European Journal of Interdisciplinary Research and Development
Volume-17 July 2023
Website: www.ejird.journalspark.org ISSN (E): 2720-5746

REDUCTION AND EXTRACTION OF FACIAL FEATURES IN IMAGES USING LINEAR DISCRIMINANT ANALYSIS (LDA) AND PRINCIPAL COMPONENT ANALYSIS (PCA)

Wafaa Razzaq
College of Nursing - University of Thi-Qar, Iraq
wafaa_razzaq@utq.edu.iq

ABSTRACT

This research paper shows the implementation of two algorithms for the reduction and extraction of characteristics in images: the Principal Component Analysis (PCA) algorithm and the Linear Discriminant Analysis (LDA) algorithm, in the public database known as Cohn-Kanade (CK +) as part of the progress of the research work on the detection and classification of facial expressions with a Support Vector Machine (SVM). These two algorithms were chosen because they are the most used in the literature, where it is empirically demonstrated that they make a good reduction in the data dimension. This paper reports a part of the methodology used for the classification and detection of facial expressions, and remarks the importance of reducing the dimension of the data in a dimensional space. It is of great relevance to know if these data are really representative of the original set, and if they contribute with images that are representative of the expressions (anger, happiness, sadness, fear, surprise, neutral) to be compared. It is shown that for this specific case, LDA carries out a better grouping of data thanks to the fact that a large number of images that represent each of the facial expressions can be provided, additionally, it is a supervised algorithm.

Keywords: IAR, Dimensionality, Reduction, Extraction, PCA, LDA.

INTRODUCTION

In recent years, various methodologies and strategies have been proposed to improve the classification of facial expressions considering 3 general steps: preprocessing, feature extraction and classification for 2D and 3D images [1], [2], [3], [17]. The recognition of facial expressions is one of the most important problems for the human being, and to which special attention has been paid nowadays. Facial expressions form a key piece in non-verbal communication [4], and have been shown to convey emotions and intentions. This area is very interesting and challenging, since said expressions are subjective and even similar in some situations; Likewise, they cause confusion when they are interpreted. In recent years, various automatic systems have been developed that help detect different facial expressions. This type of system is characterized by estimating the emotional state of a person.

Such systems carry out training from a set of images, to which a pre-processing stage is applied and, subsequently, an extraction of facial features. These characteristics are classified into six emotions considered universal: happiness, sadness, anger, disgust, surprise and fear; Finally, a prediction is made of the emotion shown by the image. This technique could be applied in advertising, to analyze a person's face while viewing either an advertisement, a television commercial or a product in a store; It would also be possible to use it as a medical tool aimed at people with autism or Alzheimer's, where it is useful to monitor a patient's facial expression when faced with a certain stimulus, with the purpose of helping in the diagnosis or follow-up during the disease. Another possible application would be as a tool in police interrogations or in the psychological evaluation of a person.

As an example of a development of a framework for the classification of facial expressions [5], the following technique is shown, which consists of two steps: first, a feature extraction and, later, its classification. From an input image, certain points of interest are located on a face with the purpose of dividing it into seven triangles, which will be the support points used as base descriptors. For the extraction of local and global characteristics, a Fourier transformation combined with the Linear Discriminant Analysis (LDA) algorithm is performed, with which the linear discriminants are obtained. Then, in the classification part, an own kernel is applied in a Support Vector Machine (SVM). This framework was evaluated with the JAFFE, Cohn-Kanade (CK) and FER-2013 databases, where different successful measurements were obtained in each of the six emotions evaluated. A way to implement a facial expression recognition system, it is through the methodology [6] , [18]. In this methodology, the Principal Component Analysis (PCA) algorithm is applied to a cropped image of the face, with the aim of obtaining its main components and reducing its dimension. Then, a feature extraction is carried out with Gabor waves or with Local Binary Patterns (LBP). Once this is done, the information is stored in a database which is then used to classify the expressions with SVM, thus predicting in real time the emotion shown in the image (captured with a webcam).

In the proposed model, different success rates are obtained in public databases. Using the Gabor wave method, a percentage of 84.17% was obtained in JAFFE, 93.00% in MMI, and 85.83% in Cohn Kanade. In the JAFFE database, 88.00 was obtained %, 88.16 % in MMI, and 96.83 % in Cohn Kanade, using the LBP feature extraction technique. It is also possible to see the implementation of an automatic facial expression recognition system for 3D faces in [7]. Here the facial feature extraction technique known as PCA is combined with an SVM classifier. In this work, a binary SVM algorithm and a multi-class one are implemented, in order to carry out the classification of facial expressions. The extraction task with 3D PCA is supported by the Mean Reference Points (MLPs) algorithm, which locates points in areas of the face and, based on these points, performs an alignment and extracts the face in a very delimited way.

In this research it is concluded that using feature extraction algorithms, as well as having a greater number of faces with more diversity in expressions, is vital to carry out a better classification and save computational time. Finally, the possibility of improving the system is shown, to classify expressions in faces showing different poses, degrees of occlusion and lighting.

In [8] a proposed three-step scheme for a facial expression recognition system is shown. The first step consists in the extraction of characteristics through the wavy transformation type II (ripplet-II), whose implementation is hybrid, since it combines the extraction of appearance and geometric characteristics. The ripplet-II algorithm returns high dimensional space coefficients, for this reason, in the second step PCA and LDA are combined in order to reduce the dimension of the data, this returns a vector of 6 features and improves computational time. As the third and last step, the features are classified using the LS-SVM algorithm, where a Radial Base Function

(RBF) kernel is used. The tests and validations of this system were carried out with public databases known as CK+ and JAFFE. Despite having obtained a 98.97 success rate % and 99.46 % respectively, in said public databases, results are not documented with images that do not have controlled lighting conditions; therefore, it is assumed that real life images affect the performance of the system, this demonstrates the existence of opportunities for improvement. The length of the data to be analyzed is a performance issue for many machine learning algorithms. That is why the main objective of this research work is to reduce the amount of data that an algorithm machine learning must process. It is said, then, that the input data set is d-dimensional, however, it seeks to project a new k-dimensional data set, where k < d, here it is important to know if the new size and feature space represent the data well. To illustrate the above, it is proposed to implement two feature extraction and reduction algorithms known as Principal Component Analysis (PCA), proposed by [9], and Linear Discriminant Analysis (LDA), proposed by [10]. The research describes each algorithm to understand its operation, and compares its results in order to have a starting point to later choose one of them and apply it in the future with a classification algorithm; such implementations are based on [11] for PCA and [12] for LDA.

METHODOLOGY

Obtaining the images

The first image acquisition was made from public databases. The first database that was used was [13], known as Cohn-Kanade (CK+). This database provides a vast collection of sequential images of people displaying an expression, both posed (non-spontaneous) and spontaneous; It starts from an expression that goes from a neutral state, until exhibiting a facial expression that denotes a specific emotion, as shown in figure 1.

Figure 1. Extract of images with signs of surprise from the public database Cohn–Kanade (CK+)

This database has a different number of sequences and expressions in each person; likewise, each individual has a different number of facial expressions denoting the same emotion. Therefore, it was decided to discretionally select the images that best represent a specific emotion. The procedure and criteria were as follows: 1) select three images of each individual that show different intensities of an emotion, 2) not take into account that the same individual appears in each type of facial expression (due to the inconsistency of the database); 3) separate the images taking into account only the facial expressions. As shown in figure 2, there are 66 separate images of the same type of emotion that reflect 3 types of intensities per person: a total of 22 different people showing the same emotion with three different intensities.

After manually extracting each of the images from the Cohn-Kanade (CK+) database, these images were separated into 6 folders named after the name of the emotions known as anger, happiness, fear, neutral, surprise, and sadness, as shown in figure 4. Thus, a new database was obtained with a total of 396 images of 640 x 490 pixels each.

Figure 2. Extraction of faces with visible features of surprise.

Figure 3. Folders where the images that make up the Cohn-Kanade (CK+) public database are stored.

Figure 4. Extraction of faces with visible features of surprise.

Face extraction

Once the images were obtained in the acquisition stage, the face was extracted from each of the images from the databases. An algorithm that performs this task automatically was developed based on the algorithm proposed by [14] , [19] , the images are stored on the local disk in a directory called Database. Inside this directory are the other folders with the manually classified images, as shown in Figure 3. This algorithm iterates through each of the folders in the database and proceeds to read each of the images that are stored. Subsequently, it detects the face and cuts out the area of interest, resizes it to 50 × 50 pixels, converts it to grayscale and stores it in the output location. When finished processing all the images in a folder, the output location is updated to save the next set of images to another folder and continue processing on the rest of the folders and images. The total collection has 396 images of 50 × 50 pixels, separated into 6 folders.

The reasons for processing the images in this way were the following:

- to. Detecting the area of the face helps to eliminate those areas of the image that are not relevant for distinguishing facial expressions.
- b. An image of 2500 pixels is a good representation of the face and relevant information that helps to identify different facial expressions is not lost.
- c. Although some images appear to be grayscale, the image actually has all 3 color channels; by converting it to grayscale, the processing time is considerably reduced.

Prepare the training set

The newly formed data set consists of 396 rows and 2500 columns. The 50x50 pixel images make up the 2500 columns, plus an additional label column is added which is essentially a class label to indicate whether the instance in each row belongs to a class (1 angry, 2 happy, 3 sad, 4 surprised, 5 scared, 6 neutral). Each component of the row contains a value between 0 and 255, this describes the intensity of each pixel, as shown in figure 5.

	labels	pixel0	pixel1	pixel2	pixel3	pixel4	pixel5	pixel6	pixel7	pixel8	...
83	2	34.0	23.0	21.0	24.0	19.0	21.0	31.0	31.0	40.0	...
84	2	58.0	73.0	53.0	31.0	35.0	28.0	19.0	19.0	19.0	...
386	6	62.0	83.0	79.0	76.0	77.0	53.0	49.0	53.0	53.0	...
387	6	62.0	80.0	79.0	81.0	76.0	59.0	48.0	51.0	53.0	...
388	6	65.0	80.0	79.0	82.0	75.0	63.0	49.0	53.0	55.0	...
389	6	53.0	65.0	67.0	66.0	56.0	36.0	15.0	14.0	13.0	...
390	6	54.0	68.0	65.0	65.0	54.0	31.0	15.0	15.0	13.0	...
391	6	55.0	67.0	66.0	62.0	50.0	24.0	14.0	15.0	16.0	...
392	6	23.0	23.0	22.0	25.0	32.0	37.0	41.0	43.0	74.0	...
393	6	24.0	19.0	20.0	23.0	27.0	34.0	40.0	39.0	71.0	...
394	6	22.0	20.0	20.0	27.0	30.0	39.0	41.0	41.0	70.0	...
395	6	127.0	134.0	87.0	63.0	63.0	55.0	39.0	64.0	88.0	...
0	1	1.0	63.0	38.0	50.0	47.0	34.0	34.0	35.0	46.0	...
1	1	64.0	82.0	56.0	40.0	50.0	48.0	33.0	30.0	41.0	...
2	1	61.0	73.0	41.0	43.0	49.0	38.0	31.0	35.0	41.0	...

Figure 5. Representation of the images from the database used in table form.

Linear Discriminant Analysis

It is one of the most popular supervised algorithms for data dimension reduction in the preprocessing of large amounts of information, which is subsequently processed by a pattern classification algorithm or machine learning. The objective of this algorithm is to project the original data into a new subspace of features that is of a lower dimension, in order to achieve a good reduction of the information and, at the same time, to achieve a better separation between the data.

Because LDA takes into account the nature of the data a priori —that is, it classifies them first—, it maximizes the separation in order to avoid overfitting in the classification algorithm and thus reduces computational costs. The implementation of the LDA algorithm can be summarized in the 5 steps shown below:

1. Calculate the d-dimensional average vectors for each class μ_i , as well as the average vector of all images μ.

$$\mu_i = \frac{1}{n_i}\sum_{n\in D_i}^{N} X_k \qquad (1)$$

Where i represents the class and X_k the current element of the class.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad (2)$$

2. Calculate the within-class scatter matrix S_B and the between-class scatter matrix S_W.

$$S_{W=}\sum_{i=1}^{c} \sum_{x\in X_i}(X_k-\mu_i)(X_k-\mu_i)^T \qquad (3)$$

$$S_{B=}\sum_{i=1}^{C} N_i(\mu_i-\mu)(\mu_i-\mu)^T \qquad (4)$$

3. Calculate the eigenvectors and their corresponding values of the dispersion matrices.

$$Av = \lambda v \qquad (5)$$

Where $A = S_W^{-1}S_B$; v is the eigenvector and λ own value.

4. Order the eigenvectors according to the descending order of their respective values and choose K the eigenvectors with the largest values in order to form the matrix W of $d \times k$, where each column of the vector W represents an eigenvector.

5. Use the matrix W to transform the examples (input data) in a new feature subspace. To achieve this, the following operation is performed: $Y = W \times X$, where of X is the d-dimensional matrix that represents the data set (n examples), and Y is the n d-dimensional transformed matrix into a new subspace.

Principal Component Analysis

Principal component analysis was first implemented for facial recognition purposes in [15], [20]; it is an unsupervised algorithm that performs a linear transformation. Said technique has various applications, be it prediction of data in the stock market, analysis of genetic expression data, facial recognition, or others. Being an algorithm, it does not Prevised, the main objective of PCA is to identify patterns in the data, since its type is not known a priori. The purpose of this algorithm is to find a correlation between the variables to separate them and reduce the size of the data. That is, it is possible to state that PCA finds the directions of maximum variance of high-dimensional data and projects them into a new, smaller feature subspace, while preserving most of the information, since the PCA method seeks to obtain the optimal directions (eigenvectors) that capture the largest variance.

Figure 6 shows the 60 directions optimal values or axes of principal components that were generated with the PCA method in the form of images for the set of data. It is impossible to visualize one way graph the first component from eigenvalue 1 to eigenvalue 60, since further addresses or components are generated complicated in the search to maximize the variance in the new feature subspace. The implementation of the algorithm of PCA can be summarized in the 4 steps which are shown below:

1. Obtain the vectors and eigenvalues of the covariance matrix or correlation matrix.
2. Sort the eigenvalues in order descending and choose the k vectors eigenvalues that correspond to the k largest eigenvalues, where k is the number of dimensions of the new entity subspace ($k < d$).
3. Build the projection matrix W of the k selected eigenvectors.
4. Transform the original data set X through W, to obtain a k-dimensional feature subspace Y.

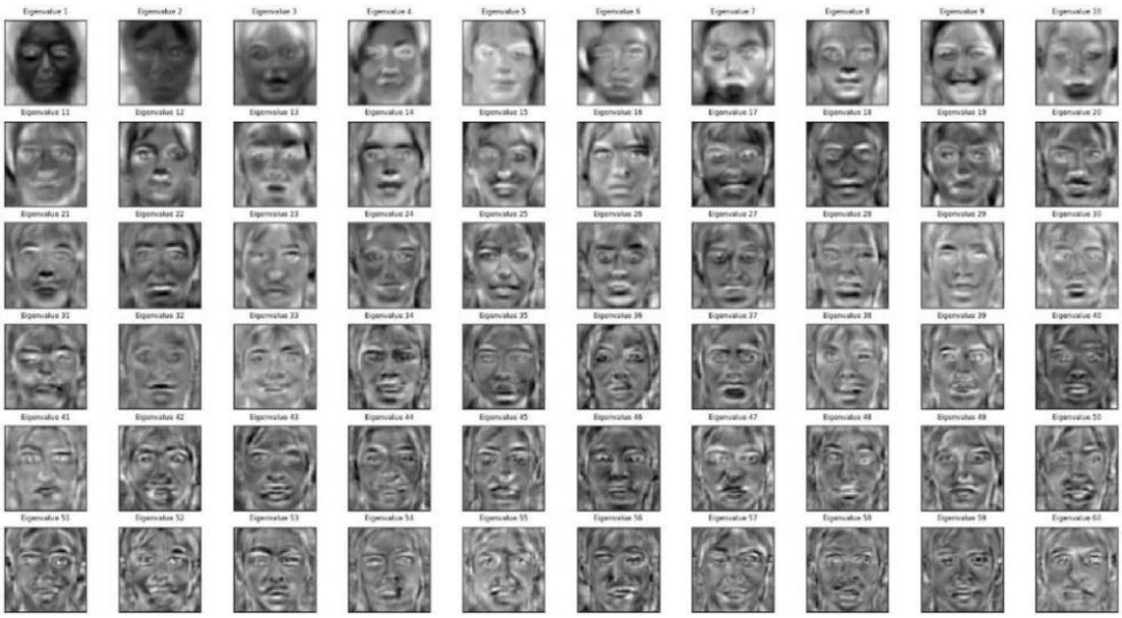

Figure 6. Optimal directions of the principal components that the algorithm returns.

The eigenvectors are the direction of the distortion and the eigenvalues are the scale factor for the eigenvectors that describe the magnitude of the distortion. The eigenvectors are important, since they are the ones that form the axes of the new feature subspace, and the eigenvalues of the magnitude are the new axes. In figure 7, of the 2500 features or columns (pixels), it is clearly shown that most of the variance is 41.39%, which can be explained only by the first principal component, the second principal component contains 11.28% of information, while the third component contains 6.67%. The three components together contain 63.72% of the information.

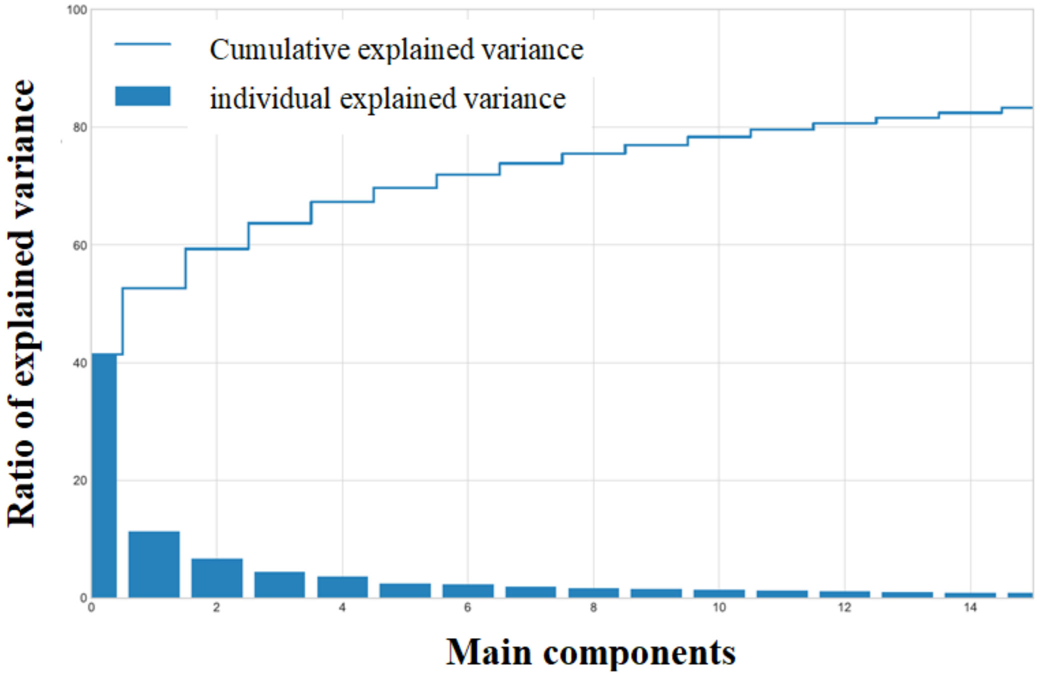

Figure 7. Distribution of the explained variances in the principal components.

RESULTS AND DISCUSSION

Both Linear Discriminant Analysis (LDA) and PCA perform a linear transformation. PCA produces directions (principal components) that maximize variance, while LDA also seeks to find directions that maximize separation (linear discriminants) between different classes. LDA can be more useful in the pattern classification problem, since PCA ignores class labels. In other words, PCA projects the entire data set onto a subspace of different characteristics, whereas LDA attempts to determine a feature subspace to distinguish between patterns belonging to different classes.

In Figures 8 and 9, it is possible to see the first 2 and 3 principal components, respectively, which represent 63.72% of the information, depending on the results to be obtained. Since PCA is known to find the axes with the maximum variance for the entire data set, while LDA tries to find the axes for the best class capability, in practice an LDA followed by a PCA for dimension reduction is often performed.

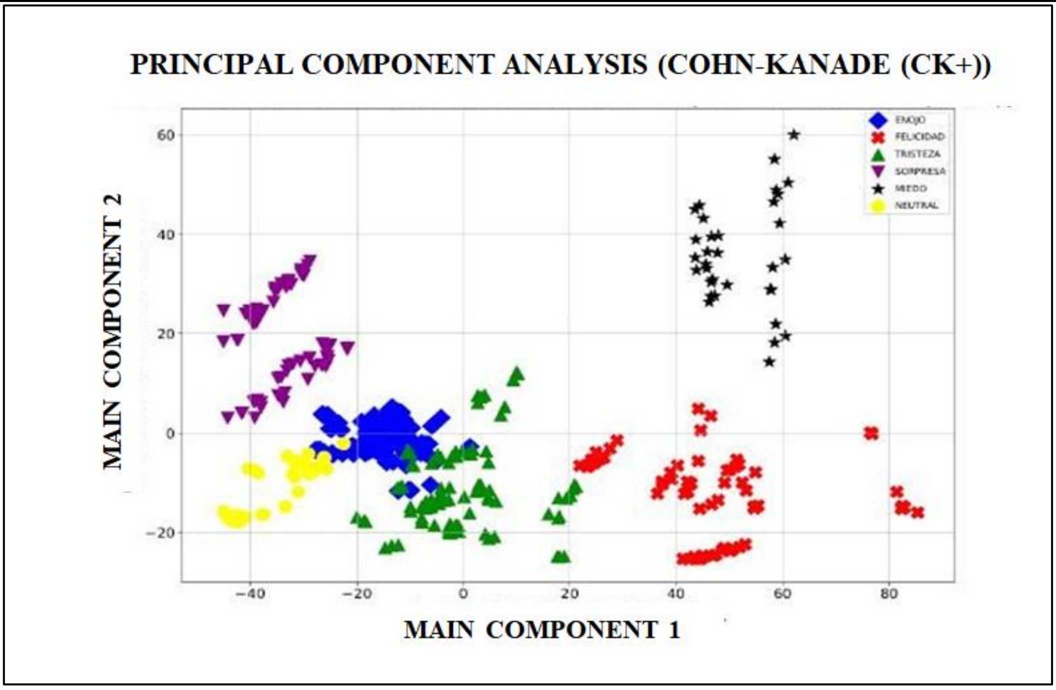

Figure 8. Cohn-Kanade (CK+) projection in a new 2-dimensional space with PCA.

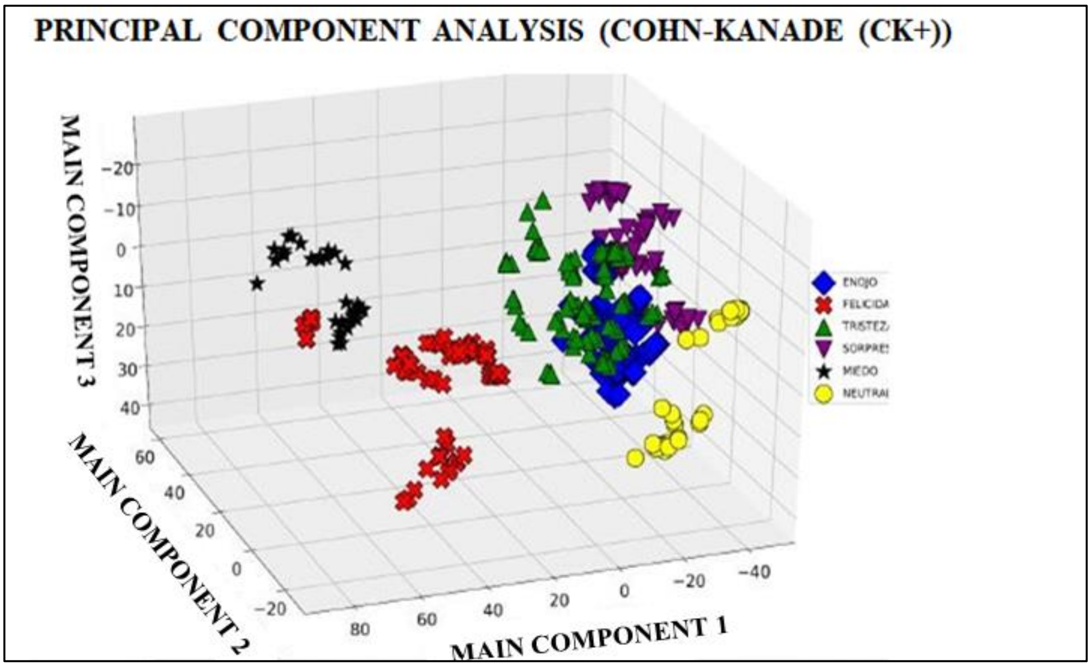

Figure 9. Cohn-Kanade (CK+) projection in a new 3-dimensional space with PCA.

The goal of LDA is to preserve class separation information while continuing to reduce the dimensions of the data set.

In Figure 10, it is possible to see that the data points cluster better when using LDA, compared to the PCA implementation with class labels; this is an inherent advantage to monitoring the method. By contrasting figures 8 and 10, it is clear that PCA represents the largest variation in the entire data set, while the algorithm LDA and its axes explain the greatest variation between individual classes. The implementation of LDA is very similar to that of PCA, so the fit and transform methods are called, which fit the LDA model to the data, and then a transformation is performed using LDA dimension reduction. However, since LDA is a supervised learning algorithm, there is a second argument that the user must supply to the method: the class labels, that is, the emotion labels.

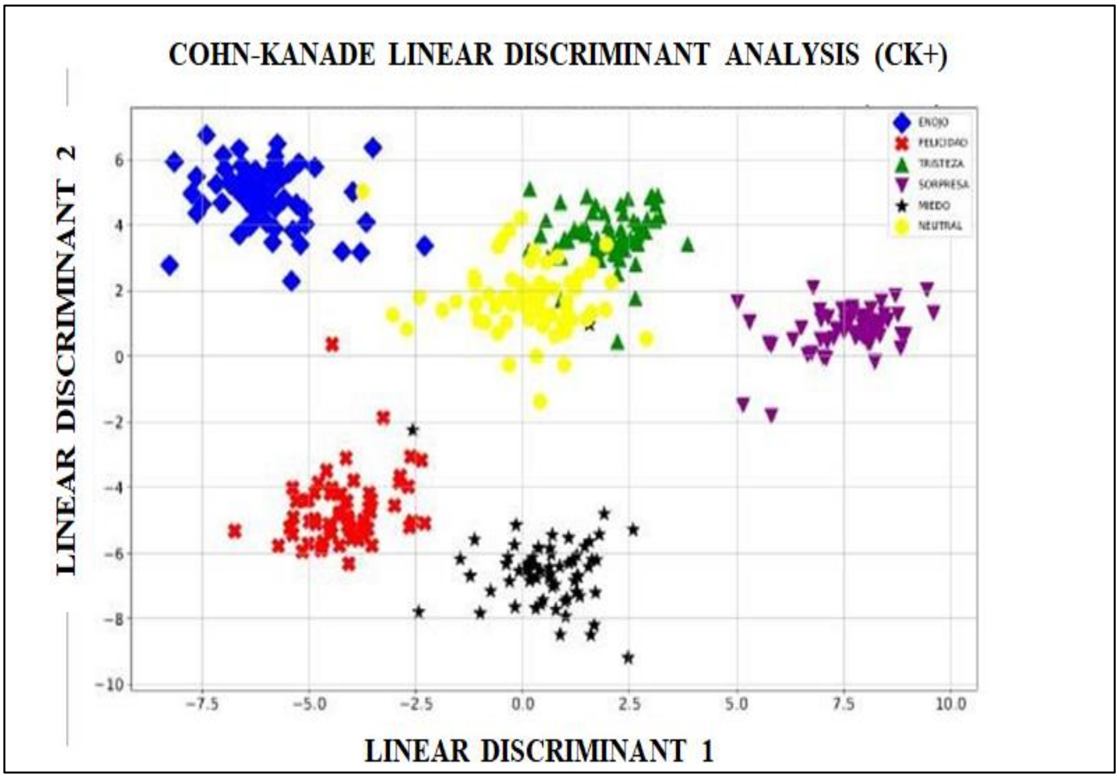

Figure 10. Cohn-Kanade (CK+) projection in a new 2-dimensional space with LDA.

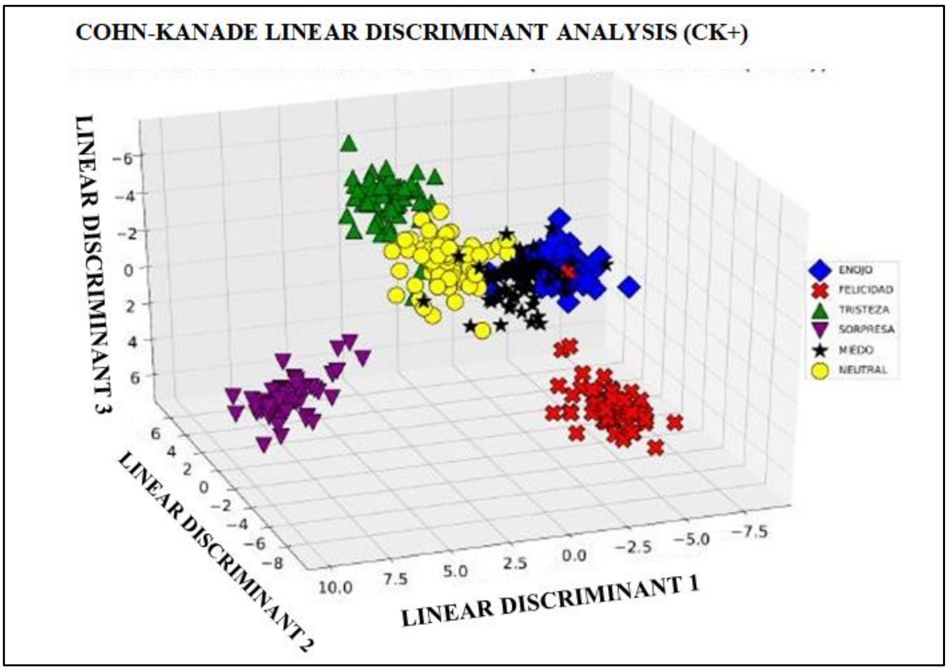

Figure 11. Cohn-Kanade (CK+) projection in a new 3-dimensional space with LDA.

CONCLUSIONS

The work demonstrated that LDA was superior to PCA when it came to reducing and extracting the characteristics of the images from the public Cohn-Kanade (CK+) database, because they know the class labels, although this is not always the case. For example, in comparisons between classification accuracies for image recognition, PCA tends to outperform LDA if the number of samples per class is relatively small [16]. Several experiments in different works show the superiority of PCA over LDA, while others show the opposite.

When PCA outperforms LDA, the number of training samples per class is small, but not atypical of the data sizes previously used by some researchers.

Because in this work only the part of a more complex methodology that seeks to detect and classify facial expressions is reported, the research focuses only on the extraction and reduction of features in images. It is proposed as future work to use this advance as a reference; The next step is to carry out a classification of facial expressions through a support vector machine with several types of kernels, and thus carry out the detection of facial expressions in a more elaborate system.

REFERENCES

[1] S. Li, W. Deng, "Deep facial expression recognition: A survey," arXiv preprint arXiv:180408348, 2018.

[2] I.M. Revina and W.S. Emmanuel: "A survey on human face expression recognition techniques", Journal of King Saud University- Computer and Information Sciences, 2018.

[3] F. Nonis, N. Dagnes, F. Marcolin and E. Vezzetti, "3d approaches and challenges in facial expression recognition algorithms: A literature review", Applied Sciences, vol. 9, no.18, p. 3904, 2009.

[4] Y. O. López, Emotional intelligence. Mexico D. F.: Plaza and Valdes, 2000.

[5] J. H. Shah, M. Sharif, M. Yasmin, and S. L. Fernandes, "Facial expressions classification and false label reduction using lda and threefold svm," Pattern Recognition Letters, 2017.

[6] S. Bellamkonda and N. Gopalan, "A facial expression recognition model using support vector machines" Int. Jour. of Math. Sci. Comp., vol. 4, pp. 56-65, 2018.

[7] M. J. Leo and S. Suchitra, "Svm based expression-invariant 3d face recognition system," Procedia computer science, vol. 143, p. 619–625, 2018.

[8] N.B. Kar, K.S. Babu, A.K. Sangaiah, and
S. Bakshi, "Face expression recognition system based on ripplet transform type ii and least square svm," Multimedia Tools and Applications, vol. 78, no. 4, p. 4789–4812, 2019.

[9] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis chemometrics and intelligent laboratory systems 2," in IEEE Conference on Emerging Technologies & Factory Automation Efta Volume, vol. 2, no. 2, 1987.

[10] R. A. Fisher, "The use of multiple measurements in taxonomic problems," Annals of eugenics, vol. 7, no. 2, p. 179–188, 1936.

[11] S. Raschka, Python machine learning. Packt Publishing Ltd, 2015.

[12] S. Raschka, "Principal component analysis in 3 simple steps," Sebastianraschka. com, (Retrieved O [13] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial

expression analysis," in Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580). IEEE, p. 46–53, 2000.

[14] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," CVPR (1), vol. 1, no. 511-518, p. 3, 2001.

[15] K. Pearson, "LIII. on lines and plans of closest fit to systems of points in space," The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, vol. 2, no. 11, p. 559–572, 1901.

[16] A. C. Kak and A. Martinez, "Pca versus lda," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, p. 228–233, 2001.ctober, 2015).

[17] Nguyen, H. D., Kim, S. H., Lee, G. S., Yang, H. J., Na, I. S., & Kim, S. H. (2019). Facial expression recognition using a temporal ensemble of multi-level convolutional neural networks. IEEE Transactions on Affective Computing, 13(1), 226-237.

[18] Ghimire, D., Jeong, S., Lee, J., & Park, S. H. (2017). Facial expression recognition based on local region specific features and support vector machines. Multimedia Tools and Applications, 76, 7803-7821.

[19] Lienhart, R., & Maydt, J. (2002, September). An extended set of haar-like features for rapid object detection. In Proceedings. international conference on image processing (Vol. 1, pp. I-I). IEEE.

[20] Huang, R., Wei, C., Wang, B., Li, B., Yang, J., Wu, S., ... & Sudakov, V. (2021, November). A comprehensive machine learning approach for quantitatively analyzing development performance and optimization for a heterogeneous carbonate reservoir in Middle East. In SPE Eastern Europe Subsurface Conference. OnePetro.