DOI: 10.48011/asba.v2i1.1118

Novelty Detection Applied in Recognition of Facial Expressions

Márcio S. P. Bove * **Jés J. F. Cerqueira** **
Eduardo F. Simas Filho **

* *Feira de Santana Higher Education Unit, Feira de Santana, Brazil. E-mails: bove@ufba.br, engenhariaeletrica@gruponobre.net.*
** *Electrical Engineering Graduate Program, Federal University of Bahia, Salvador, Brazil. E-mails:{jes, eduardo.simas }@ufba.br.*

Abstract: This research investigates the capacity of the Multilayer Perceptron (MLP) and Radial Basis Function (RBF) networks in the task of Novelty Detection (ND) in the recognition of facial expressions using video resources. The video data set used is produced by professional actors in the studio with basic affective states of the human face. The Viola-Jones, Kanade-Lucas-Tomasi (KLT) and Principal Component Analysis (PCA) algorithms are used in the pre-processing phase to extract features from the face. The results evaluate the performance of the MLP and RBF networks in the ND task, using new facial expressions compatible with those used in the training phase and also examines the capacity of the networks in ND using the faces of actors never before seen by the networks. In this process, the MLP and RBF networks have an accuracy of 98% for classification task, 68% and 98% for ND with data similar to the data from the training phase and 100% for ND with totally new data. Thus, this work brings together methods and techniques applied in ND using Artificial Neural Networks (ANN) aiming at the production of interactive cognition systems in the field of affective computing, based on techniques of Artificial Intelligence (AI) and Computer Vision.

Resumo: Esta pesquisa investiga a capacidade das redes Perceptron de Múltiplas Camadas (MLP) e Função de Base Radial (RBF) na tarefa de Detecção de Novidade (DN) no reconhecimento de expressões faciais usando recursos de vídeo. O conjunto de dados de vídeo utilizado é produzido por atores profissionais em estúdio com estados afetivos básicos do rosto humano. Os algoritmos Viola-Jones, Kanade-Lucas-Tomasi (KLT) e Análise de Componentes Principais (PCA) são usados na fase de pré-processamento para extrair atributos da face. Os resultados avaliam o desempenho das redes MLP e RBF na tarefa DN, usando novas expressões faciais compatíveis com as utilizadas na fase de treinamento e também examinam a capacidade das redes em DN usando as faces de atores nunca antes vistos pelas redes. Neste processo, as redes MLP e RBF têm uma precisão de 98% para tarefa de classificação, 68% e 98% para DN com dados semelhantes aos dados da fase de treinamento e 100% para DN com dados totalmente novos. Assim, este trabalho reúne métodos e técnicas aplicadas na DN utilizando Redes Neurais Artificiais (RNA), visando a produção de sistemas interativos de cognição no campo da computação afetiva, baseados em técnicas de Inteligência Artificial (IA) e Visão Computacional.

Keywords: Novelty Detection, Neural Networks, Viola-Jones, Kanade-Lucas-Tomasi, Principal Component Analysis.

Palavras-chaves: Detecção de Novidade, Redes Neurais, Viola-Jones, Kanade-Lucas-Tomasi, Análise de Componentes Principais.

1. INTRODUCTION

Novelty detection (ND) consists of the ability to identify new or previously unknown situations, being considered an extremely complex and important task, so that many methods are proposed for ND (Pimentel et al., 2014; Markou and Singh, 2003a,b; Domingues et al., 2018). Approaches based on different categories such as probability, reconstruction, domain, information theory and distance can be used to ND (Pimentel et al., 2014). Thus, ND can be compared to a classifier that produces results for normal patterns and another for unknown patterns, where a description of normality is learned by fitting a model to the set of normal examples, and previously invisible patterns are tested by comparing their score of novelty with some decision limit (Sameer and Markou, 2004).

In different areas, as industrial monitoring, sensor networks, robotics, signal processing, computer vision, pattern recognition, text mining, information security, diagnosis and medical supervision (Pimentel et al., 2014; Oliveira, Moisés A. et al., 2020), studies in the scope of ND has contributed to the development of intelligent systems. In the field of affective computing, natural verbal and non-verbal signals of human emotions, cognitions, perceptions and behaviors are used in the production of efficient and satisfactory interaction systems between man and machine, where the human face is used of several applications (Calvo and D'Mello, 2010; Chatterjee and Chandran, 2016).

In this work methods as Viola-Jones (Viola and Jones, 2001), Kanade-Lucas-Tomasi (KLT) (Chai and Shi, 2011; Barnouti et al., 2018), Principal Component Analysis (PCA) (Liu and Kau, 2017) and Artificial Neural Networks (ANN) (Markou and Singh, 2003b; Pimentel et al., 2014) are used in the production of a compact algorithm with low computational cost with the purpose of ND in the recognition of facial expressions in real time using video feature.

In the pre-processing phase the Viola-Jones, KLT and PCA algorithms are used to extract features from the face and produce the features vector. The Viola-Jones algorithm initially detects the face on video (Viola and Jones, 2001) and consecutively the KLT algorithm identifies facial features and tracks the face throughout the video (Chai and Shi, 2011; Barnouti et al., 2018). The PCA algorithm characterized as a statistical method used to extract the main components of a data set (Liu and Kau, 2017) is employed in this work to extract features of the data obtained with the KLT algorithm, where only the first component is used for coding the human face, enabling the production of the vector for training and network testing.

In the processing phase, ANN classifies or ND in the analysis of facial expressions in video. The Multilayer Perceptron (MLP) and Radial Basis Function (RBF) networks are evaluated for their respective performance in assertiveness to ND. The used haves video database a dynamic and multimodal set of facial and vocal expressions in English assessing the emotional authenticity of professional actors for basic affective states (Livingstone, 2018).

In this way, this article investigates ND for a multi-class approach using the MLP and RBF neural networks, evaluating facial expressions similar to those used in the training phase and evaluating entirely new facial expressions. Consecutively motivated in the integration of artificial intelligence and computer vision methods and techniques to implement a compact algorithm that detects new facial expressions or classifies facial expressions in real time video stream.

The organization of the paper is as following. In section 2 are presented considerations for the use of ANN and significant aspects to enable the MLP and RBF networks for ND. Besides main aspects of the KLT and PCA algorithms to extract facial features from video are highlighting. Section 3 presents details of the data set and also presents the methods used to enable the MLP and RBF networks for ND. Section 4 presents the results obtained with the application of the algorithm developed for ND, as well as graphical results that evaluate the MLP and RBF networks for ND using facial expressions similar to the expressions used in the training phase and facial expressions of new actors. The conclusions are presented in section 5.

2. BACKGROUND

In this section, initially the aspects for ND using ANN are described. Then, the KLT and PCA methods are presented as a resource for extracting face attributes.

2.1 ANN

ANN are conceptualized as adaptive computational systems that learn from data representative of the problem by using training to adjust their synaptic weights. Its ability to learn complex boundaries to identify classes and the ability to model implicit data autonomously makes neural networks a widely used method for ND (Markou and Singh, 2003b).

In practice, ANNs do not automatically perform ND because they act as discriminators, not as detectors (Markou and Singh, 2003b), requiring training, adjustments and tests to determine limits to perform ND (Hodge and Austin, 2004). ND using ANN considers the evaluation of the results presented in the output layer of the network compared based on a limit value that makes it possible to identify when a vector displayed at the network entrance different from the vector used in the training phase (Augusteijn and Folkert, 2002; Vasconcelos and Fairhurst, 1995; Sameer and Markou, 2004).

In applications with video processing in which the same object may change gradually during operation due to different lighting conditions, exposure times and other reasons the neural networks becomes a very useful technique (Markou and Singh, 2003b). Successively because neural networks do not need data recycling as in the statistic methods for detecting new events (Markou and Singh, 2003b). In several applications, the supervised learning neural networks most used for ND in a multi-class approach are MLP and RBF (Markou and Singh, 2003b; Barreto and Frota, 2012).

Markou and Singh (2006) search in a model for ND with the image sequence analysis using neural networks. This model uses experiments with video-based image sequence data containing several novel classes. In this process, the neural network MLP is used in three configurations, with the function softmax, without rejection filter and with rejection filter. The best results are presented by the network with the rejection filter.

Vasconcelos and Fairhurst (1995) investigate the ability to ND using standard the MLP networks, MLP with Gaussian Activation Function (GMLP) and the RBF. In this evaluation is possible to identify the reasons for the unreliability of standard MLP networks in rejecting unknown patterns and consecutively that alternative configurations such as the GMLP are candidates for a more reliable structure for ND. The researchers also present technical characteristics of the RBF network that justify its greater capacity to ND compared to the MLP and GMLP networks.

Using a example from the field of speech recognition Albrecht et al. (2000) show the functioning of a generalized RBF network that can self-organize to form a Bayesian classifier and ND. For this purpose, the researchers introduce stochastic rules that concern the centers, shapes and widths of the receptive fields of neurons allowing a joint optimization of all network parameters for ND.

2.2 MLP

MLP networks are successfully applied to solve complex problems using the backpropagation algorithm. The use of the Gaussian activation function for the MLP network forces the receptive field of neurons to be more selective, being activated only for a restricted region of the input space, optimizing the performance of the network for ND (Barreto and Frota, 2012; Vasconcelos and Fairhurst, 1995). The parameterization of the GMLP network with only a single hidden layer also improves its ability for ND, so that this technique allows to detect arbitrarily complex class limits (Hodge and Austin, 2004).

In cases where the GMLP network has a few hundred neurons in the hidden layer, its sensitivity to detect new patterns can again be amplified using the training algorithm Levenberg-Marquardt (Hagan and M.B., 1994). The Levenberg-Marquardt algorithm is an approximation to Newton's Method. The use of the softmax function in the output layer of the GMLP network makes it possible to visualize the activation values of neurons in a probabilistic way, allowing to more accurately distinguish the reference value for ND (Sameer and Markou, 2004). The reference value predefined by the user comes from the test process of the trained network with the vectors used in the training phase together with the test of vectors never before seen by the network. This testing process reveals the reference activation value for the output neurons to normal patterns and also shows the different levels of activation of the output neurons for unknown vectors. In this way, an alternative for ND can be achieved with the calculation of the distance of the winning neuron for a reference value defined by the user in the test with the normal vectors (Augusteijn and Folkert, 2002).

2.3 RBF

RBF networks have been widely used in classification problems with applications in speech recognition, medical diagnosis, handwriting recognition, image processing and fault diagnosis. In these applications, RBF networks are often used with the "Winner Takes All" (WTA) (Pont and Barrie Jones, 2002) output rule. Its only hidden layer uses functions with a radial base that make it possible to model high-dimensional spaces with higher performance for learning speed, memory requirements and generalization in comparison with MLP (Vasconcelos and Fairhurst, 1995).

The use of the Bayesian regularization training algorithm self-organizes the RBF network to form a Bayesian classifier and successively amplifies its capacity for ND (Pimentel et al., 2014; Albrecht et al., 2000). Bayesian regularization minimizes a linear combination of squared errors and weights being to able produce networks which have excellent generalization capabilities. This Bayesian regularization takes place within the Levenberg-Marquardt algorithm where it calculate the Jacobian matrix (Foresee and Hagan, 2017). Similar to the MLP network, we can use the softmax function to provide probabilistic values for the output neurons of the RBF network and define the decision threshold value for ND.

2.4 KLT

Object detection and tracking from a video-sequence are becoming an interesting subject in many computer vision applications and artificial intelligence, as bank security, border crossings, airport check-in, home monitoring, office remote meeting, prisons and factories (Barnouti et al., 2018; Chatterjee and Chandran, 2016). The KLT method to allows track a set of feature points in video frames. Its computational efficiency and robustness to scale change are relevant aspects that make its use widely feasible in the development of computer vision systems (Barnouti et al., 2018; Chatterjee and Chandran, 2016).

The structure of the KLT algorithm is composed of two phases of operation, features extraction and tracking (Chai and Shi, 2011). In the features extraction phase, initially the Viola-Jones algorithm detects the face in the video (Barnouti et al., 2018; Chatterjee and Chandran, 2016) and feature points are identified around of face. With the feature points identified on image, the task of feature tracking is to track them from frame to frame (Chai and Shi, 2011).

2.5 PCA

The human face has a complex and dynamic structure (Delac et al., 2005), being the one of the most important information in biometric sciences based on personal identification (Abbas et al., 2017). The technique of analysis and understanding of images have gained prominence in recent years with successful in applications of facial recognition (Delac et al., 2005). The PCA algorithm allows the identification of patterns in the data maintain their identification status and effectively reducing the dimensions in human face images (Abbas et al., 2017). This reducing eliminates information irrelevant or redundant for to arrives in a higher compression ratio for the first component (Liu and Kau, 2017) producing a low-dimensional representation of the input data without significant loss of the original data. The mathematical formula of PCA is based in the standard deviation, and the eigenvectors and eigenvalues, seing considered a robust technique with a process simple, fast and what works well under constrained environment for facial recognition (Abbas et al., 2017).

3. METHODOLOGY

Novelty detection or novelty recognition is based on the normal class modeling technique, that is, a description of normal data is learned by the network so that the it is able to detect new events. This methodology is appropriate for ND with static or dynamic data (Hodge and Austin, 2004). Therefore, in this section, initially are presented the details of the data set used to model the normal classes and for ND, and consecutively the methods used in the training and testing phase of the MLP and RBF networks. In Figure 1 is show the model used for ND applied in recognition of facial expressions using of video resource.

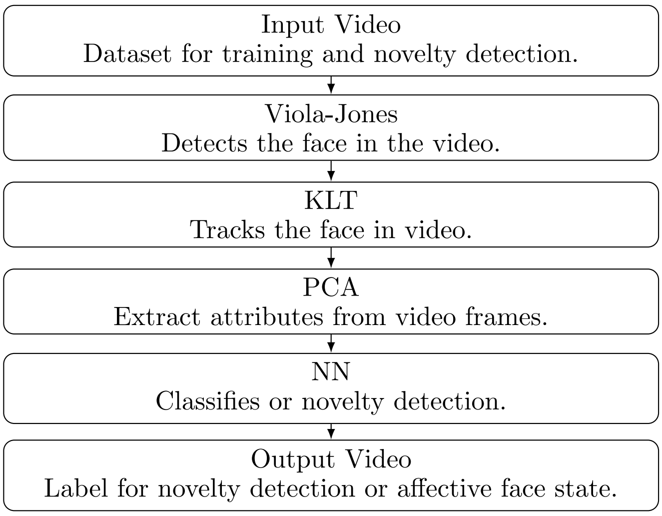

Figure 1. Model applied to novelty detection in video resource.

3.1 Dataset

The database proposed as visual resource for research, account with a selection dynamic and multimodal facial and vocal expressions assessing emotional authenticity of 24 professional actors (12 women and 12 men). The emotions calm, happy, sad, angry, fear, surprise and disgust are included. Each expression is selected on two levels of emotional arousal (normal and strong) with an additional neutral expression. All conditions are available in three data formats: audio only (16 bits, 48kHz, Type .wav), audio-video only (Width 1280 pixels, Height 720 pixels, Rate 29.97 fbs, AAC 48kHz, Type .mp4) and video-only (no sound) (Livingstone, 2018). The Figure 2 and Table 1 show the selection of video resources used in the training and testing phases of the MLP and RBF networks. In Figure 2 on the left side column, the sequence (happy, sad, angry and neutral) of actor 1 is used in the training phase of the network, in the validation of the classification of affective states and in the analysis process to determine the decision threshold for ND. On the center, the sequence (calm, fear, surprise and disgust) of actor 1 are used for ND with affective states similar to used in the training phase. On the right side, the sequence (happy, sad, calm and happy) for new actors are used for ND in totally new conditions.

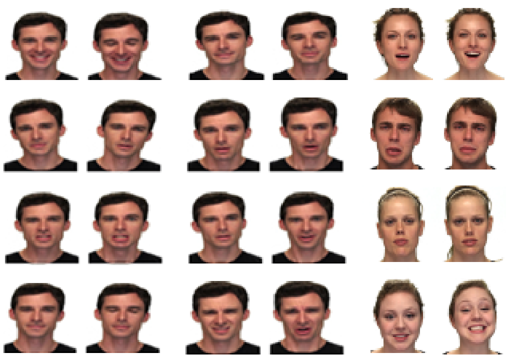

Figure 2. Actors of the Database (Livingstone, 2018) used for training and ND with MLP and RBF.

Table 1. Database used for training and ND of MLP and RBF networks.

Video	Number	Phase	Affection	Frames
Actor 1	3	Training	Happy	399
Actor 1	3	Training	Sad	426
Actor 1	3	Training	Angry	426
Actor 1	3	Training	Neutral	294
Actor 1	3	Test ND	Calm	294
Actor 1	3	Test ND	Fear	336
Actor 1	3	Test ND	Surprise	318
Actor 1	3	Test ND	Disgust	351
Actor 2	3	Test ND	Happy	312
Actor 5	3	Test ND	Sad	322
Actor 6	3	Test ND	Calm	310
Actor 10	3	Test ND	Happy	320

3.2 Training Phase

In the first phase, the face is segmented by the holistic method, that is, the face is processed as a whole, considering the facial information of the nose, eyes, mouth and hair. In this process, Viola-Jones algorithm initially detects and locates the face in a video frame and in the sequence feature points are used to map the face region.

In the second step, KLT algorithm tracks the set of points considering the bounding box around the face to estimate the position of the face frame by frame according to the dynamics of the video.

In the third phase, PCA algorithm is used to calculate the first axis of the main component by extracting and reducing the dimensionality of the frames with low loss of information.

In the fourth phase, MLP and RBF networks are parameterized and trained for ND considering a single hidden layer according to the topology of Figure 3.

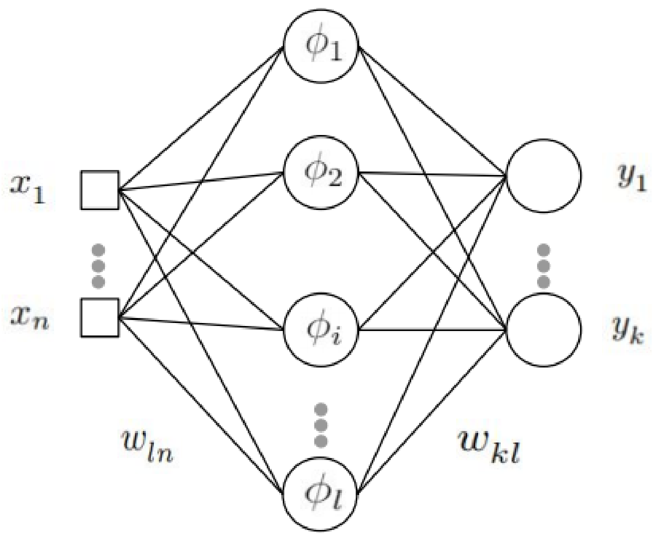

Figure 3. Topology of MLP and RBF networks.

For MLP network, the input layer own 6400 neurons, the only hidden layer own 190 neurons with the Gaussian activation function, the output layer own 4 neurons and the network synaptic weights are adjusted by the Levenberg-Marquardt training algorithm. In this case, the parameterization of the hidden layer with only 190 neurons is justified for the use of the Levenberg-Marquardt training algorithm, which has better efficiency when the network has a few hundred neurons in the hidden layer and consecutively amplifies the MLP network capacity for ND (Hagan and M.B., 1994). The induced local field of the neuron $v_i(n)$ adjusts the height of the Gaussian function and the constant γ_i adjusts the radius of the function. Where the outputs are obtained by $y_k(\mathbf{x},\mathbf{w})$ with $\mathbf{x} = [x_1, x_2, x_i, ...x_n]^T$ being the attribute vector.

$$v_i(n) = \sum_{i=1}^{n} w_{ji}(n)x_i(n) \tag{1}$$

$$\phi_i(\mathbf{x},\mathbf{w}) = exp\left[-\frac{v_i(n)^2}{\gamma_i^2} \right] \tag{2}$$

$$y_k(\mathbf{x},\mathbf{w}) = \sum_{j=1}^{l} w_{kj}\phi_i(\mathbf{x},\mathbf{w}) \tag{3}$$

For the RBF network, the input layer own 6400 neurons, the only hidden layer own 100 neurons with the Gaussian activation function, the output layer own 4 neurons and the network synaptic weights are adjusted by the Bayesian Regularization training algorithm. The Bayesian regularization use Gauss-Newton approximation to the Hessian matrix. The additional overhead of this Gauss-Newton approximation is minimal when Levenberg-Marquardt optimization algorithm is used to locate the optimal weights. This training method reduce the sum of squared errors and produces small values for synaptic weights, thus result a smoother network response that intensifier the capability to ND (Foresee and Hagan, 2017). The induced local field of the neuron $v_i(n)$ adjusts the Gaussian function at the center of the observed data. This process occurs by calculating the Euclidean distance $\parallel x_i(n) - c_i(n) \parallel$, considering the input vector $\mathbf{x} = [x_1, x_2, x_i, ...x_n]^T$ and the observed data centers $\mathbf{c} = [c_1, c_2, c_i, ...c_n]^T$. The standard deviation σ_i adjusts the radius of the Gaussian function.

$$v_i(n) = \parallel x_i(n) - c_i(n) \parallel \tag{4}$$

$$\phi_i(\mathbf{x},\mathbf{c}) = exp\left[-\frac{v_i(n)^2}{2\sigma_i^2} \right] \tag{5}$$

$$y_k(\mathbf{x},\mathbf{c}) = \sum_{j=1}^{l} w_{kj}\phi_i(\mathbf{x},\mathbf{c}) \tag{6}$$

At the end of the training phase, the function $softmax(y_i)$ is used to calculate the level of activation of the output neurons of the MLP and RBF networks to obtain values with a degree of probabilistic confidence, in order to create thresholds reliable for ND. Thus, the new outputs for ND are obtained using equation (8).

$$softmax(y_i) = \frac{e^{y_i}}{\sum_k e^{y_k}} \tag{7}$$

$$z(y_i) = softmax(y_i) - \delta_i \tag{8}$$

3.3 Testing Phase

After training the networks, they are tested with the vectors of the training phase and new vectors. Vectors of the training phase are used to check the accuracy of the networks in the classification task and consecutively to define the threshold vector for ND. The limit vector $\delta(n) = [\delta_1, \delta_2, \delta_i, ...\delta_k]^T$ for ND is defined by looking at the activation levels of the winning neurons of each class. The lowest level of activation among the winning neurons of each class is defined as the decision threshold for ND. The new vectors are used this process to check the sensitivity of the networks to ND. In this condition, the respective outputs present an activation level for the winning neuron lower than the level defined as threshold for ND. Thus, when all outputs $z(y_i)$ simultaneously have an activation level lower than zero, a novelty is detected.

4. RESULTS

In this section, are presented the results obtained with the proposed algorithm for classification and ND in recognizing facial expressions with video resource. In Figure 4 at the top the Viola-Jones and KLT algorithms are used to extract features from the face by the holistic method. At the bottom, the PCA algorithm is used to calculate the first axis of the main component, extracting and reducing the dimensionality of the frames.

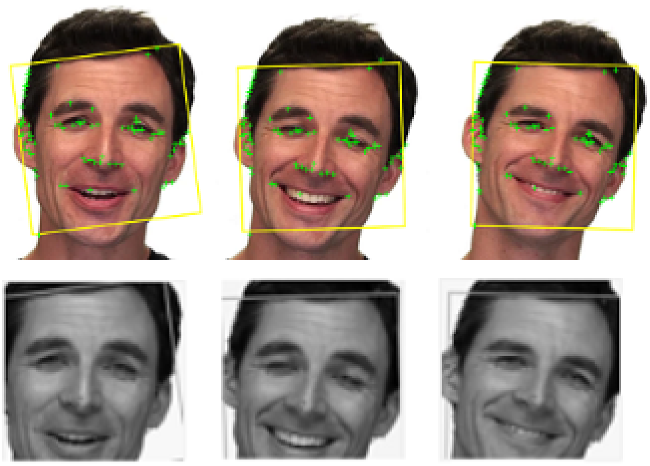

Figure 4. Viola-Jones, KLT and PCA extracting face attributes in video.

The Figures 5 and 6 show the results obtained for the classification test with the MLP and RBF networks considering the affective states (happy, sad, angry and neutral) of the actor 1 as described in Table 1. These results represent the learning of normal patterns, qualifying the MLP and RBF networks to classify the affective states of actor 1 with 98% accuracy when evaluating video frames. Confusions occur between sad and angry affective states owing to the similarity of some facial expressions.

In Figure 7, the proposed algorithm evaluates actor 1 facial expressions. Application of the tag respective affective state is produced in real time, confirming the efficiency in the integration of Viola-Jones, KLT, PCA and ANN methods.

	Happy	Sad	Angry	Neutral
Happy	399	0	0	0
Sad	0	411	15	0
Angry	0	15	411	0
Neutral	0	0	0	294

Figure 5. Results of the MLP network for the classification of Happy, Sad, Angry and Neutral affective states represented by the confusion matrix.

	Happy	Sad	Angry	Neutral
Happy	399	0	0	0
Sad	0	408	12	0
Angry	0	18	414	0
Neutral	0	0	0	294

Figure 6. Results of the RBF network for the classification of Happy, Sad, Angry and Neutral affective states represented by the confusion matrix.

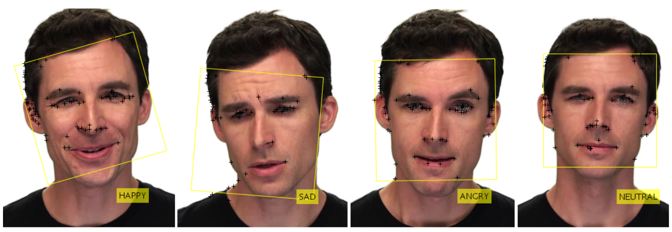

Figure 7. Classification in video resource for facial expressions to Actor 1 (Happy, Sad, Angry and Neutral).

In Figure 8, the algorithm evaluates actor 1 facial expressions for ND that are similar to the facial expressions used in the training phase. This selection of frames of actor 1, introduces in the MLP and RBF networks small disturbances that are more difficult to be recognized as novelty. At the bottom of the Figure 8, the selection of frames presents results for ND with facial expressions of new actors that produce into the MLP and RBF networks gross disturbances.

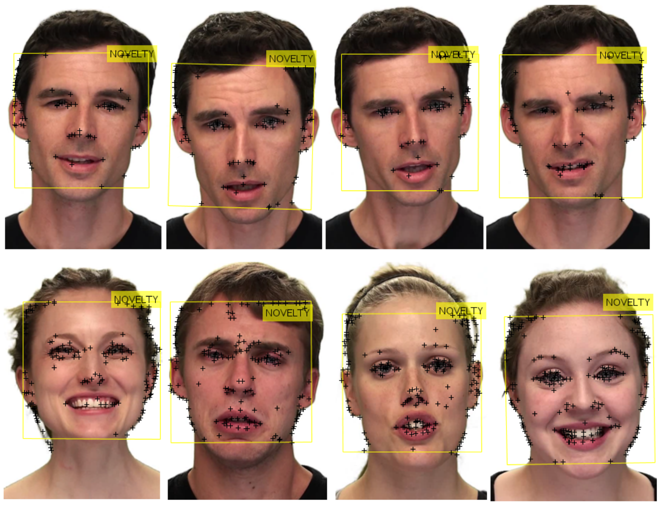

Figure 8. Novelty detection in video resource for facial expressions to Actor 1 (Calm, Fear, Surprise and Disgusted), Actor 2 (Happy), Actor 5 (Sad), Actor 6 (Calm) and Actor 10 (Happy).

The results presented in Figures 9 and 10 for ND are obtained using the MLP network based on the data described in Table 1. The algorithm produces results with an efficiency of 100% for the detection of new actors. However, when evaluating actor 1's videos, this efficiency drops for 68%. This less accurate result is represented by video frames that have not been recognized as new. The video frames not recognized as new are classified with the closest facial expression used in the training phase. These results confirm the robustness of the MLP network for ND when input data produces gross errors and stability for a range when the input data produces a small disturbance.

That way, a system can perfectly adjust the noise to the data provided, leading to the phenomenon of overfitting. In overfitting situations, the learning process may be able to achieve lower training error levels because the learned function has tried to fit all data as close as possible (Chen-Chia Chuang et al., 2000). However, another set of data, which is not used in any way in the training process, have errors significantly be increased because the noise in the training patterns is different from that in the testing patterns (Chen-Chia Chuang et al., 2000).

The experimental results produced by Markou and Singh (2006) for ND in the analysis of video image sequences based on MLP neural networks show different levels of performance according to the Z metric. These results are generated using four different video sequences, where the best performance of Z with 92% occurs in the experiment with the second video sequence. Considering the results produced in this research and the results obtained in the experiment carried out by Markou and Singh (2006) it is confirmed that the MLP network can be adapted and adjusted to be used as an efficient alternative to perform ND.

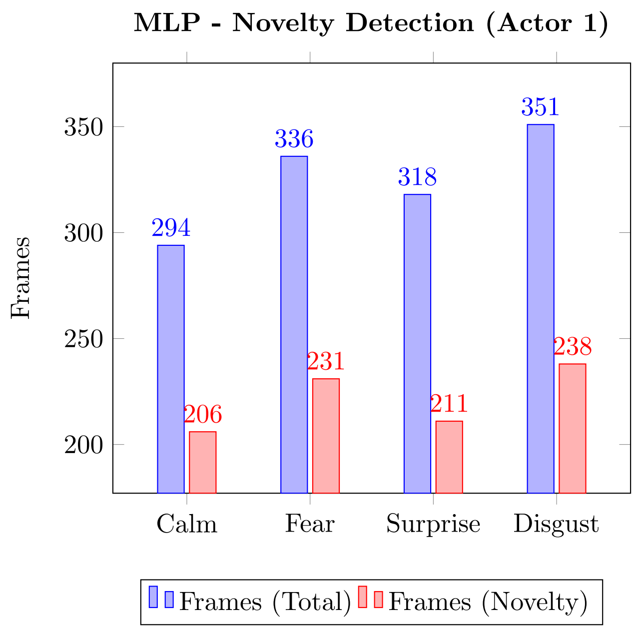

Figure 9. ND for video frames with Actor 1 (Calm, Fear, Surprise and Disgust).

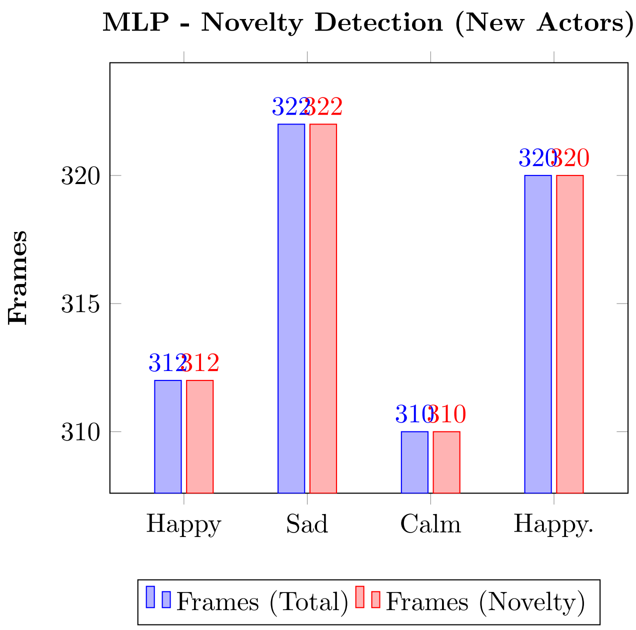

Figure 10. ND for video frames with Actor 2 (Happy), Actor 5 (Sad), Actor 6 (Calm) and Actor 10 (Happy).

In Figures 11 and 12 the algorithm produces results with an efficiency of 100% for the detection of new actors, but when evaluating actor 1's videos, this efficiency reduces for 98% owing the classification of facial expressions that are similar to those used in the training phase.

Different on the MLP network, each hidden unit in the RBF network responds to a receptive field located in the input space according to the Euclidean distance calculation. As a result, the network output reaches its maximum when the input pattern is close to the centroid and decreases monotonically when it is further away from the centroid, making it ideal for ND (Markou and Singh, 2003b). The self-organized Bayesian training algorithm updates the synaptic weights according to the Levenberg-Marquardt optimization (Foresee and Hagan, 2017), minimizing the errors and consecutively producing a robust response in the analysis of gross errors or small.

The results produced for ND by Albrecht et al. (2000) with the RBF network for voice recognition are based on a cumulative distribution function of log-likelihood. Where new uniformly distributed voice patterns are safely rejected and new patterns similar to training data are not safely detected for phoneme transitions. These results confirm the efficiency of the RBF network as a Bayesian classifier for ND.

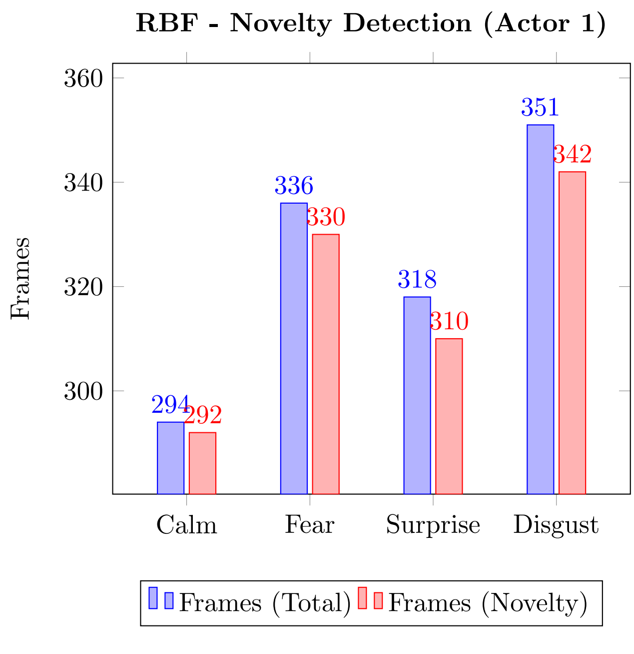

Figure 11. ND for video frames with Actor 1 (Calm, Fear, Surprise and Disgust).

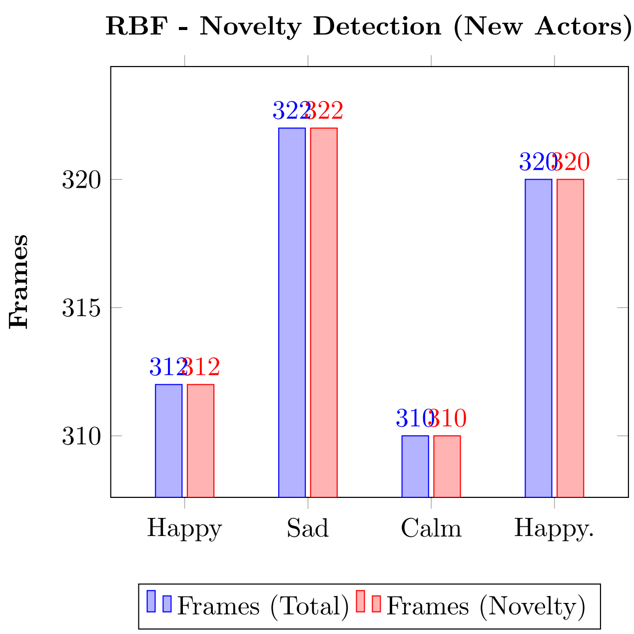

Figure 12. ND for video frames with Actor 2 (Happy), Actor 5 (Sad), Actor 6 (Calm) and Actor 10 (Happy).

5. CONCLUSION

This research investigated the capacity of MLP and RBF networks in the task of ND by evaluating facial expressions in video resources. The attributes of facial expressions are obtained by the holistic method, using the Viola-Jones, KLT and PCA algorithms, which extract attributes of the face in real-time video. The results confirm the 98% efficiency of the RBF network in ND for video frames with attributes close to the frames used in the training phase and efficiency of 100% for video frames not similar to the training data. The results of the MLP network are also validated in 100% for ND for frames without similarity to the frames used in the training phase of the network. However, when similar frames are presented for the MLP network, their efficiency reduces to a success rate of 68% to ND. Thus, this work with the integration of Viola-Jones, KLT, PCA and ANN methods confirms the ND in the evaluation of facial expressions in video resources in order to contribute to the development of intelligent systems in the field of affective computing.

REFERENCES

Abbas, E.I., Safi, M.E., and Rijab, K.S. (2017). Face recognition rate using different classifier methods based on pca. In *2017 International Conference on Current Research in Computer Science and Information Technology (ICCIT)*, 37–38.

Albrecht, S., Busch, J., Kloppenburg, M., Metze, F., and Tavan, P. (2000). Generalized radial basis function networks for classification and novelty detection: self-organization of optimal bayesian decision. *Neural Networks*, 13, 1075–1076.

Augusteijn, M.F. and Folkert, B.A. (2002). Neural network classification and novelty detection. *International Journal of Remote Sensing*, 23, 2891–2893.

Barnouti, N.H., Al-Mayyahi, M.H.N., and Al-Dabbagh, S.S. (2018). Real-time face tracking and recognition system using kanade-lucas-tomasi and two-dimensional principal component analysis. *2018 International Conference on Advanced Science and Engineering (ICOASE), Duhok*, 24–28.

Barreto, A. and Frota, A. (2012). A unifying methodology for the evaluation of neural network models on novelty detection tasks. *Pattern Analysis and Applications*, 16, 85–89.

Calvo, R. and D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1, 1.

Chai, Z. and Shi, J. (2011). Improving KLT in embedded systems by processing oversampling video sequence in real-time. *2011 International Conference on Reconfigurable Computing and FPGAs*, 297–298.

Chatterjee, D. and Chandran, S. (2016). Comparative study of camshift and KLT algorithms for real time face detection and tracking applications. *2016 Second International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata*, 63–64.

Chen-Chia Chuang, Shun-Feng Su, and Chin-Ching Hsiao (2000). The annealing robust backpropagation (arbp) learning algorithm. *IEEE Transactions on Neural Networks*, 11(5), 1067–1077.

Delac, K., Grgic, M., and Grgic, S. (2005). Independent comparative study of pca, ica, and lda on the feret data set. *International Journal of Imaging Systems and Technology*, 15, 252–254.

Domingues, R., Filippone, M., Michiardi, P., and Zouaoui, J. (2018). A comparative evaluation of outlier detection algorithms: Experiments and analyses. *Pattern Recognition*, 74, 407–408.

Foresee, F.D. and Hagan, M.T. (2017). Gauss-newton approximation to bayesian learning. *Proceedings of International Conference on Neural Networks (ICNN'97)*, 3, 1930–1935.

Hagan, M. and M.B., M. (1994). Training feedforward networks with the marquardt algorithm. *IEEE Transactions on Neural Networks*, 5, 989–993.

Hodge, V.J. and Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22, 85–122.

Liu, Y. and Kau, L. (2017). Scalable face image compression based on principal component analysis and arithmetic coding. *2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)*, 256.

Livingstone, R. (2018). The ryerson audio-visual database of emotional speech and song. . URL `https://doi.org/10.1371/journal.pone.0196391`.

Markou, M. and Singh, S. (2003a). Novelty detection: a review—part 1: statistical approaches. *Signal Processing*, 83, 2481.

Markou, M. and Singh, S. (2003b). Novelty detection: a review—part 2: neural network based approaches. *Signal Processing*, 83, 2499–2503.

Markou, M. and Singh, S. (2006). A neural network-based novelty detector for image sequence analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10), 1664–1675.

Oliveira, Moisés A., Simas Filho, Eduardo F., Albuquerque, Maria C. S., Santos, Ygor T. B., Da Silva, Ivan C., and Farias, Cláudia T. T. (2020). Ultrasound-based identification of damage in wind turbine blades using novelty detection. *Ultrasonics*, 1–2.

Pimentel, A., Clifton, D., Clifton, L., and Tarassenko, L. (2014). A review of novelty detection. *Signal Processing*, 99, 217–232.

Pont, L.Y. and Barrie Jones, N. (2002). Improving the performance of radial basis function classifiers in condition monitoring and fault diagnosis applications where 'unknown' faults may occur. *Pattern Recognition Letters*, 23, 569–572.

Sameer, S. and Markou, M. (2004). An approach to novelty detection applied to the classification of image regions. *IEEE Transactions on Knowledge and Data Engineering*, 16, 396–397.

Vasconcelos, G.C. and Fairhurst, D.B. (1995). Investigating feedforward neural networks with respect to the rejection of spurious patterns. *Pattern Recognition Letters*, 16, 208–209.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1, 511.