ISSN: 1692-7257 - Volume 1 – Number 43 - 2024

RCTA
Revista Colombiana de Tecnologías de Avanzada
UNIPAMPLONA

Digital Object Identifier: 10.24054/rcta.v1i43.2888

Face presentation attack detection based on siamese-LSTM and analysis of optic flow and facial landmarks

Detección de ataques de presentación facial basado en siamese-LSTM y el análisis del flujo óptico y puntos de referencia facial

Ing. Arnold Jiménez Vargas [1], PhD. Rubiel Vargas Cañas [2]
PhD. Carlos Alberto Cobos Lozada [1], PhD. Humberto Loaiza Correa [3]

[1] ***Universidad del Cauca,*** *Facultad de Ingeniería Electrónica y Telecomunicaciones, Grupo de I+D en Tecnologías de la Información, Popayán, Cauca, Colombia.*
[2] ***Universidad del Cauca,*** *Facultad de Ciencias Naturales, Exactas y de la Educación, Grupo de Investigación en Sistemas Dinámicos, Instrumentación y Control, Popayán, Cauca, Colombia.*
[3] ***Universidad del Valle,*** *Facultad de Ingeniería, Grupo de Percepción y Sistemas Inteligentes, Santiago de Cali, Valle del Cauca, Colombia.*

Correspondence: *rubiel@unicauca.edu.co*

Received: *November 8, 2023.* ***Accepted:*** *January 10, 2024.* ***Published:*** *April 30, 2024.*

How to Cite: *A. J. Jimenez Vargas, R. Vargas Cañas, C. A. Cobos Lozada, and H. Loaiza Correa, "Face presentation attack detection based on siamese-LSTM and analysis of optic flow and facial landmarks", RCTA, vol. 1, no. 43, pp. 125–133, Apr. 2024. Recovered from https://ojs.unipamplona.edu.co/index.php/rcta/article/view/2888*

Abstract: Facial biometrics authentication has become essential in verifying the identity of individuals in online transactions, as classic mechanisms like username and password authentication have proven to be unreliable since users often choose easily memorable passwords. However, advances in model manufacturing with materials such as latex, improvements in print quality, and enhancements in screen resolutions have demanded that fraud detection systems quickly adapt to new conditions. This paper presents a proposal to address the problem of detecting presentation attacks by extracting optical flow and facial landmarks and analyzing them through a Siamese-LSTM network. To evaluate the proposed model, three datasets were used: Rose-youtu, Replay-attack, and Replay-mobile, and the metrics used were HTER and EER

Keywords: Biometrics, anti-spoofing, Siamese neural network, LSTM network, optical flow, facial landmark points.

Resumen: La autenticación por medio de la biometría facial se ha vuelto fundamental para verificar la identidad de las personas en transacciones en línea, ya que mecanismos clásicos como la autenticación por nombre de usuario y contraseña han demostrado ser poco fiables, ya que los usuarios suelen escoger contraseñas que son fáciles de recordar. Sin embargo, el avance en la fabricación de modelos con materiales como el látex, el aumento en la calidad de las impresiones y la mejora en las resoluciones de las pantallas han exigido que los sistemas de detección de fraude se adapten rápidamente a las nuevas condiciones. El presente trabajo muestra una propuesta para abordar el problema de la detección de ataques de presentación por medio de la extracción del flujo óptico y los puntos de referencia facial y su análisis por medio de una red siamese. Para evaluar el modelo propuesto, se utilizaron

tres data sets: Rose-youtu, Replay-attack y Replay-mobile, y las métricas HTER y EER.

Palabras clave: Biometría, contra la suplantación de identidad, red neuronal siamese, red LSTM, flujo óptico, puntos de referencia faciales.

1. INTRODUCTION

The massive use of mobile devices, such as tablets or smartphones, has allowed services to be offered through mobile applications that require the unequivocal identification of users. In scenarios such as financial services or security-oriented services, authentication by username and password is not the most suitable and it is in these contexts where biometric authentication is used, which is based on the extraction and processing of physical traits of people to verify their identity, traits that can be extracted among others from the voice, fingerprints and face. The advantage of facial biometrics lies in the fact that more characteristics can be extracted from the face than from the voice or fingerprints, but one must deal with the problem of presentation attacks [1]. One of the most challenging problems in this area is determining whether the device is capturing a person and not a photograph or a screen displaying a video or a photo, fraud modalities that are shown to be the most common and are known as presentation attacks [2].

Presentation attack detection methods are basically divided into two: Classic and deep learning-based [3]. Classic methods include motion analysis using techniques such as DMD (Dynamic Mode Decomposition) [4] used in other areas such as fluid dynamics, and texture analysis through manually extracted features, such as LBP (local binary pattern) [5] or Sobel [6]. The problem with texture-based approaches is that they are sensitive to lighting conditions, capture devices, among other factors as shown in while motion-based approaches are sensitive to attacks that include eye or lip movement, such as in print or latex mask attacks [7]. Deep learning-based methods, which have gained notoriety due to the rapid advancement of this technology, allow training models capable of autonomously extracting features, i.e., without supervision and through the review of dataset samples, this allows models to infer features that are not entirely perceptible to the human eye, and through techniques such as transfer learning [8], learn new features from previously trained models.

This work addresses the problem of presentation attack detection through a siamese network composed of two branches of LSTM networks, which are fed with two types of data: the optical flow calculated for pairs of frames of short-duration videos which are separated from each other by 300 milliseconds, and 68 facial landmark points that are extracted from each of the frames. Siamese networks have been used in image comparison and consist of two identical branches of a neural network that share the same weights and are fed with the two images to compare. The output of both branches is concatenated and used to determine if the images of, for example, faces or signatures, correspond to the same person [9]. This ability of neural networks to determine the degree of similarity between the two inputs was applied to the process of presentation attack detection by combining it with LSTM networks, which are widely used in tasks of sequential data processing such as time series [10] and documents or text [11]. Temporal data is constituted by the optical flow and the facial landmark points, where the optical flow provides information about the apparent movement of the objects in the video and the facial landmark points provide information about the movements of the key points of the face.

The remainer of this paper is organized as follows, in section 2, the most recent and relevant previous works related to fraud detection in facial biometrics are presented. In Section 3, basic concepts necessary to understand the proposed model are presented, including the architecture of the siamese network, the LSTM network, the siamese-LSTM network, the optical flow, and the representative facial points. In Section 4, the proposed model is presented, providing the necessary details to understand and replicate it. In Section 5, the experimentation carried out is presented, starting from the description of the selected datasets to train and evaluate the model, the comparison metrics, and the results obtained by the proposed model compared to those reported in the state of the art. Finally, in Section 6, the conclusions and future work expected to be addressed on the topic are presented.

2. RELATED WORKS

In this section, some previous works in the field of presentation attack detection are reviewed, where we find that some of them combine different approaches, such as those oriented to texture analysis, motion, or deep learning.

In [12], we find a model based on optical flow and texture analysis that consists of extracting the optical flow (section 3.4) from video sequences to describe the direction and intensity of movement, and then integrate them with texture information. The method also introduces local attention and channel mechanisms to adaptively assign weights to different regions and channels, respectively. In [13], the authors propose a learnable gradient operator, designed to extract efficient fine-grain information, such as the spatial gradient magnitude, for images as fraudulent or genuine. This approach offers a data-based solution, unlike methods like the Sobel operator that is handcrafted and uses fixed weights to calculate the gradient magnitude of an image.

There are models in which customer information is used in the process, for example, in [14], the authors propose an approach for presentation attack detection using classifiers specific to each client and convolutional neural networks. According to the authors, presentation attack detection systems obtain relevant information from each class, while client-specific methods take into account the characteristics of each individual, since authentication systems are ultimately used to guarantee access to users individually, thus, with this approach, the system is able to adapt to the unique characteristics and patterns of each individual, improving detection accuracy. In [15], the authors propose an approach that uses information about the client's identity, first recognizing the face and then using the identity to assist in presentation attack detection. The client's identity is used to select a real image of the subject with which a pair is formed with the input image, this pair is used as input in a Siamese network that determines if it is a positive pair (both images are genuine), or if the pair is negative (one of the images is fraudulent).

3. BASIC CONCEPTS OF THE MODEL

3.1. Siamese neutral network

The Siamese Neural Network is a neural network architecture designed to compare two or more inputs and determine the level of similarity. This architecture consists of two or more identical branches, where each one takes an input and processes it independently, the results are joined in an output layer in which, based on a distance metric, the similarity between the inputs is calculated.

Figure 1 shows a Siamese network with two inputs and its components, which are described below.

•Input: This network receives two inputs marked as Input1 and Input2, which represent the data to be compared. The data must be pre-processed to convert them into a representation usable by the neural networks that make up the branches.
•Network Branch: Branch composed of layers of neurons. This network consists of two or more identical branches, which share the same architecture and parameters. It can be said that there is a single neural network that is replicated according to the number of branches that the Siamese network has.
•Encodings: These are the features extracted by the neural networks of the branches when processing the inputs, for example Encoding1 and Encoding2.
•Distance function: It is the layer responsible for comparing the encodings. Here functions such as the Euclidean distance are used.
•Classification: It is the layer responsible for performing the classification according to the result delivered by the distance function. It is generally composed of a series of dense layers.
•Output: The output of the neural network varies according to the problem being solved, classification, regression or a measure of similarity between the inputs. In image classification, it can, for example, say if two input images belong to the same class.

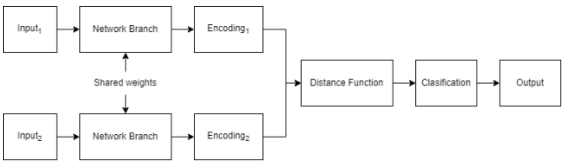

Fig. 1. *Siamese Neural Network Architecture*
Source: *(adapted from [1])*

3.2. Long short-Term memory networks (LSTM)

An LSTM neural network is a type of recurrent neural network (RNN) that is designed to solve the problem of gradient vanishing in traditional RNNs, which occurs when the gradients used to update the weights in the network become very small, making it difficult for the network to learn long-term dependencies. It does this by controlling the flow of

information through gates. Figure 2 shows an LSTM network and its components.

The input of an LSTM network consists of:

•Input (Xt): Input vector at time t. The dimensionality of this vector depends on the domain of the problem and the specific task to be performed.
•Previous hidden state (Ht-1): Output from the previous step t-1. The hidden state is a vector that summarizes the network's memory regarding previous inputs.
•Previous cell state (Ct-1): The cell state is also an output from the previous step t-1. It represents the internal state of the cell and behaves as a form of memory.
The gates in the LSTM are used to selectively learn and forget information over time, allowing them to retain information about longer data sequences. Below are the main types of gates:
•Forget gate (Ft): Determines what information from the previous hidden state and the current input should be forgotten. Its output is a value between 0 and 1 for each element of the hidden state, where zero is used to "completely forget" and 1 to "completely keep or remember".
•Input gate (It) or Update gate: Determines what new information from the current input should be added to the hidden state. Its output is a value between 0 and 1 for each element of the hidden state, where 0 means "completely ignore the new information" and 1 means "completely add the new information".
•Output gate (Ot): Determines what information from the current hidden state should come out. Its output is a value between 0 and 1 for each element of the hidden state, where 0 means "completely ignore the information" and 1 means "completely emit the information".

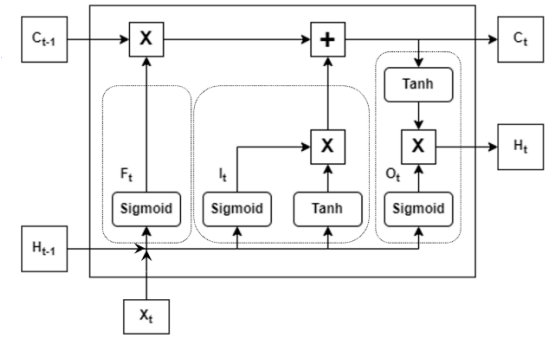

Fig. 2. *LSTM Neural Network Architecture*
Source: *(adapted from [2])*

Figure 3 shows an unrolled LSTM network for a three-element input at times t, t+1, and t+2. The input to the cell at time t consists only of the value Xt, while for the second it is the value Xt+1 and the vectors Ht and Ct, this is because the memory state is non-existent at the initial time t. The output of the LSTM cell at each moment Ht, constitutes the encodings that can then be used, for example, in a dense network for other tasks.

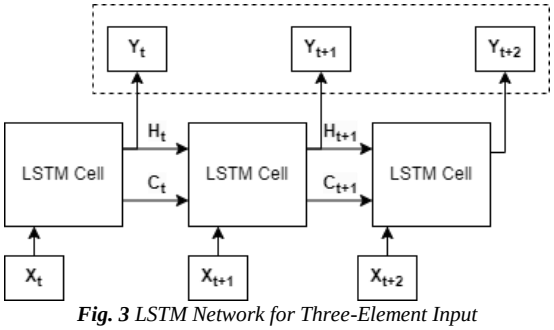

Fig. 3 *LSTM Network for Three-Element Input*
Source: *(adapted from [3])*

3.3. Siamese-LSTM neural network

In a Siamese-LSTM network, the branches are made up of identical LSTM networks, as they share the same parameters as seen in section 3.1. These networks are useful when you want to compare pairs of temporal data, such as time series or videos. Figure 4 shows the architecture of a Siamese-LSTM network, where X1 and X2 correspond to the input vectors of sequential nature, which could be phrases for example. The LSTM Cell component is a simplified representation of the LSTM neural network in which a feedback loop has been represented indicating that part of the input of an LSTM cell is constituted by the output at the previous time of the LSTM cell, each branch delivers the encodings represented by Y1 and Y2 that pass through a distance function that calculates the similarity that finally passes, in our case, through a dense network responsible for performing the classification task.

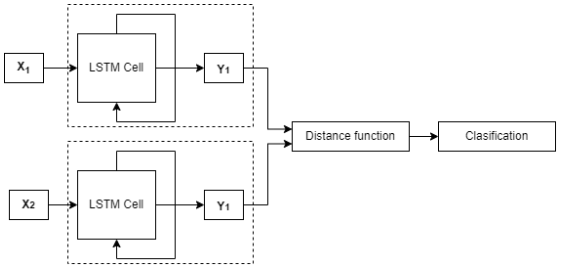

Fig. 4. *Siamese-LSTM Network Architecture*
Source: *(adapted from [4])*

3.4. Optical flow

Optical flow is a technique used to determine the apparent motion pattern of objects between two

images or video sequence, caused by the movement of the camera or the objects in the scene. It is used in the context of video processing to estimate the movement of objects in consecutive frames, analyzing changes in intensity and colors of pixels in frames over time. Optical flow algorithms can estimate the displacement of objects in a scene and create a vector field that describes the direction and magnitude of movement at each point in the frame. Figure 5 shows the RAFT model (Recurrent All Pairs Field Transforms for Optical Flow) that calculates the optical flow from two images (top left) (bottom right), which was used in this work.

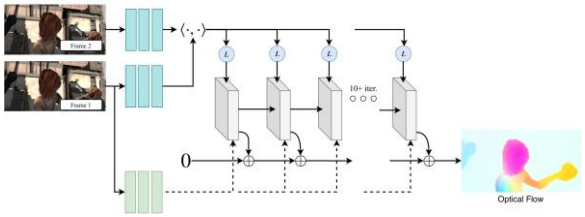

Fig. 5. *RAFT Model Architecture*
Source: *(taken from [5])*

3.5. Facial Landmark points

Facial landmark points are specific points or features on the face that can be used to identify and measure different facial features. Facial landmark points include: the tip of the nose, the outer edges of the eyes, the corners of the lips, the corners of the jaw, and the bridge of the nose and eyebrows. These points are used to measure the distance between different areas of the face, the angulation of facial features, the relationship between different parts of the face, and facial symmetry. Figure 6 shows the facial landmark points calculated for a frame from one of the videos in the Rose-youtu dataset. Generally, work is done with 68 facial points, although there is also the option to use 106 points.

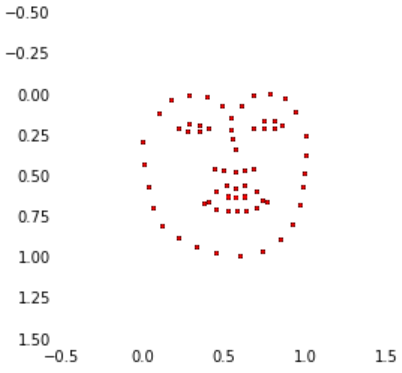

Fig. 6. *Facial Landmarks*
Source: *Own elaboration.*

4. PROPOSED MODEL

The proposed model for detecting presentation attacks in facial biometric authentication consists of a Siamese-LSTM network fed with records composed of optical flow and facial landmark information. Fig. 7 shows the general process, which involves extracting frames to obtain facial landmarks and optical flow, which are then combined to form the input records. Next, the Siamese-LSTM network is executed with the generated records to obtain the encodings, and finally, the classification process is performed to indicate whether the two videos belong to the same class (genuine or attack).

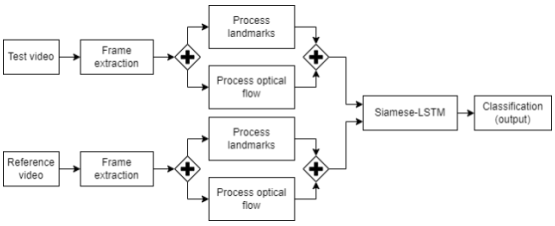

Fig. 7. *General Process*
Source: *Own elaboration.*

Each video is sampled at 30 frames per second, with 30 frames taken at intervals of 300 milliseconds. Facial landmarks are calculated for each of these frames, and optical flow is calculated for each pair of consecutive frames. Then, the facial landmarks and optical flow are concatenated into a vector of dimension 30 x 5544 (details on where these values come from are shown below). Since there is no previous frame for the first frame to calculate optical flow, the first element of the record consists of the facial landmarks of the first frame and a vector of zeros. The i-th element is composed of the facial landmarks of the i-th frame and the optical flow between the i-th frame and the previous frame.

The 68 facial landmarks are calculated using the SBR model [6], which takes an image as input and generates a vector of dimension 2x68, which is flattened into a vector of dimension 1x136.

Optical flow is calculated using RAFT. It takes two images, fj and fj-1, of the same size, which are the i-th frame and the previous frame, and generates a vector of dimension 52x52x2, which is flattened into a vector of dimension 1x5408.

To train the model, a list of pairs of videos is generated as follows: The entire dataset is traversed, and if the index of the i-th video (sample) is odd, a video of the same class (genuine or fraudulent) is randomly selected from the entire dataset. If the

index is even, the selection is made from all records of the same subject. With these videos, two pairs are generated, one consisting of the dataset sample and a reference video of the same class, and the other consisting of the sample and a video of the opposite class. The pair of samples from the same class is labeled with 1, and the other pair of samples from different classes is labeled with 0. In the end, the training dataset containing N samples generates a dataset of size 2*N. Figure 8 shows how two pairs are generated from a dataset sample (G-Sample0): one consisting of G-Sample0 and a random sample from the dataset G-SampleR with label 1, and the other consisting of G-Sample0 and a sample from the opposite class S-SampleR with label 0.

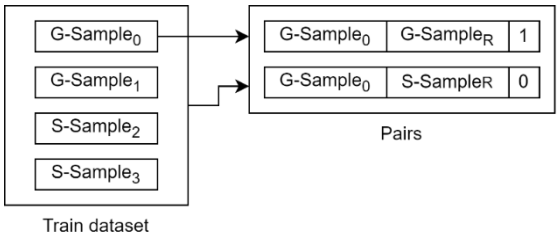

Fig. 8. *Pair Selection*
Source: *Own elaboration.*

5. EXPERIMENTATION

5.1. Data sets

To evaluate the effectiveness of the proposed model, experiments were conducted using the following datasets:

Rose-youtu: The Rose-youtu dataset [7] is a collection of videos for liveness detection created at Nanyan University. It covers a wide range of lighting conditions, camera models, and types of attacks. It consists of 4,225 videos of 25 people, with each person having between 150 and 200 video clips, averaging 10 seconds in duration. The data was collected using five different mobile devices: Hasee, Huawei, iPad 4, iPhone 5s, and ZTE, all with front cameras and a distance of 30-50 cm between the face and the camera. Impersonation attacks include attacks with printed paper, video replay attacks, and masking attacks.

Replay-mobile: The Replay-Mobile dataset [8] is a collection of 1190 videos of photos and presentation attacks from 40 people under different lighting conditions (controlled, adverse, direct, lateral, and diffuse), recorded using an iPad mini2 and an LG-G4 smartphone. The dataset is divided into 4 groups: Training, development, testing, and enrollment.

Replay-attack: The Replay-attack dataset [9] is a collection of 1300 videos of photos and presentation attacks from 50 people under different lighting conditions. It is divided into 4 groups: Training, development, testing, and enrollment. The videos were recorded with a Macbook in mov format.

5.2. Experimentation protocol

In the experimentation phase, the test sets of the replay-attack and replay-mobile datasets were used. However, the rose-youtu dataset does not have a separate test set. To address this limitation, the test set was generated using clients 13 to 18 and 20 to 23 from the dataset. For each record in the selected test sets, random pairs were formed, assigning a label of 1 if the pairs corresponded to the same class and a label of 0 if they belonged to different classes. It was ensured that for each record, a positive and a negative pair were always generated.

5.3. Metrics

Here are the metrics used to evaluate the effectiveness of the model compared to other approaches reported in the literature:

EER (Equal error rate) [10]: This is a commonly used metric for evaluating biometric systems, as well as other classification systems. It is the point on the ROC (Receiver Operating Characteristic) curve [11] where the False Acceptance Rate (FAR) is equal to the False Rejection Rate (FRR). The ROC curve is a graph that shows the True Positive Rate (TPR) versus the False Positive Rate (FPR) at different classification thresholds. TPR is the percentage of true positives correctly classified as positive, while FPR is the percentage of false positives incorrectly classified as positive.

HTER (Half Total Error Rate): Also known as the Average Classification Error Rate, it is defined as the average of the False Acceptance Rate (FAR) and the False Rejection Rate (FRR), where FAR is the proportion of negative instances classified as positive and FRR is the proportion of positive instances classified as negative [16].

5.4. Results

Table 1 shows the EER results obtained by the proposed model for each dataset, as well as those reported in the literature. Additionally, Table 2 shows the HTER results.

Table 1: *Comparison of results using EER as metric*

Dataset	Model	EER%
Replay-attack	Work in [13]	2.72
	Work in [17]	1.26
	Work in [17]	4.70
	Proposed	9.15

Source: *Own elaboration*

Tabla 2: *Comparison of results using HTER as metric*

Dataset	Model	HTER%
Rose-youtu	Work in [14]	8.13
	WA(GA+MMS+PS) [19]	5.12
	Proposed	13.24
Replay-mobile	Localised MKL [20]	5.60
	Work in [21]	8.58
	Proposed	6.70
Replay-attack	Work in [21]	0.00
	Proposed	3.75

Source: *Own elaboration*

5.5. Discussion

It can be observed from Table 1 and Table 2 that the proposed model does not exhibit competitive advantages compared to other state-of-the-art models. In Table 1, it can be seen that this approach ranks last among the 4 works using the same Replay-attack dataset, with a difference of 7.89 percentage points from the best model and a difference of 4.45 percentage points from the next model. In Table 2, it can be appreciated that the difference with the best model is 8.12 percentage points for the Rose-youtu dataset, 3.75 percentage points for the Replay-attack dataset, and 1.1 percentage points for the Replay-mobile dataset. These results indicate that the proposed model fails to outperform existing models on the various evaluated datasets.

6. CONCLUSION AND FUTURE WORK

Siamese networks have shown excellent results in scenarios where comparing two inputs and determining their degree of similarity is required, such as signature validation systems or individual identity validation systems. In these cases, an input image is compared to a well-known image, i.e., one that represents a genuine sample. Generally, these types of networks are trained and tested with positive pairs, consisting of two genuine samples, and negative pairs, consisting of one genuine sample and one fraudulent sample. In this work, we experimented by generating two pairs for each video, where the positive pairs were each formed by a record of the same class, either genuine or fraudulent, and the negative pairs were each formed by one genuine and one fraudulent record. This allowed us to experiment with the ability of this type of architecture to determine the degree of similarity between two fraudulent records, which in terms of a biometric system in production would be equivalent to, given an input video, randomly selecting a record from the test dataset, which could be genuine or fraudulent, and using these records to determine if the input is genuine or not, according to the label of the randomly selected record. However, we see from the results that this is not an appropriate approach, since at the time of registering a client, what is stored in the database is a genuine record, in the case that the registration of said client is done in a controlled environment. This leads us to another important point, which is the need for well-known records that can be taken as a reference when verifying a video, as seen in works such as those presented in [14] and [15], where the authors show the usefulness of having specific client information. And since facial biometric authentication systems are ultimately used for specific clients, previously registered in the system, it makes sense for training and testing to be done with client information. Additionally, strategies can be developed for pair selection, as can be seen in [22]. By combining the Siamese network architecture with LSTM, we aimed to build a model focused on analyzing dynamic information about subjects and their environment through facial landmarks, which identify key points of the face, and optical flow, which describes the apparent movement of objects in the scene and the camera. However, according to the results, it is necessary to work on a representation of the input records that better reflects the relationship between key points on the face and the movement of the environment. It is important to obtain a model that, from the data, identifies patterns in how facial landmarks change between frames in a genuine face and how the movement of the environment relates to the face in question. Likewise, the representation must take into account facial movements that may be influenced by the individual's own characteristics, such as excessive blinking, partial facial paralysis, among others.

ACKNOWLEDGEMENTS

The work presented in this article was partially supported by the Information Technology Research and Development Group (GTI) of the University of

Cauca. We thank Colin McLachlan for assisting with the translation of the article.

REFERENCES

[1] S. Jia, G. Guo, Z. Xu, and Q. Wang, "Face presentation attack detection in mobile scenarios: A comprehensive evaluation," Image and Vision Computing, vol. 93, p. 103826, Jan. 2020, doi: 10.1016/j.imavis.2019.11.004.

[2] S. Kumar, S. Singh, and J. Kumar, A Comparative Study on Face Spoofing Attacks. 2017.

[3] Y. Xin et al., "A survey of liveness detection methods for face biometric systems," Sensor Review, vol. 37, no. 3, pp. 346–356, Jul. 2017, doi: 10.1108/SR-08-2015-0136.

[4] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, "On Dynamic Mode Decomposition: Theory and Applications," Journal of Computational Dynamics, vol. 1, no. 2, pp. 391–421, Dec. 2014, doi: 10.3934/jcd.2014.1.391.

[5] L. Li, X. Feng, Z. Xia, X. Jiang, and A. Hadid, "Face spoofing detection with local binary pattern network," Journal of Visual Communication and Image Representation, vol. 54, pp. 182–192, Jul. 2018, doi: 10.1016/j.jvcir.2018.05.009.

[6] Z. Wang et al., "Deep Spatial Gradient and Temporal Depth Learning for Face Anti-spoofing." arXiv, Mar. 18, 2020. doi: 10.48550/arXiv.2003.08061.

[7] X. Tu et al., "Learning Generalizable and Identity-Discriminative Representations for Face Anti-Spoofing," arXiv:1901.05602 [cs], Jan. 2019, Accessed: Nov. 10, 2019. [Online]. Available: http://arxiv.org/abs/1901.05602

[8] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A Survey on Deep Transfer Learning," in Artificial Neural Networks and Machine Learning – ICANN 2018, V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, and I. Maglogiannis, Eds., Cham: Springer International Publishing, 2018, pp. 270–279.

[9] V. Ruiz, I. Linares, A. Sanchez, and J. F. Velez, "Off-line handwritten signature verification using compositional synthetic generation of signatures and Siamese Neural Networks," Neurocomputing, vol. 374, pp. 30–41, Jan. 2020, doi: 10.1016/j.neucom.2019.09.041.

[10] A. Niknam, H. K. Zare, H. Hosseininasab, and A. Mostafaeipour, "Developing an LSTM model to forecast the monthly water consumption according to the effects of the climatic factors in Yazd, Iran," Journal of Engineering Research, vol. 11, no. 1, p. 100028, Mar. 2023, doi: 10.1016/j.jer.2023.100028.

[11] A. Al Hamoud, A. Hoenig, and K. Roy, "Sentence subjectivity analysis of a political and ideological debate dataset using LSTM and BiLSTM with attention and GRU models," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 10, Part A, pp. 7974–7987, Nov. 2022, doi: 10.1016/j.jksuci.2022.07.014.

[12] L. Li, Z. Xia, J. Wu, L. Yang, and H. Han, "Face presentation attack detection based on optical flow and texture analysis," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 4, pp. 1455–1467, Apr. 2022, doi: 10.1016/j.jksuci.2022.02.019.

[13] C. Wang, B. Yu, and J. Zhou, "A Learnable Gradient operator for face presentation attack detection," Pattern Recognition, vol. 135, p. 109146, Mar. 2023, doi: 10.1016/j.patcog.2022.109146.

[14] S. Fatemifar, S. R. Arashloo, M. Awais, and J. Kittler, "Client-specific anomaly detection for face presentation attack detection," Pattern Recognition, vol. 112, p. 107696, Apr. 2021, doi: 10.1016/j.patcog.2020.107696.

[15] M. Pei, B. Yan, H. Hao, and M. Zhao, "Person-Specific Face Spoofing Detection Based on a Siamese Network," Pattern Recognition, vol. 135, p. 109148, Mar. 2023, doi: 10.1016/j.patcog.2022.109148.

[16] C. Yuan, Q. Cui, X. Sun, Q. M. J. Wu, and S. Wu, "Chapter Five - Fingerprint liveness detection using an improved CNN with the spatial pyramid pooling structure," in Advances in Computers, A. R. Hurson and S. Wu, Eds., Elsevier, 2021, pp. 157–193. doi: 10.1016/bs.adcom.2020.10.002.

[17] X. Cheng, J. Zhou, X. Zhao, H. Wang, and Y. Li, "A presentation attack detection network based on dynamic convolution and multi-level feature fusion with security and reliability," Future Generation Computer Systems, Apr. 2023, doi: 10.1016/j.future.2023.04.012.

[18] X. Shu, X. Li, X. Zuo, D. Xu, and J. Shi, "Face spoofing detection based on multi-scale color inversion dual-stream convolutional neural network," Expert Systems with Applications,

vol. 224, p. 119988, Aug. 2023, doi: 10.1016/j.eswa.2023.119988.

[19] S. Fatemifar, S. Asadi, M. Awais, A. Akbari, and J. Kittler, "Face spoofing detection ensemble via multistage optimisation and pruning," Pattern Recognition Letters, vol. 158, pp. 1–8, Jun. 2022, doi: 10.1016/j.patrec.2022.04.006.

[20] S. R. Arashloo, "Unknown Face Presentation Attack Detection via Localised Learning of Multiple Kernels." 2022.

[21] "How do Siamese Networks Work in Image Recognition? | Baeldung on Computer Science." https://www.baeldung.com/cs/siamese-networks (accessed Apr. 09, 2023).

[22] G. He, F. Li, Q. Wang, Z. Bai, and Y. Xu, "A hierarchical sampling based triplet network for fine-grained image classification," Pattern Recognition, vol. 115, p. 107889, Jul. 2021, doi: 10.1016/j.patcog.2021.107889.