer images. Column (b) the L2a layer as processed from edge detection Gabor filters, column (c) the L2b layer processed from texture grating cells and column (d) the combined layers L2a and L2b after spatial summation. The richness and depth of edge-texture information in the original image (top row) is apparent.   .

3 EXPERIMENTS

All presentation attack detection experiments were conducted with MATLAB using standard computer hardware. The databases employed for this work as well as the different spoofing attacks that were explored, are explained further in section 3.1 below. It is important to note that in all experiments for both genuine and impostor attacks, only one photo per person was used from the entire video sequences. Since our model currently does not perform any liveness detection method, successive video frames are not being considered. For the purpose of homogeneity and statistical accuracy between datasets, train and test data were divided with the cross-validation technique, bypassing the original train/test data split found in the CASIA database.

3.1 Databases

The CASIA Face Anti-Spoofing (Zhang et al. 2012) database is a database from the Chinese Academy of Sciences (CASIA) Centre for Biometrics and Security Research (CASIA-CBSR). This database contains videos at 10 seconds of real-access and spoofing attacks of 50 different subjects, divided into train and test sets with no overlap. All samples were captured with three devices at different resolutions: a) low resolution with an old 640x480 webcam, b) normal resolution with a more up-to-date 640x480 webcam and c) high resolution with a 1920x 1080 Sony NEX-5 camera. Three different attacks were considered, a) warped, spoofing attacks are performed with curved copper paper hardcopies of high-resolution digital photographs from genuine users, b) cut, attacks are performed using hardcopies of high-resolution digital photographs from genuine users, with the eye areas cut out to simulate eye blinking, c) video, genuine user videos are replayed in front of the capturing device using a tablet.

The MSU Mobile Face Spoofing Database or MFSD (Wen et al. 2015) for face spoof attacks, consists of 280 video clips of photo and video attack attempts of 35 different users. This database was produced at the Michigan State University Pattern Recognition and Image Processing (PRIP) Lab, in East Lansing, US. The MSU database has the following properties, a) mobile phones were used to acquire both genuine faces and spoofing attacks, b) printed photos were generated as high-definition prints and their authors claim that these have much better quality than printed photos in other databases of this kind. Two types of cameras were used in this database, a) built-in camera in MacBook Air at a resolution of 640x480, and b) front-facing camera in the Google Nexus 5 Android phone at a resolution of 720x480. Spoofing attacks were generated using a Canon SLR camera, recording at 18.0Mpixel photographs and 1080p high-definition video clips and iPhone 5S back-facing camera, recording 1080p video clips.

3.2 Results

The PAD biologically-inspired model explained in section 2.5 was evaluated against two databases, CASIA and MFSD. The main concern of our experiments was the detection success rate of spoofing attacks made by potential impostors. Keeping this in mind, consideration was primarily given to whether a fake access could be successfully detected or if the subject’s image was a genuine access attempt. This was treated as a two-class classification problem. The applied biometric evaluation procedures are defined for the spoofing False Acceptance Rate (sFAR) and False Rejection Rate (FRR) as:

$$sFAR = \frac{\text{Impostor attacks seen as genuine}}{\text{Total number of attacks}} \tag{13}$$

$$FRR = \frac{\text{Rejected genuine access attempts}}{\text{Total number of genuine access attempts}} \tag{14}$$

Moreover, overall performance of the presentation attack detection is further presented according to (SC37 ISO/IEC JTC1 and Biometrics 2014) with an additional measure, Average Classification Error Rate (ACER). The average of impostor attacks incorrectly classified as genuine attempts and normal presentation incorrectly classified as impostor attacks is given by:

$$ACER = \frac{\text{sFAR+FRR}}{2} \tag{15}$$

All scores were obtained using 10 folds in the cross-validation technique and in order to further testify performance scores, the L4 feature vectors were applied under two different classification schema. A Support Vector Machine (SVM) with a linear kernel and k-nearest neighbour classifier of *n=2* nearest neighbours with Euclidean distance as a distance measure. In reality, the number of neighbours varies according to the dataset but for the two class problem here out of all *n* values, 2 produced the best average on both datasets as found through cross-validation.

Figure 4 and Figure 5, show the overall classification or detection accuracy rates for the two classifiers SVM and KNN. These accuracy rates are defined as the number of images for each database correctly classified as genuine or fake, i.e true positives and true negatives. It is quite evident in both figures that SVM portrays a desirable and consistent detection rate which persists at rates over 85%. Depending on the choice of training and testing data, it is also apparent that significant deviations in results can occur. This is largely due to the small sample sizes from both databases, leading to significant statistical variance. The average classification accuracy scores from all trials in table 1, also highlight the large differences between the two classifiers.

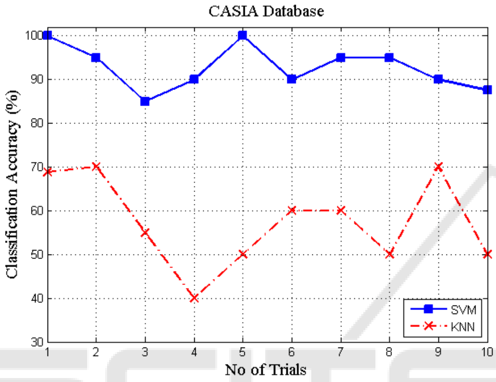

Figure 4: Overall classification/detection accuracy rate of 10 trials with two different classifiers, SVM (blue) and KNN (red) for the CASIA database. –

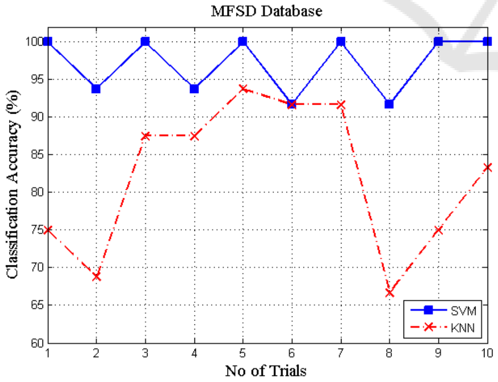

Figure 5: Overall classification/detection accuracy rate of 10 trials with two different classifiers, SVM (blue) and KNN (red) for the MFSD database.

From SVM scores in table 1 it can be deduced that our PAD model performs well under both databases. Better performance is achieved for the MFSD database which is not entirely surprising since it consists of higher resolution images and high resolution print attacks. In addition, there is greater variability of features from the subjects in the MFSD database

Table 1: The average classification percentages (%) and standard deviation values of 10 trials with cross-validation.

Dataset	***μ SVM***	***μ KNN***	***σ^2 SVM***	***σ^2 KNN***
CASIA	92.75	57.37	5.06	10.18
MFSD	97.08	82.08	3.82	9.97

The detection accuracy rates in table 1 provide an insight into the overall ability of the PAD model to detect spoofing attacks. From these results it is seen that the model can achieve a high detection rate at 97% with relatively acceptable standard deviation value of 3.82 for the SVM case in the MFSD database. The KNN classifier with the CASIA database has shown the worst performance overall. This result indicates how important feature selection and classifier choice can be for presentation attack detection. If there is a significant overlap in feature attributes then KNN portrays sensitivity which requires fine-tuning. KNN sensitivity is also illustrated in the large differences between trials in Figure 5. The relatively low number of features in higher dimensional space in our case is a problem better suited for a SVM. While conclusions from table 1 can be useful, biometric evaluation becomes more meaningful when measured in terms of sFAR and FRR and table 2 more effectively capture the nature of error.

Table 2: Average sFAR and FRR percentage scores with SVM classifier after 10 trials.

Dataset	***sFAR***	***FRR***	***ACER***
CASIA	2.77	14.58	8.67
MFSD	3.44	5	4.22

Table 2 shows that error percentages are relatively small and comparable with other state-of-the-art algorithms that have used these databases in the past. The sFAR percentages for both databases are comparable but there is a significant difference between the two databases in their FRR percentages. Naturally, this is also reflected onto the ACER percentages.

A significant difference of 9.58% between FRR percentages indicates the difficulty of distinguishing attacks from genuine access attempts in the CASIA database. In effect, this proves the importance of image quality in terms of both verification and presentation attack cases since the CASIA database consists of photos with lower quality than MFSD.

We further wanted to investigate the impact V1 and V2 operations of edge and texture have on the overall performance of presentation attack detection. These tests were only performed for the SVM linear kernel case. In table 3 below, we provide V1 PAD values alongside values of V1 and V2 of the complete model (table 1) for comparison. As seen from this table, not only overall PAD scores drop for V1 only operations but standard deviation values across all trials indicate a worsening performance. While these values are indicative in these early stages of experimentation, another study on optimum parameterisation for each layer may yet reveal a more important relationship between edge and texture features for accurate presentation attack detection.

Table 3: The average classification percentages (%) and standard deviation values of 10 trials with cross-validation for V1 and V2 operations.

Dataset	*μ* *V1 only*	*μ* *V1&V2*	*σ^2* *V1*	*σ^2* *V1&V2*
CASIA	90	92.75	8.6	5.06
MFSD	95.63	97.08	6.25	3.82

Furthermore, the performance between the two datasets can be viewed from the Detection Error Tradeoff (DET) curve as shown in Figure 6. The DET curve for the CASIA dataset contains more samples and therefore is slightly larger, yet superior performance of the model is noticeable for the MFSD database.

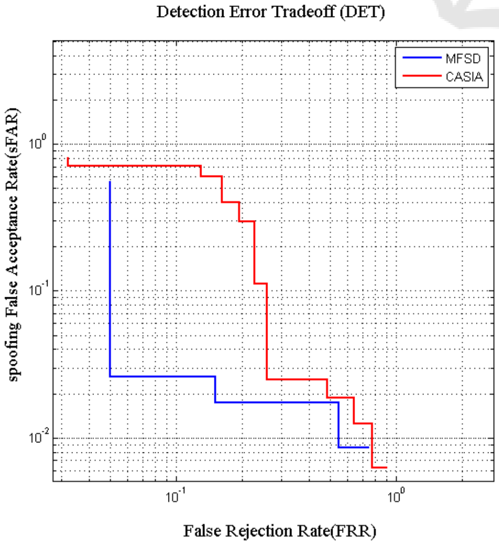

Figure 6: DET curve of SVM scores for the CASIA (red) and MFSD (blue) databases.

4 CONCLUSIONS

In this work we have presented a novel presentation attack detection hierarchical algorithm that relies on the extraction of edge and texture biologically-inspired features, by mimicking processes of the visual cortex and particularly of areas V1 and V2.

By using two different datasets for the most common and basic attacks, we have achieved results that average at 94% detection rates. ACER scores for both databases indicate a tolerable error performance that will in the future be compared with other existing state-of-the-art algorithms. It was further obvious from our experiments that the nature of data is well separated in classification by a SVM classifier.

Overall, the results have been promising and the proposed model can serve as the foundation for further enhancements. Future work will include refinement of the biological-like operations to increase performance and speed, optimisation for video and real time processes, experimentation with more datasets, different types of attacks such as 2D and 3D masks, and experimentation in dynamic – real world scenarios.

ACKNOWLEDGEMENTS

This work has been part of the ABC4EU project and has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement No 312797.

REFERENCES

Anjos, A., Chakka, M.M. & Marcel, S., 2014. Motion-based counter-measures to photo attacks in face recognition. *IET Biometrics*, 3(3), pp.147–158.

Axelrod, V. & Yovel, G., 2012. Hierarchical Processing of Face Viewpoint in Human Visual Cortex. *Journal of Neuroscience*, 32, pp.2442–2452.

Chakraborty, S. & Das, D., 2014. An overview of Face Liveness Detection. *International journal of Information Theory*, 3(2), pp.11–25.

Chingovska, I., Anjos, A. & Marcel, E., 2012. On the effectiveness of local binary patterns in face anti-spoofing. *International Conference of the Biometrics Special Interest Group*.

D, F. & D, V.E., 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1), pp.1–47.

Daugman, J.G., 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of Optical Society of America*, 2(7), pp.1160–1169.

Engel, S., Zhang, X. & Wandell, B., 1997. Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature*, 388(6637), pp.68–71.

Fukushima, K., Miyake, S. & Ito, T., Neocognitron: a neural network model for a mechanism of visual pattern recognition. In *IEEE Transactions on Systems, Man, and Cybernetics*. p. 826—834.

Galbally, J., Marcel, S. & Fierrez, J., 2014. Image quality assessment for fake biometric detection: Application to Iris, fingerprint, and face recognition. *IEEE Transactions on Image Processing*, 23(2), pp.710–724.

Grigorescu, S.E., Petkov, N. & Kruizinga, P., 2002. Comparison of texture features based on Gabor filters. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 11(10), pp.1160–1167.

Hegdé, J. & Van Essen, D.C., 2000. Selectivity for complex shapes in primate visual area V2. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 20(5), p.RC61.

Hermosilla, G. et al., 2012. A comparative study of thermal face recognition methods in unconstrained environments. *Pattern Recognition*, 45(7), pp.2445–2459.

Hubel, D.H. & Wiesel, T.N., 1967. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195(1), p.215–243.

Van Kleef, J.P., Cloherty, S.L. & Ibbotson, M.R., 2010. Complex cell receptive fields: evidence for a hierarchical mechanism. *Journal of Physiology*, 588(18), pp.3457–3470.

Kose, N., Apvrille, L. & Dugelay, J.-L., 2015. Facial makeup detection technique based on texture and shape analysis. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. Ljubljana: IEEE, pp. 1–7.

Lampl, L. et al., 2004. Intracellular Measurements of Spatial Integration and the MAX operation in complex cells of the cat primary visual cortex. *Journal of Neurophysiology*, 92, pp.2704–2713.

LeCun, Y. et al., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86, pp.2278–2324.

Lei, Z. et al., 2007. Face recognition with local gabor textons. *Advances in Biometrics*, pp.49–57.

Li, J. et al., 2004. Live face detection based on the analysis of fourier spectra. In *Defense and Security*. pp. 296–303.

Li, M. et al., 2013. Face recognition using early biologically inspired features. In *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*. pp. 1–6.

Lyons, M. et al., 1998. Coding facial expressions with Gabor wavelets. *Proceedings - 3rd IEEE International Conference on Automatic Face and Gesture Recognition, FG 1998*, pp.200–205.

Maatta, J., Hadid, A. & Pietikäinen, M., 2011. Face spoofing detection from single images using micro-texture analysis. In *2011 International Joint Conference on Biometrics (IJCB)*. pp. 1–7.

Marcelja, S., 1980. Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America*, 70, pp.1297–1300.

McAdams, C.J. & Reid, R.C., 2005. Attention modulates the responses of simple cells in monkey primary visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 25, pp.11023–11033.

Meyers, E. & Wolf, L., 2008. Using biologically inspired features for face processing. *International Journal of Computer Vision*, 76(1), pp.93–104.

Pan, G., Wu, Z. & Sun, L., 2008. Liveness detection for face recognition. *Recent Advances in Face Recognition*, (December), p.236.

Perlibakas, V., 2006. Face Recognition using Principal Component Analysis and Log-Gabor Filters. *Analysis*, 3(February 2008), p.23.

Petkov, N. & Kruizinga, P., 1997. Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: bar and grating cells. *Biological cybernetics*, 76, pp.83–96.

Pisharady, P.K. & Martin, S., 2012. Pose invariant face recognition using neuro-biologically inspired features. *International Journal of Future Computer Communications*, 1(3), pp.316–320.

Prokoski, F.J. & Riedel, R.B., 2002. Infrared identification of faces and body parts. *Biometrics*, pp.191–212.

Raghavendra, R., Raja, K.B. & Busch, C., 2015. Presentation Attack Detection for Face Recognition Using Light Field Camera. *Image Processing, IEEE Transactions on*, 24(3), pp.1060–1075.

Ramon, M., Caharel, S. & Rossion, B., 2011. The speed of recognition of personally familiar faces. *Perception*, 40(4), pp.437–49.

Riesenhuber, M. & Poggio, T., 1999. Hierarchical models of object recognition in cortex. *Nat. Neurosci.*, (2(11):1019-25).

Riesenhuber, M. & Poggio, T., 2000. Models of object recognition. *Nature Neuroscience*, 3, pp.1199–1204.

Rolls, E.T., 2012. Invariant Visual Object and Face Recognition: Neural and Computational Bases, and a Model, VisNet. *Front Comp Neurosci*, 6, p.35.

Rose, N., 2006. Facial Expression Classification using Gabor and Log-Gabor Filters. In *7th International Conference on Automatic Face and Gesture Recognition, 2006. FGR 2006*. pp. 346–350.

Rust, N.C. et al., 2005. Spatiotemporal elements of macaque V1 receptive fields. *Neuron*, 46, pp.945–956.

Van De Sande, K., Gevers, T. & Snoek, C., 2010. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9), pp.1582–1596.

SC37 ISO/IEC JTC1 & Biometrics, 2014. *Information Technology—Presentation Attack Detection—Part 3: Testing, Reporting and Classification of Attacks*,

Schmid, A.M., Purpura, K.P. & Victor, J.D., 2014.

Responses to orientation discontinuities in V1 and V2: physiological dissociations and functional implications. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 34(10), pp.3559–78.

Seal, A. et al., 2013. Automated thermal face recognition based on minutiae extraction. *International Journal of Computational Intelligence Studies*, 2(2), pp.133–156.

Serrano, Á. et al., 2011. Analysis of variance of Gabor filter banks parameters for optimal face recognition. *Pattern Recognition Letters*, 32, pp.1998–2008.

Serre, T. et al., 2007. Robust Object Recognition with Cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3), pp.411–426.

Serre, T. & Riesenhuber, M., 2004. Realistic Modeling of Simple and Complex Cell Tuning in the HMAX Model , and Implications for Invariant Object Recognition in Cortex. *Methods*, p.017.

Slavkovic, M. et al., 2013. Face recognition using Gabor filters, PCA and neural networks. In *2013 20th International Conference on Systems, Signals and Image Processing (IWSSIP)*. pp. 35–38.

Wang, S. et al., 2013. Aging face identification using biologically inspired features. In *2013 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC 2013)*. pp. 1–5.

Wang, Y. & Chua, C., 2005. Face recognition from 2D and 3D images using 3D Gabor filters. *Image and Vision Computing*, 23(11), pp.1018–1028.

Webster, M.A. & De Valois, R.L., 1985. Relationship between spatial-frequency and orientation tuning of striate-cortex cells. *Journal of the Optical Society of America. A, Optics and image science*, 2, pp.1124–1132.

Wen, D., Han, H. & Jain, A.K., 2015. Face spoof detection with distortion analysis. *IEEE Transaction on Information Forensics and Security*, 10(4), pp.746–761.

Wu, H.-Y. et al., 2012. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics*, 31(4), pp.1–8.

Yan, J. et al., 2012. Face liveness detection by exploring multiple scenic clues. In *12th International Conference on Control Automation Robotics & Vision (ICARCV)*. pp. 188–193.

Yokono, J.J. & Poggio, T., 2004. *Rotation Invariant Object Recognition from One Training Example*.

Zhang, W. et al., 2005. Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A novel non-statistical model for face representation and recognition. *Proceedings of the IEEE International Conference on Computer Vision*, I, pp.786–791.

Zhang, Z. et al., 2012. A face antispoofing database with diverse attacks. *Proceedings - 2012 5th IAPR International Conference on Biometrics, ICB 2012*, pp.26–31.