I. ABSTRACT:

AL FINAL LO ESCRIBIMOS

This paper presents a multispectral detection of spoofing attacks using the learned features of a Convolutional Neural Network. Visible light, near-Infrared and thermal images are used with this purpose. The design of the CNN is the result from a search of the parameters which defines a Neural Network. Different classifiers have been tested to get the one whose performance is the finest.

Index Terms: Biometrics, Presentation Attack Detection. Antispoofing. Automatic Border Crossing systems. Convolutional Neural Network, bio-inspired systems.

II. INTRODUCTION:

Face biometrics systems are commonplace by society, such as unlock mobile devices, as a key to enter to private places, to identify criminal people in public places e. g. airports or subways or just to ascertain the identity of an individual \cite{survey2}, \cite{PaperDavid}.

One recent application of face verification systems in real environments and with huge security constrains is Automated Border Crossing (ABC) systems. In the borders of several countries border guards are helped by ABC systems to perform routine tasks such as passport control and face verification. Queue time delays are reduced by the using of these systems, mainly devoted to speed up border crossing for “bona fide” travelers.

Nowadays the performance of the face verification systems has achieved remarkable results, but other aspects like the robustness against attacks are not usually considered or treated with certain lack of detailed. In the case of ABC systems security aspects are absolutely crucial and all the system vulnerabilities in this aspect should be eliminated or at least minimized. In this sense, face anti-spoofing is a topic whose importance is increasing in the face biometrics area, and the quantity of references devoted to presentation attack detection (PAD) or detecting spoofing attacks has grown in the recent years. Some recent conferences focused specifically in this areas has appeared, like the 16th International Workshop on Digital Forensics and Watermarking, IWDW 2017, held in Magdeburg, \cite{Congress}.

Anti-spoofing capabilities are extremely important in the design of an ABC systems. In these systems, face verification is done in an almost unsupervised scenario. Border crossing is a high security environment and border guards make their efforts to guarantee that only passport holder can cross the border. Identity suplantation in a border crossing can be used by criminal organizations and therefore, automatic systems should be ready to process not only bona fide traveler’s identity but also to detect presentation attacks to the ABC system.

People could try to hack the system with the determination of not recognize a user. The attack can be produced at sensor level, feature level or score level, depending on where is produced in the system \cite{Spoofing\_survey}. Attacks to the system, at the sensor level, are commonly labeled “spoofing attacks” and preventing or detecting those attacks is denominated “anti-spoofing” or “Presentation Attack Detection” (PAD) \cite{ISO\_PAD}.

Sensor level attacks (in face biometrics) are commonly produced by tools to impersonate people as masks or printed photos positioned in front of the sensor (camera) so that the sensor would capture the photo instead of the correct person face. Could be highlighted that the equipment necessary to spoof systems is low-price or easy to be acquired (camera) \cite{MFSDMSUdatabase} , what is more, facial images are actually easy to obtain through social networks \cite{attacks}.

Automated border crossing is a highly harmonized and defined systems. In Europe, Frontex agency has publish technical guidelines and recommendations for ABC face verification procedures, Standardization groups both in Europe and in international levels also defined the face verification procedure based on a matching one live image with passport store image. In a similar way, PAD attacks will be detected based on alone image processing. PAD based on images facilitates its integration in real ABC systems since PAD can be performed in a similar way in which face processing and verification is done. Considering this requirement, the PAD system presented in this work has been designed only for one-image approach, avoiding video approximations not suitable for this environment.

Regarding the data considered, the face anti-spoofing researched work is commonly addressed using visible light images, therefore, the public available face anti-spoofing databases for researching purposes contains only RGB images. We can mention CASIA \cite{CASIAdatabase}, REPLAY ATTACK \cite{Replaydatabase} and MFSD-MSU databases \cite{MFSDMSUdatabase.

Nowadays there are available in the market other kind of sensors that can works at a different range of the spectra and can aquire threedimensional information too. Our method aims to use different image sensors (visible light camera, thermal and near-infrared sensors) for PAD with the purpose of using them separately or at the same time considering costs and computations restrictions. The attacks considered are performed using the following artifacts: printed photo, printed mask, printed mask with the eyes cropped or an image displayed on a tablet.

Several methods to detect presentation attacks from images have been proposed, mainly they used texture and color characteristics following a classifier that can separate authentic images form attack images. One of the most important trends in processing and classification during last years is the used of deep learning and deep neural networks. The solutions presented in this work follows a novel approach for PAD systems using a deep neural network, more specifically, Convolutional Neural Networks (CNN). The CNN are trained with the purpose of learning images features that encourage the classification task.

The remainder of the paper is organized as follows: an abstract which summarizes the object of this article in \ref{sec:XX}, an introduction of multispectral face-antispoofing in section \ref{sec:XXX}, an analysis of the researched work in this article field (section \ref{sec:XX}), a description of the used method to develop this project in section \ref{sec:xxx} in which the utilized database is described, as well as the algorithm and experiments description. In section \ref{sec:XXX} the results obtained in each experiment are presented and discussed, in section \ref{sec:XXX} the conclusions and future work is detailed followed of the acknowledgments. The last section \ref{sec:XXX} points out the references.

III. PREVIOUS WORK

Different techniques have been presented recently to determine if a user is being impersonated or not, those techniques could be used with image or video sequences. A general state of the art in PAD research is out of the scope of this paper since this is a huge research area with new and promising methods. This state of the art is focused on the Deep Learning for face anti-spoofing detection using multispectral images.

Adding different biometrics systems improves the performance of the system because the weakness of each acquisition system are neutralized by the others \cite{multimodal}. Using others acquisition system could be other types of sensors or other biometric systems such as fingerprint or iris.

In \cite{multispectral\_handbookbiometrics} authors use visible light (VIS) images and Near-Infrared (NIR) images. The authors present a visible subsystem and a NIR subsystem which attack independently and a multi-spectral system (VIS and NIR). Authors attack both subsystems with visible and NIR images independently, but they attack de multi-spectral system with pairs of visible and NIR images. The color of the visible images is analyzed to detect if it is an attack while the texture is analyzed in case of NIR images. The multi-spectral system is composed by two-step based on color analysis, NIR photos are rejected because of the color information because NIR photo has no color and VIS photo attack is detected because of the reflectance of the light. Despite obtaining a 100% accuracy, the multi-spectral system is not a unit, attacks are being fallen apart during the process, our system is a whole and it is at the end when the classifier classifies the sample, in addition, they only use photo attacks.

In \cite{multispectral\_wavelength} authors present a method for liveness detection (not PAD, nevertheless is related to). The aim of detecting liveness is that attacks are produced with artificial products. Authors use two different wavelengths to photograph and record reals users and attacks with. The database is not uniform, the genuine cluster is formed by 120 users, the attacks subsets does not have the same number of samples, depends on the attacks which are planar images, and different minds of mask faces. Attacks are not used simultaneously, each attack is compared with genuine users. The reflectance obtained from the two wavelengths are used to train a SVM. Authors obtain a 100% accuracy when planar attacks videos are being used and 92% accuracy on planar attacks photos vs genuine users. If genuine users are compared with mask face attacks, 89% accuracy is obtained. Authors affirm that their method depends on the material which covers the faces. This method does not compare all the attacks with genuine users at the same time, our method does.

In \cite{multispectral\_gradient} authors purpose is detecting liveness (not anti-spoofing detection) their paper is based on the gradient of the captured images by a multispectral system with different wavelengths. Four gradient-based vectors, obtained from the reflectance, are used as input to a SVM to classify the samples into genuine or attack. Experiments are not realized with the same database, one is formed by 70 positive and 132 negative samples, the other is formed by 59 subjects (three orientations) for positive database and the negative database includes 53 photos, 26 PVC mannequins, and 33 silica gel masks. For the classification task, SVM classifier is used to learn the gradient-based features of genuine and fake faces. Best results authors obtain area 98.3 TPR (%) and 98.7 TNR (%).

Our work differs from the described ones in that we use three different sensors (visible, thermal, and infrared) to capture the database and its corresponding attacks (printed photo, printed mask, printed mask with the eyes hole cropped and tablet attacks). We compare the results of using each sensor separately or all together, whereas, in other works, they use as much two different sensors and not using the same method.

Convolutional Neural Networks are, nowadays, widely used in recognition and classification tasks obtaining satisfying results \cite{CNNexample1,CNNexample2,CNNexample3}. Detecting spoofing attacks, with CNN as main tool, is a field that has started to be researched. However, using CNN for multispectral PAD detection is not investigated yet and it is this paper purpose. The latest developed investigations on anti-spoofing with visible light cameras and CNN are the described following ones:

In \cite{LearnCNN}, authors implement a canonical CNN structure for face anti-spoofing. Authors use videos recorded by a visible light camera. The CNN is adopted from the architecture which won ImageNet 2012 called AlexNet \cite{Imagenet]. Authors use CASIA (formed by 55 users) and REPLAY-ATTACK (50 subjects) databases. The face is detected in each sample and cropped with different scale-ratio. Authors use from one to three video frames to augment the data temporally. Authors assert that in spite of not carefully selecting the parameters of the CNN, the obtained results are successful (HTER lower than 5%). The inter-test results do not have good performance.

In \cite{LSTM-CNN}, authors implement a recurrent neural network with Long Short Term Memory units above a CNN architecture (which is based AlexNet). The determination of this paper is detecting face anti-spoofing. CASIA database (50 subjects: 20 for training, and 30 for testing) has been utilized and a 5.9% HTER (EER is 5.17%) is obtained. Authors base its research in learning the temporal features from videos and then, classify them using logistic regression. LSTM units are a very powerful tool but our method uses images, not video.

A new architecture is developed in \cite{TransferLearningCNN} which is denominated FASnet and is based on VGG-16 architecture \cite{VGG-16}. The neural network is pre-trained with ImageNet database. Authors use the transfer learning approach, more specifically, the fine-tuning and use the CNN to address face anti-spoofing methods. Authors obtain 100 % ACC and 0% HTER, when 3DMAD database is used, and 99.04% ACC and 1.20% HTER when REPLAY-ATTACK.

Methods referenced in this state of the art are based only on visible light cameras or RGB images, in any case with limited number of sensors. In the best of our knowledge, there are not available anti-spoofing databases whose samples are obtained with more than two sensors. In the present work, four different sensors have been considered: NIR (near infrared), thermal, visual and range or depth domains has been considered.

IV. DATABASE DESCRIPTION

The new database name “FRAV-Attack” database was acquired in order to evaluate the designed system. The purpose of this database is to cover as much as possible the different facial representations provided by several sensors.

The capture process was realized in several days, and users were volunteers which signed the correspondent consent form. In addition, the database agrees with LOPD regulation of the Spanish Agency for Data Protection and GDPR (General Data Protection Regulation) regulation of the European Union. The database is formed by 185 users. 62% of users are men and 48% are women, with an age range where the 75% of users are between 18 and 40 years and a 25% are more than 40 years.

Each user has been captured with the following sensors: a visible light camera SONY ILCE-6000Y, a Near Infrared camera, a thermal camera FLIR and a depth camera (Real Sense camera). With each camera, samples of genuine users and attacks to those users have been captured.

The attack samples considered are a wide range of artifacts: a simple printed photo attacks, printed photo mask, printed photo mask with eyes area cropped (to emulete liveness activity), and a tablet shown image.

To sum up, we use three subsets (visible light, NIR and thermal) which proceed to five subsets: visible light, NIR, thermal those three subsets added at classification level and added at the feature level.



An example of visible lights genuine users and its corresponding attacks is represented in figure \ref{fig:XXX} is which could be seen genuine user (a), photo attack (b), paper mask(c), paper mask with the eye holes cropped (d) and tablet attack (e).

The same user is represented in figure \ref{fig:} in the Near-Infrared subset, where can be visualized genuine user (a), photo attack (b), paper mask(c), paper mask with the eye holes cropped (d) and tablet attack (e).



The thermal images (grayscale mode) are represented in figure \ref{fig:XXX}, where at the same than in the others subsets, is represented genuine user (a), photo attack (b), paper mask (c), paper mask with the eye holes cropped (d) and tablet attack (e).



The face in the samples has not been selected and cropped, because a) databases are only constituted by facial images and b) it has been demonstrated in previous works \cite{LearnCNN} and \cite{LSTM-CNN} that background contributes to the spoofing identification.

The three particular subsets have been processed in the same procedure. Due to the fact that images (of the same database) do not share the same shape, images are read independently and resize to 128x128 pixels with the objective of using them together and separately.

The samples have been captured under uniform and controlled illumination: two T4 fluorescent tubes at 6000K and 12W has been used.

Part of the database has been published in the Idiap Research Institute, in the dataset distribution portal \cite{idiap}.

V. METHOD AND EXPERIMENT DESCRIPTION

In this article multispectral images for PAD has been investigated: different types of images are used visible light images, Near-Infrared and thermal images.

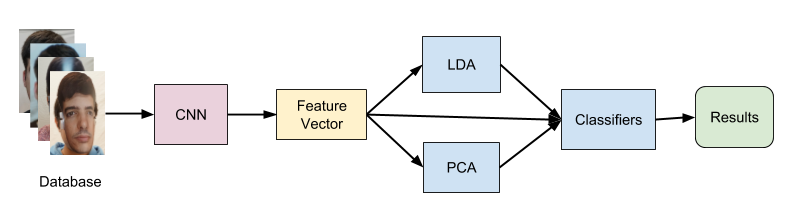
Anti-spoofing is being investigated in each kind of images separately and together, hence added in feature level and in classification level.

A. Method

The general architecture of the experiment is using as input of a convolutional neural network, whose hyperparameters have been carefully selected, the three subsets (visible, NIR and thermal) together or separately).

The convolutional neural network learns its own filters through which the best characteristics of the images (feature vector) are obtained with the intention of using them as input of the classifiers.

The classification task is formed by three processes, five different classifiers are trained and tested with the outcome CNN data, PCA and LDA algorithms are trained too with the features vector of the CNN (independently), the output of PCA and LDA is used as input to the classifiers, as long as PCA and LDA techniques are dimensionality reduction algorithms. From these experiments, three groups of results are obtained for each database: CNN + classifiers, CNN + PCA + classifiers and CNN + LDA + classifiers as is represented in figure \ref{fig:XXX}.



The classifiers used in the classification task are Support Vector Machine (SVM) with RBF and linear kernel, K-Nearest Neighbour (KNN), Decision Tree and Logistic Regression. Each one has been trained independently. A sample is considered as an attack or a user in this last step.

Furthermore, LDA and PCA techniques have been used between the CNN and the classifier in order to reduce the dimensionality of the data because the CNN returns samples with a characteristic vector whose dimensionality is 500.

1. CNN Architecture: The architecture of the convolutional neural network is based on AlexNet \cite{Imagenet}, some parameters and the architecture has been modified trying to improve the performance. The layers, that constitute the own architecture, are the following ones:

1. Convolutional Layer (11x11) + MaxPool layer (2x2) + Normalization layer

2. Convolutional Layer (4x4)

3. Convolutional Layer (3x3)

4. Convolutional Layer (3x3)

5. Convolutional Layer (3x3) + MaxPool layer (2x2)

6. Dropout Layer

7. Dropout Layer

8. Fully connected Layer

The number of kernels used for each convolutional layer are: 56 for the first layer,156 for the second, 256, 254 and 106 for the last convolutional layer . At the dropout layer 2512 neurons have been used and 500 at the fully connected.

The utilized activation function is REctified Linear Unit (RELU). Logistic regression function has been used to train the network; the negative likelihood has been used to compute the cost function at training.

With respect to the weights initialization, Gaussian (with std = 0.01 and mean value = 0) and uniform distribution for initialization (also called `Xavier initialization’) have been tested, the greatest results obtained are with Gaussian weight initialization.

Different learning rates have been tried: 0.01 static learning rate; dynamic learning rate with an initial value of 0.0095 and decay proportional to 0.995 per epoch. The best learning rate configuration and the used in the described experiments is dynamic learning rate with an initial value of 0.009 and a decay proportional to 0.995 at each epoch. 30 samples have been used per batch. The training procedure has been run by 200 epochs.

The code has been developed using Theano framework \cite{theano} with the purpose of optimizing the model on GPU.

2. Classification: Samples are tagged as 0 or 1 depending on if they are bona fide or attacks samples, thus its is a binary classification task, no distinction is made among attacks. As a result, the number of negative samples is higher than positives.

After the CNN training procedure, classifiers are trained and validated getting the hyperparameters that suits better the classification task.

B. Experiment Description.

1. Experiment 1:

In this experiment, each subset (visible, thermal and NIR) is tested individually. Each one is trained independently in the CNN, and the classifiers are trained too. The test samples are tested after the classification task: first the features are gained with the trained CNN and are used to feed classifiers, LDA and PCA techniques.

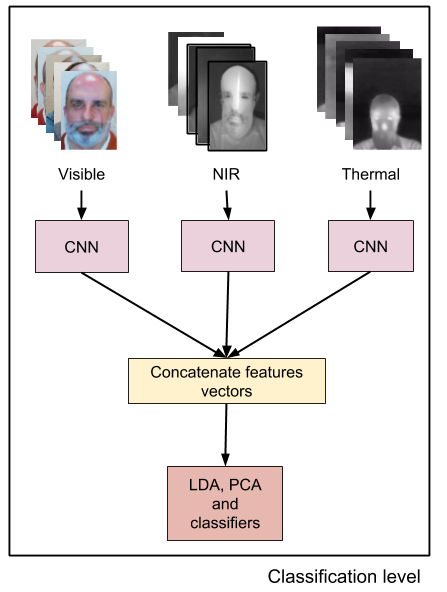
From this experiment, three different and independent clusters of results are obtained, one per subset.

2. Experiment 2

In the previous experiments description, databases are used independently, however, using the three databases at the same time have been tested too. Using the three databases could be realized in two approaches: concatenating the three databases before the CNN procedure (feature level) and concatenating the features vectors after each independent CNN procedure (classification level). In this experiment, the classification level is described.

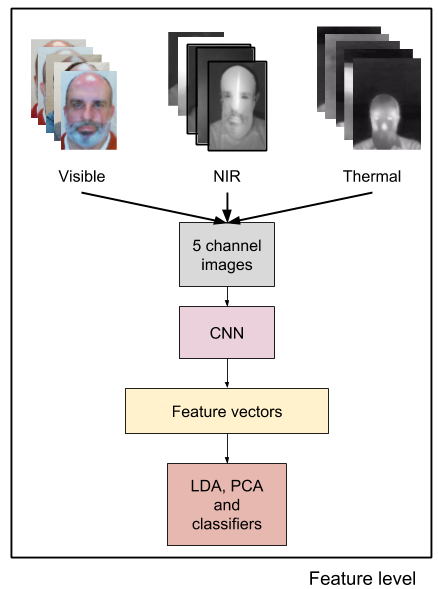
Each subset is used as input in three CNN (same architecture, though different CNN), the neural network is trained. Before using as input the three outputs of each neural network, features are concatenated resulting in a characteristic vector whose dimension is the sum of the three subsets dimensions as is represented in figure \ref{XXX}. Classifiers are fed with the concatenated feature vector.

As the same way as Experiment 1, the classifiers (and PCA and LDA) are trained and then, tested.



3. Experiment 3:

Experiment 3 describes the feature level. Three subsets are concatenated, due to visible light images have three channels (RGB) and thermal and NIR are in grey-scale images (1 channel), the result of concatenate images is a 5 channels image as can be visualize in figure \ref{XXX}, and is used as input to the neural network.



Summarizing, five clusters of results would be acquired: visible, NIR and thermal (independently each one), three databases added at feature level and added at classification level. From each cluster of results, there would be results of using only classifiers, PCA + classifiers and LDA + classifiers.

VI. RESULTS AND DISCUSSION

The results obtained with each database are exposed in this section and are represented with the metric APCER, BPCER and ERR value:

- Attack Presentation Classification Error Rate (APCER): is defined as the proportion of presentation attacks that have been classified incorrectly (as bona fide) \cite{ISO\_APCR}.

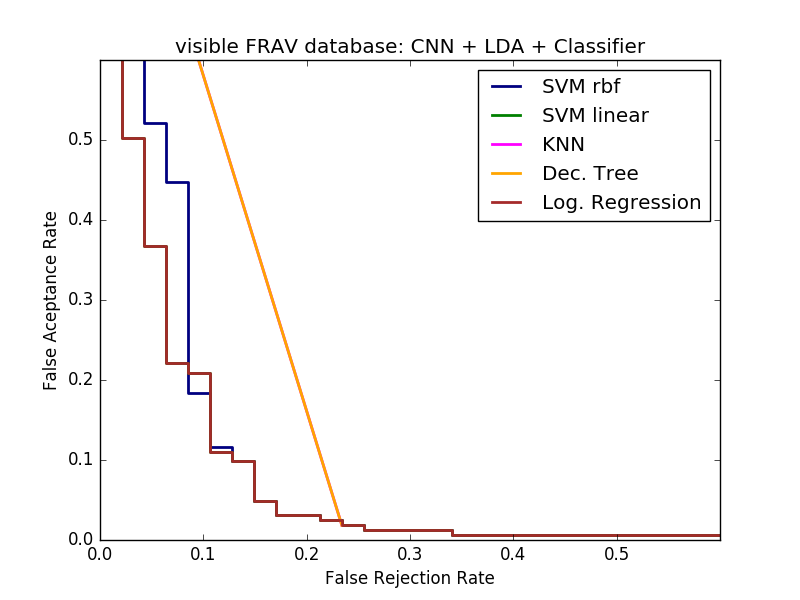
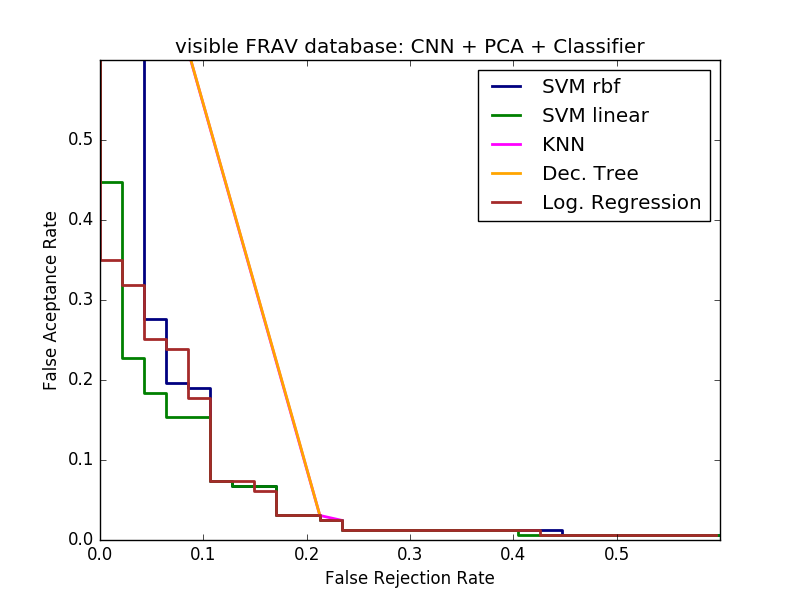
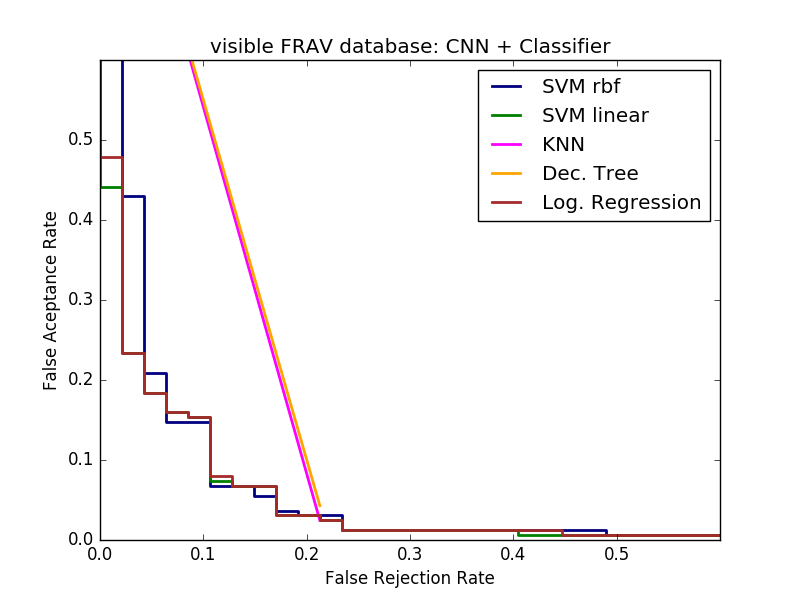
- Bona Fide Presentation Classification Error Rate (APCER): is defined as the proportion of bonafide presentation incorrectly classified as presentation attacks \cite{ISO\_APCR}.

- Equal Error Rate (EER %) is defined as the point where the FAR curve and the FRR curve get the same value \cite{EER}.

- ROC curves (False Rejection Rate (FAR) / False Acceptance Rate (FAR)) \cite{Intro\_biometrics2}.

The subsets have been analyzed, so that, five cluster results are obtained: visible light, NIR and thermal independently, added in classification level and added in feature level.

1. Results of experiment 1:
2. Test on Visible light subset: The model is tested on visible light subset, using the different classifiers and LDA and PCA algorithms. In table \ref{table:XXX} is summarized the APCR and BPCR parameters obtained with the best performance of each classifier. For details, in figure \ref{fig:XXX) the DET curves are presented. As we can see using LDA improve slightly the results. In the three cases, the are more negative misclassified samples than positives. Best result is obtained when CNN+SVM (linear) has been used.



In table \ref{table:XXX} are described the parameters that characterize the classifiers with which results are enhanced.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Visible subset | CNN + Classifier | | | CNN + PCA + Classifier  PCA = 3 components | | | CNN + LDA + Classifier  LDA = 1 component | | |
| APCER | BPCER | EER(%) | APCER | BPCER | EER(%) | APCER | BPCER | EER(%) |
| SVM RBF | 0,0245 | 0,234 | 10,5 | 0,0245 | 0,234 | 10,5 | 0,0245 | 0,234 | 11,1 |
| SVM lineal | 0,0245 | 0,2127 | 10,5 | 0,0245 | 0,2128 | 10,5 | 0,0184 | 0,234 | 10,8 |
| KNN | 0,0245 | 0,2127 | 11,8 | 0,0245 | 0,234 | 12 | 0,0184 | 0,234 | 12,6 |
| Dec. Tree | 0,0429 | 0,2127 | 12,78 | 0,0307 | 0,2128 | 16 | 0,0184 | 0,234 | 12,6 |
| Log. Regression | 0,0306 | 0,0741 | 10,8 | 0,0307 | 0,1702 | 10,5 | 0,1043 | 0,1277 | 10,8 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | SVM RBF  (C - Gamma) | SVM lineal  (C) | KNN  (Neighbours) | Decision Tree  (Depth) | Log. Regression  (learning rate) |
| CNN + Classifier | 0.05, - 0.0046 | 0.001 | 1 | 2 | same as training CNN |
| CNN + LDA + Classifier | 0,05 - 0,0046 | 0.001 | 3 | 2 | 0,005 |
| CNN + PCA + Classifier | 0.01 - 0.0046 | 0.001 | 1 | 2 | 0,005 |

The best EER result obtained with visible subset is 10,5% while in \cite{LearnCOnvolutionalNN}, authors obtain 4,64% EER when the database that they use is formed by 50 users (three frames per video); in our case our database is formed by 185 users and is composed by photos, no videos.

1. Test on NIR subset

With the NIR subset, the APCER and BPCER results are reported with regard to each classifier. In table \ref{table:XXX} we present the results. Coming out of the table results, we can observe that the performance has been greatly improved compared with visible light results, entire positive samples are correctly classified, the number of negative samples incorrectly classified is very low.

With regard to the classifiers performance, SVM (linear and RBF) and logistic regression are those whose results are exceed to others. Using LDA or PCA algorithms do not remarkably improve the test results. Best EER value is 0% which has been collected practically is each performance. The most unfavorable EER value obtained is 0,9% which is indeed an acceptable performance. Because that the probabilities obtained in each classifier are close to 0 or 1, the FAR-FRR plot draws a perfect curve.

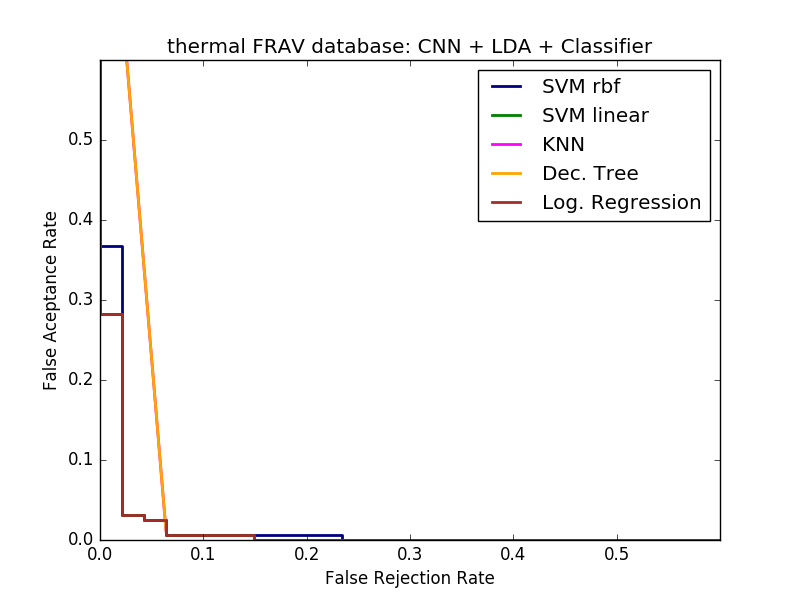
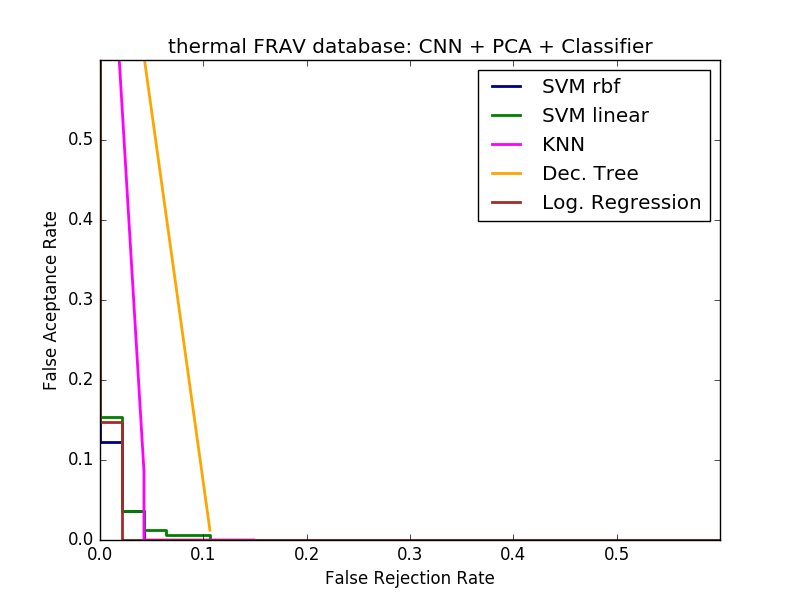
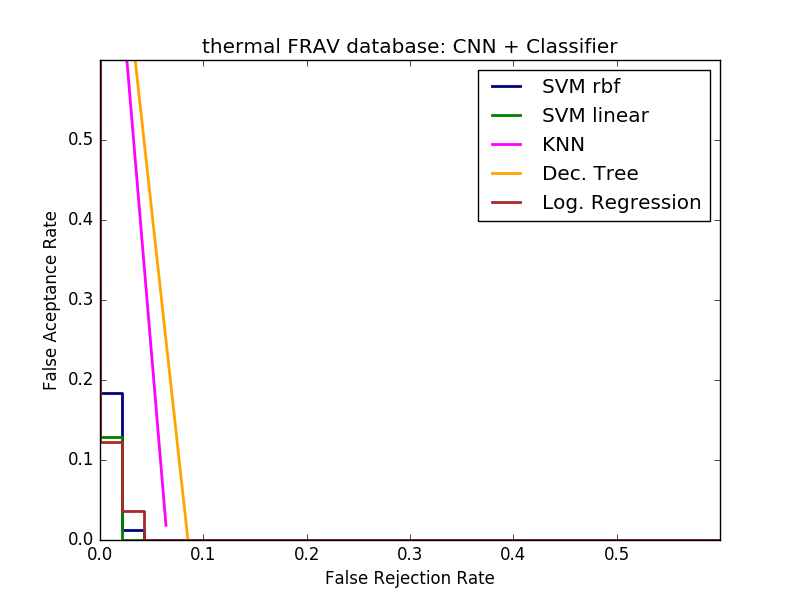
In table \ref{table:XX) are exposed the parameters of the classifiers with which the results are obtained.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NIR subset | CNN + Classifier | | | CNN + PCA + Classifier  PCA = 13 | | | CNN + LDA + Classifier  LDA = 1 component | | |
| APCER | BPCER | EER(%) | APCER | BPCER | EER(%) | APCER | BPCER | EER(%) |
| SVM RBF | 0,0123 | 0 | 0 | 0,0123 | 0 | 0 | 0,0123 | 0 | 0 |
| SVM lineal | 0,0123 | 0 | 0 | 0,0123 | 0 | 0 | 0,0123 | 0 | 0 |
| KNN | 0,0123 | 0 | 0,03 | 0,0123 | 0 | 0,03 | 0,0123 | 0 | 0,06 |
| Dec Tree | 0,0184 | 0 | 0,09 | 0,0184 | 0 | 0,06 | 0,0123 | 0 | 0,06 |
| Log. Regression | 0,0061 | 0 | 0 | 0,0184 | 0 | 0 | 0,0307 | 0 | 0 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | SVM RBF  (C - Gamma) | SVM lineal  (C) | KNN  (Neighbours) | Decision Tree  (Depth) | Log. Regression  (learning rate) |
| CNN + Classifier | 0.05, - 0.0046 | 0.001 | 9 | 6 | same as training CNN |
| CNN + LDA + Classifier | 0,05 - 0,0046 | 0.001 | 3 | 2 | 0,005 |
| CNN + PCA + Classifier | 0.01 - 0.0046 | 0.001 | 1 | 2 | 0,005 |

1. Test on thermal subset

With regard to thermal subset, table \ref{table:XXX} represents the results obtained when this subset has been employ. In figure \ref{fig:XXX} the DET curves are represented, in which is possible to visualize that thermal results outperforms visible light results, though they are poorest that NIR subset.



EER is almost zero in each test, occurring the lowest value at CNN + SVM (linear and RFB) 1,98% and the higher value at CNN + PCA + Decision Tree 5,932%.

PCA slightly improve the performance of CNN + classifier and LDA results improve PCA performance. The best execution is obtained with CNN + SVM (linear and RBF), achieving a 1,984%.

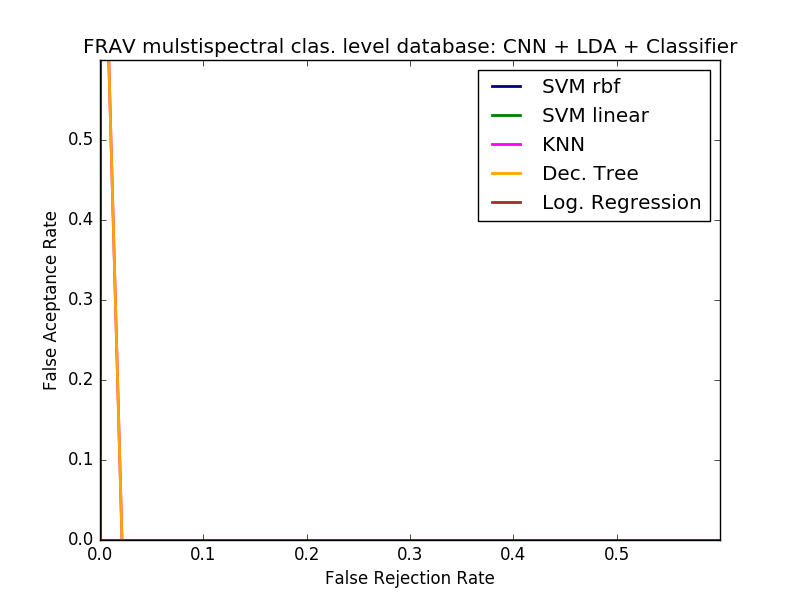
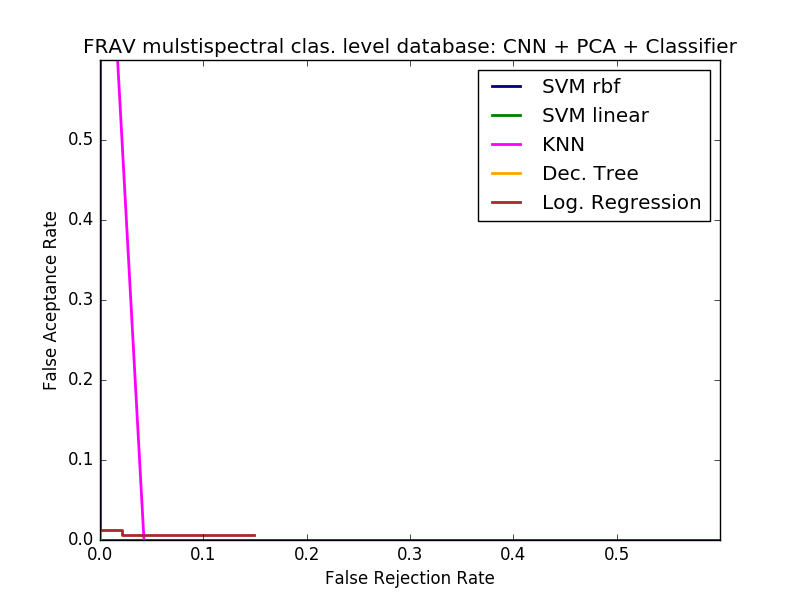
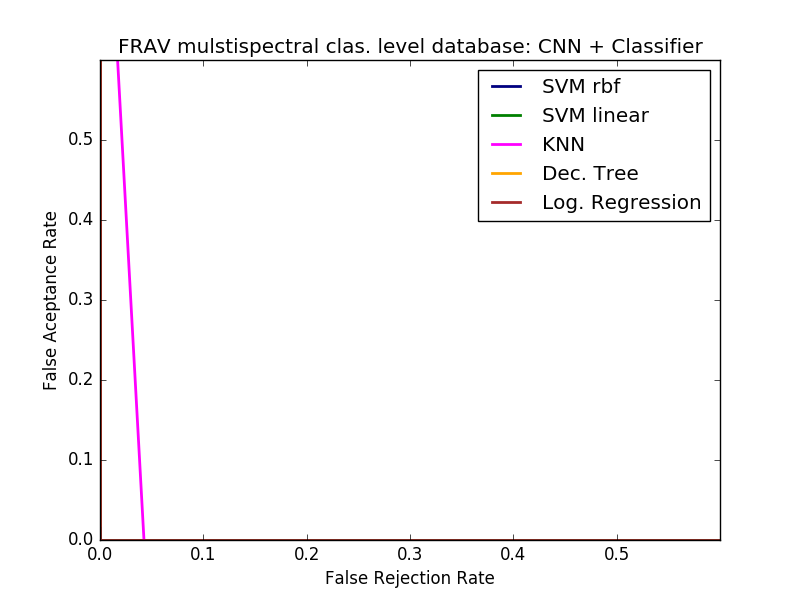
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Thermal subset | CNN + Classifier | | | CNN + PCA + Classifier  PCA = 13 | | | CNN + LDA + Classifier  LDA = 1 component | | |
| APCER | BPCER | EER(%) | APCER | BPCER | EER(%) | APCER | BPCER | EER(%) |
| SVM RBF | 0 | 0,0638 | 1,984 | 0 | 0,1277 | 3,968 | 0,0123 | 0,0638 | 2,597 |
| SVM lineal | 0 | 0,0638 | 1,984 | 0,0061 | 0,0638 | 3,968 | 0,0123 | 0,0638 | 2,597 |
| KNN | 0,0184 | 0,0638 | 4,112 | 0 | 0,1064 | 3,968 | 0,0123 | 0,0638 | 3,805 |
| Dec Tree | 0 | 0,0851 | 4,255 | 0,0123 | 0,1064 | 5,932 | 0,0123 | 0,0638 | 3,805 |
| Log. Regression | 0 | 0,0426 | 3,968 | 0 | 0,0213 | 1,984 | 0,1534 | 0,995 | 2,598 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | SVM RBF  (C - Gamma) | SVM lineal  (C) | KNN  (Neighbours) | Decision Tree  (Depth) | Log. Regression  (learning rate) |
| CNN + Classifier | 0.05, - 0.0046 | 0.05 | 1 | 2 | same as training CNN |
| CNN + LDA + Classifier | 0,05 - 0,00021 | 0.5 | 17 | 20 | 0,005 |
| CNN + PCA + Classifier | 0.001 - 0.0046 | 0.001 | 1 | 2 | 0,005 |

B. Results of experiment 2:

The model is tested when the three subsets are used at the same time added at classification level, that means, the output features of each independent subset are concatenated and trained.

The results are exposed in table \ref{table:XXX} from which could be realized that are satisfying, EER gets zero value in each execution and APCER and BPCER acquire low values or zero. In figure \ref{fig:XXX} the DET curve is detailed. From the table results and DET curve, it is possible to visualize that despite the fact that a perfect performance is obtained in PCA + Decision Tree performance, SVM and logistic regression works better.



The parameters that define each classifier are exposed in table \ref{table:XXX}. From the results, must be pointed out the perfect classification obtained when PCA and Decision tree is used, in others cases, classification task is almost perfect.

Most of the cases, entirely attack samples are classified correctly, genuine users are the inaccurate classified samples. Moreover, EER values get zero essentially in all cases, the poorest value is 2,12% obtained with CNN+KNN.

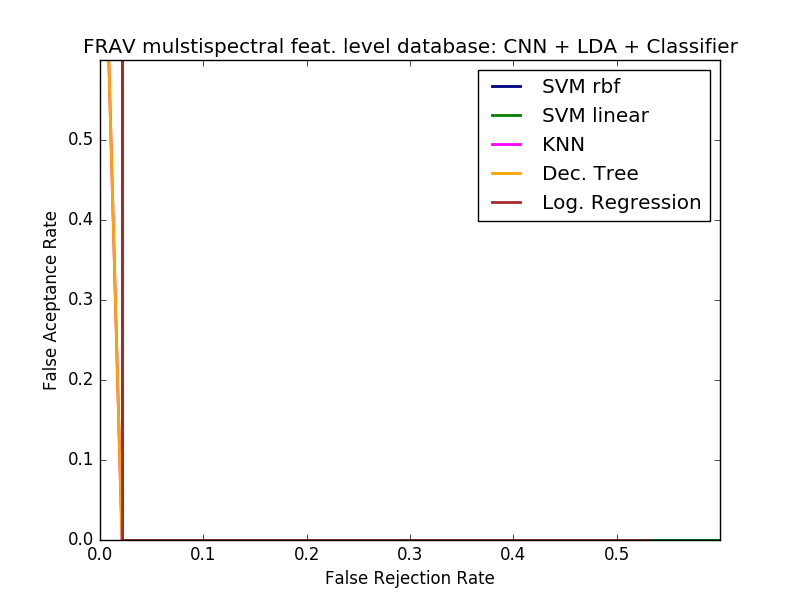
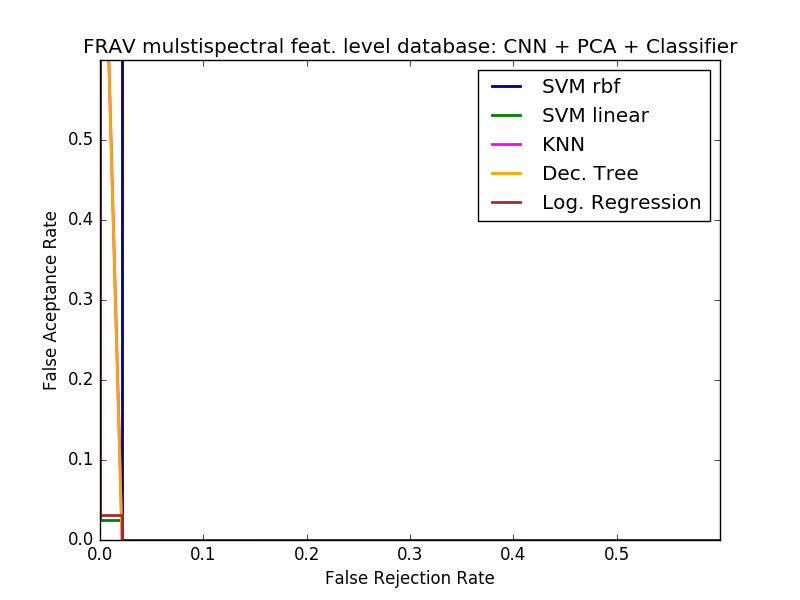
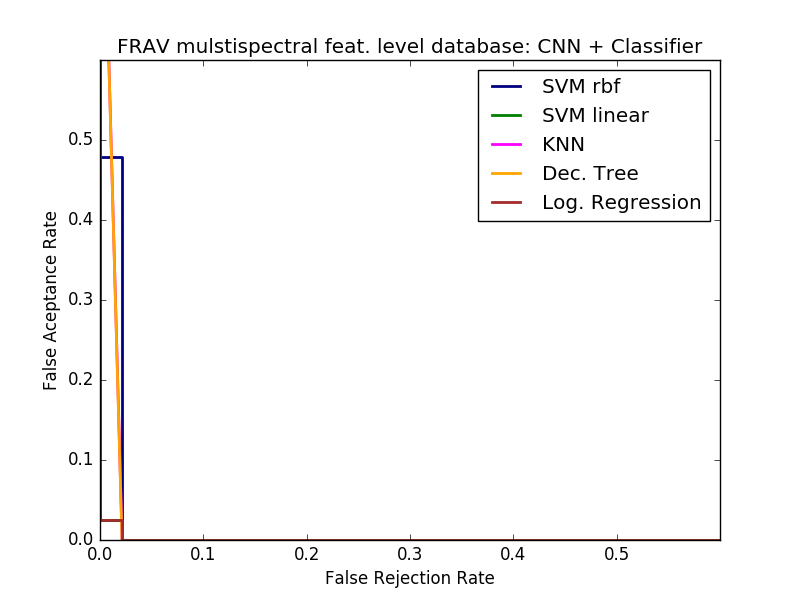
If three subsets are utilized at the same time the classification task improve their results when it is compared with visible light subset performance. NIR subsets results outperform the obtained with three subsets.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| All subset clas. level | CNN + Classifier | | | CNN + PCA + Classifier  PCA = 353 components | | | CNN + LDA + Classifier  LDA = 1 component | | |
| APCER | BPCER | EER(%) | APCER | BPCER | EER(%) | APCER | BPCER | EER(%) |
| SVM RBF | 0 | 0,0638 | 0 | 0 | 0,0638 | 0 | 0 | 0,0213 | 0 |
| SVM lineal | 0 | 0,0213 | 0 | 0 | 0,0213 | 0 | 0 | 0,0213 | 0 |
| KNN | 0 | 0,0426 | 2,12 | 0 | 0,0426 | 2,12 | 0 | 0,0213 | 1 |
| Dec Tree | 0,0184 | 0 | 0,92 | **0** | **0** | **0** | 0 | 0,0213 | 1 |
| Log. Regression | 0 | 0,0213 | 0 | 0,0123 | 0 | 1,68 | 0,0123 | 0 | 0 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | SVM RBF  (C - Gamma) | SVM lineal  (C) | KNN  (Neighbours) | Decision Tree  (Depth) | Log. Regression  (learning rate) |
| CNN + Classifier | 0.05, - 0.00021 | 0.001 | 1 | 2 | 0,005 |
| CNN + LDA + Classifier | 0,05 - 0,00021 | 0.001 | 1 | 2 | 0,05 |
| CNN + PCA + Classifier | 0.05 - 0.0046 | 0.001 | 1 | 2 | 0,005 |

C. Results of experiment 3:

The last cluster of results corresponds to the three subsets added at feature level, meaning, images are added before training a CNN. The outcome of the model is presented in table \ref{table:XXX} for each used classifier (the parameters that describe each classifier are specify in table \ref{table:XXX}) and the DET curves are represented in figure \ref{fig:XXX}, where it could be seen that results are as finest as in classification level test.



Best results are achieved with LDA algorithm, nevertheless, the difference with PCA or CNN+classifier is minor.

APCER values at feature level are near zero, likewise the obtained at classification level. Bona fide samples are erroneous classified. Lowest EER value is 1% and the higher is 1.67%, thus the difference is not remarkably, the number of incorrectly labeled samples is not decidedly contrastive.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| All subset feat. level | CNN + Classifier | | | CNN + PCA + Classifier  PCA = 3 components | | | CNN + LDA + Classifier  LDA = 1 component | | |
| APCER | BPCER | EER(%) | APCER | BPCER | EER(%) | APCER | BPCER | EER(%) |
| SVM RBF | 0 | 0,0638 | 1,37 | 0 | 0,0213 | 1,677 | 0 | 0,0213 | 1,67 |
| SVM lineal | 0 | 0,0213 | 1,677 | 0 | 0,0213 | 1,677 | 0 | 0,0213 | 1,67 |
| KNN | 0 | 0,0213 | 1,37 | 0 | 0,0213 | 1,37 | 0 | 0,0213 | 1 |
| Dec Tree | 0 | 0 | 1,063 | 0 | 0,0213 | 1 | 0 | 0,0213 | 1 |
| Log. Regression | 0,0184 | 0 | 1,677 | 0 | 0,0213 | 1,98 | 0,0061 | 0,0213 | 1,9 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | SVM RBF  (C - Gamma) | SVM lineal  (C) | KNN  (Neighbours) | Decision Tree  (Depth) | Log. Regression  (learning rate) |
| CNN + Classifier | 0.05, - 0.0046 | 0.001 | 3 | 2 | same as training CNN |
| CNN + LDA + Classifier | 0,05 - 0,00021 | 0.001 | 5 | 2 | 0,05 |
| CNN + PCA + Classifier | 0.01 - 0.0021 | 0.001 | 1 | 2 | 0,005 |

1. Final discussion:

Thus far, the proposed model has been evaluated with different subsets which differs in the capture system. From the results, it can be noticed that the ones obtained with the visible subset are the worst, and they are improved when NIR or thermal subsets are tested. Moreover, when the three subsets are analyzed at the same time (added at classification level or feature level) the behavior is as desirable as the obtained when NIR or thermal subsets are tested.

Despite proving that thermal and NIR results are more exceptional than visible light results, using the three subsets all at once, the NIR performance is superior.

Best result is obtained when CCN + PCA + Decision Tree are used at classification level are better than expected, as a matter of a fact, all samples are correctly classified.

Regardless of the fact that the best performance is obtained when PCA and Decision tree are used, SVM linear and Logistic regression perform a better classification assignment.

Nevertheless of looking for a model that suits better for this issue, a better model can be achieved if a further analysis is realized.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a new approach to use a CNN in order to learn features for multispectral face anti-spoofing, being the first time that CNN and multispectral face anti-spoofing have been used concurrently. We have adjusted an existing CNN architecture to adapt it to our problem. We have trained the neural network with different spectral images (visible, NIR and thermal) and adding those three subsets at characteristic level and feature level. We have trained different classifiers with the purpose of selecting the one which best conforms our problem.

Towards this point, our future work is: increasing the number of users of the own database, so thus, adding samples to the database would help to get a better training performance and reliable results at the testing procedure.

VIII. ACKNOWLEDGEMENTS

IX. REFERENCES:

@article{LearnCNN,

author = {Jianwei Yang and

Zhen Lei and

Stan Z. Li},

title = {Learn Convolutional Neural Network for Face Anti-Spoofing},

journal = {CoRR},

volume = {abs/1408.5601},

year = {2014},

}

@INPROCEEDINGS{survey2,

author={Kresimir Delac and Mislav Grgic},

booktitle={Proceedings. Elmar-2004. 46th International Symposium on Electronics in Marine},

title={A survey of biometric recognition methods},

year={2004},

pages={184-193}

}

@ARTICLE{MFSDMSUdatabase,

author={Di Wen and Hu Han and Jain, A.K.},

journal={IEEE Trans. Information Forensic and Security},

title={{ Face Spoof Detection with Image Distortion Analysis}},

year={2015},

month={April},

volume={10},

number={4},

pages={746-761},

}

@INPROCEEDINGS{CASIAdatabase ,

author={Z. Zhang and J. Yan and S. Liu and Z. Lei and D. Yi and S. Z. Li},

booktitle={2012 5th IAPR International Conference on Biometrics (ICB)},

title={A face antispoofing database with diverse attacks},

year={2012},

pages={26-31},

doi={10.1109/ICB.2012.6199754},

ISSN={2376-4201},

month={March},}

@unknown{TransferLearningCNN,

author = {Lucena, Oeslle and Junior, Amadeu and Hugo G Moia, Vitor and Souza, Roberto and Valle, Eduardo and Lotufo, Roberto},

year = {2017},

pages = {27-34},

title = {Transfer Learning Using Convolutional neural networks for Face Anti-Spoofing},

}

@article{VGG-16

author = {Karen Simonyan and

Andrew Zisserman},

title = {Very Deep Convolutional Networks for Large-Scale Image Recognition},

journal = {CoRR},

volume = {abs/1409.1556},

year = {2014} }

@ARTICLE{theano,

author = {Theano Development Team},

title = {Theano: A {Python} framework for fast computation of mathematical expressions},

journal = {arXiv e-prints},

volume = {abs/1605.02688},

year = 2016,

url = {http://arxiv.org/abs/1605.02688}

}

@techreport{ISO\_PAD,

type = {Standard},

key = { **ISO/IEC 30107-1:2016**

month = 1,

year = {2016},

title = {{ Information technology -- Biometric presentation attack detection -- Part 1: Framework }},

institution = {International Organization for Standardization},

}

@techreport{ISO\_APCR,

type = {Standard},

key = { **ISO/IEC 30107-3:2016**

month = 5,

year = {2016},

title = {{ Information technology -- Biometric presentation attack detection – Part 3: Testing and reporting}},

institution = {International Organization for Standardization},

}

Equal Error Rate (EER). (n.d.). *Biometrics Glossary*. Retrieved September 21, 2017, from website: <http://www.expertglossary.com/definition/equal-error-rate-eer>.

@inproceedings{Spectra\_example,

title={Live face detection based on the analysis of fourier spectra},

author={Li, Jiangwei and Wang, Yunhong and Tan, Tieniu and Jain, AK},

booktitle={Proc. of SPIE Vol},

volume={5404},

pages={297},

year={2004}

}

@inproceedings{eyeblink,

title={Eyeblink-based anti-spoofing in face recognition from a generic webcamera},

author={Pan, Gang and Sun, Lin and Wu, Zhaohui and Lao, Shihong},

booktitle={Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on},

pages={1--8},

year={2007},

organization={IEEE}

}

@unknown{antispoofing-Spectral,

author = {Steiner, Holger and Kolb, Andreas and Jung, Norbert},

year = {2016},

month = {06},

pages = {1-8},

title = {Reliable face anti-spoofing using multispectral SWIR imaging}

}

@incollection{multispectral\_handbookbiometrics,

author = {Yi, Dong and Lei, Zhen and Zhang, Zhiwei and Li, Stan Z.},

booktitle = {Handbook of Biometric Anti-Spoofing},

isbn = {978-1-4471-6523-1},

pages = {83-102},

publisher = {Springer},

series = {Advances in Computer Vision and Pattern Recognition},

title = {Face Anti-spoofing: Multi-spectral Approach.},

year = 2014

}

@inproceedings{multispectral\_wavelength,  
 title={Face liveness detection by learning multispectral reflectance distributions},  
 author={Zhang, Zhiwei and Yi, Dong and Lei, Zhen and Li, Stan Z},  
 booktitle={Automatic Face \& Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on},  
 pages={436--441},  
 year={2011},  
 organization={IEEE}  
}

@article{multispectral\_gradient,  
 title={Multispectral face liveness detection method based on gradient features},  
 author={Hou, Ya-Li and Hao, Xiaoli and Wang, Yueyang and Guo, Changqing},  
 journal={Optical Engineering},  
 volume={52},  
 number={11},  
 pages={113102--113102},  
 year={2013},  
 publisher={International Society for Optics and Photonics}  
}

@inproceedings{PaperDavid,  
 title={Face Recognition-based Presentation Attack Detection in a Two-step Segregated Automated Border Control e-Gate - Results of a Pilot Experience at Adolfo Su{\'a}rez Madrid-Barajas Airport},  
 author={David Ortega del Campo and Cristina Conde and {\'A}ngel Serrano and Isaac Mart{\'i}n de Diego and Enrique Cabello},  
 booktitle={SECRYPT},  
 year={2017}  
}

@inproceedings{LSTM-CNN,

author = {Zhenqi Xu and

Shan Li and

Weihong Deng},

title = {Learning temporal features using {LSTM-CNN} architecture for face anti-spoofing},

booktitle = {3rd {IAPR} Asian Conference on Pattern Recognition, {ACPR} 2015, Kuala Lumpur, Malaysia, November 3-6, 2015},

pages = {141--145},

year = {2015},

doi = {10.1109/ACPR.2015.7486482}

}

@article{Spoofing\_survey,

author = {Javier Galbally and Sébastien Marcel and Julian Fiérrez},

journal = {IEEE Access},

pages = {1530-1552},

title = {Biometric Antispoofing Methods: A Survey in Face Recognition.},

volume = {2},

year = {2014}

}

@ARTICLE{Intro\_biometrics2,

author={Anil K. Jain and Arun Ross and Salil Prabhakar},

journal={IEEE Transactions on Circuits and Systems for Video Technology},

title={An introduction to biometric recognition},

year={2004},

volume={14},

number={1},

pages={4-20}

}

@inproceedings{laptop,  
 author = {Li, Yan and Xu, Ke and Yan, Qiang and Li, Yingjiu and Deng, Robert H.},  
 title = {Understanding OSN-based Facial Disclosure Against Face Authentication Systems},  
 booktitle = {Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security},  
 year = {2014},  
 pages = {413--424},  
 numpages = {12},  
 doi = {10.1145/2590296.2590315},  
 publisher = {ACM}  
}

@INPROCEEDINGS{Replaydatabase,  
 author = {Chingovska, Ivana and Anjos, Andr{\'{e}} and Marcel, S{\'{e}}bastien},  
 keywords = {biometric, Counter-Measures, Local Binary Patterns, Spoofing Attacks},  
 month = september,  
 title = {On the Effectiveness of Local Binary Patterns in Face Anti-spoofing},  
 journal = {IEEE BIOSIG 2012},  
 year = {2012},  
}

@Inbook{attacks,

author={de Freitas Pereira, Tiago

and Anjos, Andr{\'e}

and De Martino, Jos{\'e} Mario

and Marcel, S{\'e}bastien},

title={LBP{\thinspace}−{\thinspace}TOP Based Countermeasure against Face Spoofing Attacks},

bookTitle={Computer Vision - ACCV 2012 Workshops: ACCV 2012 International Workshops, Daejeon, Korea, November 5-6, 2012, Revised Selected Papers, Part I},

year={2013},

pages={121--132},

doi={10.1007/978-3-642-37410-4\_11}

}

@Inbook{multimodal,

author={Akhtar, Zahid

and Kale, Sandeep},

title={Security Analysis of Multimodal Biometric Systems against Spoof Attacks",

bookTitle="Advances in Computing and Communications: First International Conference, ACC 2011, Kochi, India, July 22-24, 2011. Proceedings, Part II},

year={2011},

publisher={Springer Berlin Heidelberg},

pages={604--611},

doi={10.1007/978-3-642-22714-1\_62},

}

@article{CNNexample1,  
 title={3D convolutional neural networks for human action recognition},  
 author={Ji, Shuiwang and Xu, Wei and Yang, Ming and Yu, Kai},  
 journal={IEEE transactions on pattern analysis and machine intelligence},  
 volume={35},  
 number={1},  
 pages={221--231},  
 year={2013},  
 publisher={IEEE}  
}

@inproceedings{CNNexample2,  
 title={Best practices for convolutional neural networks applied to visual document analysis.},  
 author={Simard, Patrice Y and Steinkraus, David and Platt, John C and others},  
 booktitle={ICDAR},  
 volume={3},  
 pages={958--962},  
 year={2003}  
}

@inproceedings{CNNexample3,  
 title={Flexible, high performance convolutional neural networks for image classification},  
 author={Ciresan, Dan C and Meier, Ueli and Masci, Jonathan and Maria Gambardella, Luca and Schmidhuber, J{\"u}rgen},  
 booktitle={IJCAI Proceedings-International Joint Conference on Artificial Intelligence},  
 volume={22},  
 number={1},  
 pages={1237},  
 year={2011},  
 organization={Barcelona, Spain}  
}

@book{Congress,

author = {Kraetzer, Christian and Shi, Yun-Qing and Dittmann, Jana and Joong Kim, Hyoung},

year = {2017},

month = {01},

pages = {},

title = {Digital Forensics and Watermarking: 16th International Workshop , IWDW 2017, Magdeburg, Germany, August 23-25, 2017, Proceedings},

isbn = {978-3-319-64184-3}

}

@misc{idiap,

title = {Idiap dataset distribution portal},

howpublished = {\url{https://www.idiap.ch/dataset}},

note = {Accessed: 2018-02-27}

}