



Affective Computing: Using Covariance Descriptors for Facial Expression Recognition

Ashaylin Naidoo¹, Jules Raymond Tapamo^{1(✉)}, and Rethabile Khutlang²

¹ University of KwaZulu-Natal, Durban 4000, South Africa
ashaylin.naidoo@gmail.com, tapamoj@ukzn.ac.za

² Council for Scientific and Industrial Research, Pretoria 0184, South Africa
RKhutlang@csir.co.za

Abstract. Facial expression contains a rich variety of affective information. Facial Expression Recognition (FER) is an emerging field that has broad applications in the fields of human-computer interaction, psychological behaviour analysis and image understanding. However, FER presently is not fully realized due to the lack of an effective facial feature descriptor. This paper will examine the performance of a novel facial image descriptor referred to as Local Directional Covariance Matrices (LDCM). The descriptor will consist of fusing features, such as location, intensity, filter responses and incorporate local texture patterns. Tests were done on both posed and spontaneous facial expression datasets to evaluate the performance of the proposed model on real-world applications. Results demonstrate the effectiveness of the covariance feature descriptors compared to standard methods.

Keywords: Local Directional Pattern · Facial expression recognition
Covariance descriptor · Local Binary Pattern

1 Introduction

Affective computing [13] has recently gained momentum in the computer vision community. Facial expression recognition (FER) contributes towards the foundation of affective computing. The difficulty in automatic FER is the discrimination of different facial expressions based features extracted from the face. Facial features from one subject can exhibit similar properties to different expressions; and facial features from two or more subjects with the same expression may vary drastically from each other. Additionally, some expressions like sad and fear tend to be very similar [8]. A simple example is shown in Fig. 1, where six subjects each displaying a happy expression show considerable variation, not only in the way that the subjects convey their expression, but also in lighting, brightness, pose, and background.

Facial features are extracted using two different approaches: geometric-feature-based and appearance-based methods. Geometric feature based [7] methods encode the locations and shapes of unique facial components or elements



Fig. 1. The happy expression among six different subjects. The images are from the following databases: JAFFE [10], CK+ [9], ISED [3]

such as measuring relative positions between eyes, nose, mouth and ears then combining it into a single feature vector describing the face. Appearance based methods differ in that they focus on the individual pixel values rather than relative distance or shape of feature components [14]. They use image filters to get features based on a holistic approach, that is, using the whole face or a region of interest (ROI) of the face to create local features [11]. To strengthen robustness against factors such as occlusion, illumination and pose variations in facial expression recognition many techniques have been researched, such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA) [1]. These methods have been used holistically or locally to extract facial appearance changes [5]. To overcome the shortcomings of LBP newer methods such as Local Ternary Pattern(LTP)[6] and Local Directional Pattern (LDP) were introduced.

Another local feature descriptor that is gaining popularity concerning facial expression recognition is the Region Covariance Matrix (RCM)[12]. The RCM fuses multiple pixel-level image features like coordinate, colour, first-order gradient, etc., into a single more robust covariance matrix structure that becomes the new region descriptor. A benefit to using RCM is that it is scale and illumination independent. In this paper, we integrate RCM and LDP features for facial expression recognition referred to as Local Directional Covariance Matrices (LDCM). This approach aims to investigate the effectiveness of utilizing the LDP with the covariance structure.

The paper is structured as follows. Section 2 briefly reviews the principles of RCM. Section 3 discusses the local patterns encoding and describes the proposed LDCM algorithm in detail. In Sect. 4, we present the experimental results which demonstrate the effectiveness of the proposed method, and finally in Sect. 5 the concluding remarks are drawn.

2 Region Covariance Matrices

Tuzel et al. originally proposed the RCM feature descriptor [12]. Let I be a one-dimensional intensity (grayscale) or three-dimensional colour image (rgb, hsv infrared, depth images) and F be the $W \times H \times d$ dimensional feature image extracted from I , we have

$$F(x, y) = \phi(I, x, y), \quad (1)$$

where the function ϕ can be any mapping such as intensity, colour, gradients, filter responses, etc. For a given rectangular region $R \subset F$, let $z = \{Z_i\}_{i=1..S}$ be the d - dimensional features vector inside R . The region R is represented with the $d \times d$ covariance matrix of feature points

$$C_R = \frac{1}{S-1} \sum_{i=1}^S (z_i - \mu)(z_i - \mu)^T, \quad (2)$$

where μ is the mean of the feature vector z ,

The covariance matrix structure represents the diagonal entries as the variance of each feature, and the non-diagonal entries are their respective correlations. This inherent representation provides multiple advantages to region covariance descriptors. It allows the fusing of different types of features that share some correlation to each other. Its robustness allows matching in different views and poses from a single covariance matrix extracted from a region. Noise from the sample is reduced considerably during the computation of the covariance due to the average filter.

The use of common machine learning methods on a standard covariance matrix is prohibited as it does not lie on Euclidean space [12]. To overcome this problem, Forstner and Moonen [2] proposed a method to calculate the distance between feature points of two covariance matrices C_1 and C_2 , which is defined as

$$\rho(C_1, C_2) = \sqrt{\sum_{i=1}^d \ln^2 \lambda_i(C_1, C_2)} \quad (3)$$

where $\{\lambda(C_1, C_2) \mid i = 1, 2, \dots, d\}$ are the generalized eigenvalues of C_1 and C_2 , computed from

$$\lambda_i C_1 u_i = C_2 u_i, \quad i = 1, 2, \dots, d \quad (4)$$

and $u_i \neq 0$ are the generalized eigenvectors.

3 Local Directional Covariance Matrix

In this section, after a brief review of the Local Directional Pattern (LDP), we introduce the Local Directional Covariance Matrices (LDCM) which incorporate RCM and LDP into a single descriptor.

3.1 Local Directional Pattern

The LDP describes local image feature by computing the edge response values to all its neighbours, i.e. in all 8 directions at each pixel position. It then generates a code from the relative strength magnitude, from [4] it is established that edge responses are more stable than intensity values in the presence of noise and non-monotonic illumination changes. LDP therefore performs superior in these environments as compared to its predecessor LBP.

The LDP is made up of an eight bit binary code assigned to each pixel of an input image. This pattern is encoded using edge response value of a pixel in different directions. There are different edge detectors such as Kirsch, Prewitt and Sobel that can be used for this regard. The Kirsch edge detector is more proficient at detecting directional edge responses because it considers all eight neighbours as compared to the others [4]. Each mask $(M_i)_{i=0,1,\dots,7}$ represents a different orientation. For each mask M_i we compute the response m_i , in total we obtain a response value m_0, m_1, \dots, m_7 , each representing the edge significance in its respective direction. The higher the response value the more significant the edge is in that direction. The Local Directional Pattern code is generated by using the k most prominent directions. The most significant k directional bit responses $|b_i|$ are set to 1 and the remaining bits(8-k) are set to 0. The code LDP_k is then computed as

$$LDP_k = \sum_{i=0}^7 b_i(m_i - m_k) \times 2^i \tag{5}$$

$$b_i(n) = \begin{cases} 1 & n \geq 0 \\ 0 & n < 0 \end{cases} \tag{6}$$

where m_k is the k-th most significant response.

3.2 Local Directional Covariance Matrix

The success of a region covariance matrix as a descriptor relies on the pixel wise features chosen for its specified operation. The LDP and RCM operators are designed to detect textures. Facial expression of a person can be regarded as a texture of the face. Pixel location and intensity are used in the RCM as it improves its discrimination ability. The pixelwise mask of the LDP generated image will also be incorporated into the RCM. Thus, we form a new mapping function based on local directional feature defined as

$$\phi(I, x, y) = [x \quad y \quad I(x, y) \quad LDP(x, y)]^T \tag{7}$$

The feature vector in region R can now be defined as $z_k = \phi(I, x_R, y_R)$, $z_k \in R^d, k = 1, 2, \dots, n$, and the covariance matrix C_R can be obtained by substituting (7) into (2).

The LDCM mapping has a total dimension of $d = 4$ and the resulting covariance matrices are of size 4×4 . This feature descriptor is considerably smaller

than other methods such as LDP or LBP. The advantage of LDCM is that it is more compact than traditional LBP or LDP, the incorporation of the LDP features versus LBP makes it more stable in presence of noise, and the inherent structure of region covariance matrix makes it rotation and scale invariant.

4 Experimental Results

In this section we will review the performance of the proposed algorithm for facial expression recognition on JAFFE [10], Extended Cohn-Kanade [9] and ISED [3] facial expression databases. Firstly, we examine the performance of various covariance based features tested against the whole face region. Then we conduct a test to determine the impact of segmenting the face into regions and lastly we focus on using special landmarks on the face. The method used for classification throughout all tests is a medium KNN classifier using distance described in (3) and leave one out cross validation per expression class. The LDCM uses $k = 3$ for most prominent directions.

4.1 Global Face Covariance Features

In this experiment we use LDCM to analyse the face holistically to determine its effectiveness against different facial expressions in the above mentioned databases. We also incorporate other feature patterns into the covariance matrix like LBP and Sobel mask and compare them to conventional LDP and LBP methods which use histograms as feature vectors.

JAFFE Database. The Japanese Female Facial Expression (JAFFE) database contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models. All experiments carried out on the JAFFE database use an average of 30 images per class tested against an average of 60 random images consisting of 7 classes. The images are cropped automatically to make two eyes align at the same position and are then resized to 160×160 .

Extended Cohn-Kanade Database. The extended Cohn-Kanade Database (CK+) consists of 593 sequences from 123 subjects. The sequences start from a neutral position and ends with the peak of the expression. The database comes with 327 validated emotion labels consisting of six basic (anger, disgust, fear, happy, sadness and surprise) plus contempt expressions. In our analysis, contempt is left out. We choose 25 images per class and test against an average of 75 random images consisting of 6 classes. The images are cropped to make two eyes align and are then resized to 256×256 .

ISED Database. The database consists of 428 segmented video clips of the spontaneous facial expressions of 50 participants. The database consists of labelled peak expressions of 4 classes: happy, sadness, disgust and surprise. The

database features mixed images of people with glasses, non-cohesive pose and other varying uncontrolled environmental factors. The images are cropped by using a facial detector and then resized to 256×256 . An average of 48 images per class were tested against an average of 93 random images consisting of 4 classes.

Table 1. Global face covariance features FER accuracy across datasets

Features	JAFFE								CK+								ISED			
	%Acc.	Neut.	Hap.	Sad	Sur.	Ang.	Dis.	Fear	%Acc.	Ang.	Dis.	Fear	Hap.	Sad	Sur.	%Acc.	Hap.	Sur.	Sad	Dis.
LDCM	90	90	100	90	90	87	79	91	71	76	76	84	56	72	64	97	100	100	98	92
LBCM	89	90	100	90	90	87	76	91	73	76	76	88	60	76	60	96	100	96	96	92
LDP+Sobel+COV	88	90	100	87	90	83	72	91	68	60	76	88	64	68	52	96	100	96	96	94
LBP+Sobel+COV	86	87	100	87	83	83	72	91	71	76	76	88	52	76	56	96	100	94	96	94
Sobel+COV	86	90	100	84	83	83	72	91	70	76	72	84	68	76	44	96	100	98	94	94
LDP	92	100	97	81	87	93	100	84	85	100	68	88	96	76	80	94	92	94	96	94
LBP	93	97	100	87	87	93	93	91	87	100	80	88	96	84	72	91	96	83	96	88

Table 2. Segmented image regions FER accuracy across datasets

Segments	JAFFE								CK+								ISED			
	%Acc.	Neut.	Hap.	Sad	Sur.	Ang.	Dis.	Fear	%Acc.	Ang.	Dis.	Fear	Hap.	Sad	Sur.	%Acc.	Hap.	Sur.	Sad	Dis.
1 x 2	90	90	100	84	87	93	76	97	57	56	72	80	40	56	40	88	90	85	94	83
2 x 1	89	90	94	84	87	83	93	94	53	40	64	64	52	48	52	89	90	92	94	79
2 x 2	96	97	100	100	90	90	97	97	54	36	60	80	44	64	40	88	88	94	94	77
3 x 3	93	100	100	94	87	93	86	91	54	36	64	76	40	60	48	88	90	85	90	85

Table 3. Special landmark regions FER accuracy across datasets

Method	JAFFE								CK+								ISED			
	%Acc.	Neut.	Hap.	Sad	Sur.	Ang.	Dis.	Fear	%Acc.	Ang.	Dis.	Fear	Hap.	Sad	Sur.	%Acc.	Hap.	Sur.	Sad	Dis.
1 - MinDist	95	97	100	90	97	97	90	94	75	84	76	76	68	80	64	94	96	96	92	92
2 - MinSum	91	93	97	97	97	83	69	100	82	92	88	80	80	76	76	95	96	98	92	94

Table 1 shows the results when using the global face experiment. This experiment establishes the effectiveness of the LDCM method compared to the original based LBP and LDP methods. The method LDCM gives good performance accuracy of 90% and 71% using JAFFE and CK+ datasets, respectively. The LDCM outperformed by LBP and LDP on CK+ database. With JAFFE database LDCM is also outperformed, but marginally. The LDCM performs the best with an impressive 97% in the ISED database, the covariance feature based methods performed marginally better than LBP and LDP methods. This can be due to the fact that the ISED database contains more random type images that have partial occlusions and more pronounced pose variations of the face. The covariance descriptor proves to be more robust for these conditions. It is also

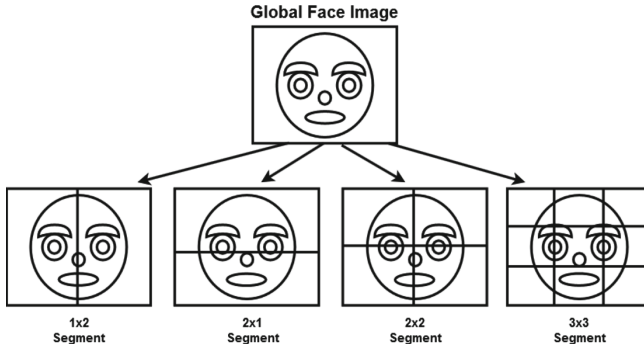


Fig. 2. Segmentation of face into different regions

noteworthy that the LBP and LDP feature vector consists of $[1 \times 16348]$ feature points versus the $[4 \times 4]$ feature descriptor of the covariance matrix. The covariance descriptor is able to produce similar or more effective results at a far lower computational cost in terms of feature size.

4.2 Segmented Face Regions

In this experiment, we test the component-based approach using LDCM. The global face image is segmented into equal sized regions of $[1 \times 2]$, $[2 \times 1]$, $[2 \times 2]$, $[3 \times 3]$. Figure 2 demonstrates a representation of how the face is divided. To classify between segments each region in the test image is compared to its like region in the training images and the region with the minimum distance is chosen for classification.

The results from Table 2 show that the holistic approach performs better than the component-based approach using LDCM in CK+ and ISED databases. This could be due to the fact that when the face is divided into smaller random segments it loses important discriminable information. However, the JAFFE database performed the best using this method compared to the holistic approach receiving a recognition accuracy of 96%. It is also evident that certain regions exhibit greater performance than other regions. Across the three datasets, regions of different segments outperformed. In CK+ the best performing segment was the $[1 \times 2]$ split whereas in the JAFFE dataset it was the $[2 \times 2]$ split and the ISED dataset the $[2 \times 1]$ split. The information from the random segments can be improved upon by targeting specific regions of the face.

4.3 Special Landmark Regions

Intuitively certain regions possess more discriminative properties than others. Based on the later, in this experiment we use a eye, nose and mouth detector to first extract the regions of interest from the face. The pair of eyes including the eyebrows is used, the extracted eye region is expanded to get the eyebrows. The

special regions from the databases are segmented as follows: CK+ dimensions: Eye- 80×160 , Nose- 56×60 , Mouth- 50×90 JAFFE dimensions: Eye- 60×130 , Nose- 40×50 , Mouth- 40×60 ISED dimensions: Eye- 70×200 , Nose- 80×80 , Mouth- 70×120

The LDCM descriptor is then applied to each region. Minimum covariance distance and minimum sum of total regions of covariance matrices are then used to perform the classification.

Table 3 shows that the proposed method achieved the best results on CK+ database. Tracking the performance on CK+ dataset, from the global face we get an accuracy of 71% versus the split segments achieving 57% and finally attaining 82% using special landmark regions. For all datasets a high facial expression recognition accuracy is achieved using special landmark regions and LDCM. JAFFE and ISED scored an average of 94% across both classification methods. The minimum sum classification method achieved a mean of 89% across all datasets. It outperformed the minimum distance classification method for CK+ and ISED databases shown in Table 3.

5 Conclusion

We have proposed a novel local facial expression feature based on LDP codes and region covariance matrices. Results obtained establish that the proposed descriptor achieves high level of performance for FER at a reduced feature size. It was shown that, focusing on special regions of the face such as eyes, nose and mouth promising and stable results were achieved across different datasets and environments. Covariance descriptors are limited with regards to using standard machine learning methods. Transforming covariance structure to accommodate standard machine learning methods is an interesting future research direction.

References

1. Bartlett, M.S., Movellan, J.R., Sejnowski, T.J.: Face recognition by independent component analysis. *IEEE Trans. Neural Netw./Publ. IEEE Neural Netw. Council.* **13**(6), 1450–1464 (2002)
2. Forstner, W., Moonen, B.: A metric for covariance matrices. *Quo vadis geodesia* **66**, 113–128 (1999)
3. Happy, S.L., Patnaik, P., Routray, A., Guha, R.: The indian spontaneous expression database for emotion recognition. *IEEE Trans. Affect. Comput.* **8**(1), 131–142 (2017)
4. Jabid, T., Kabir, M.H., Chae, O.: Local Directional Pattern (LDP); a robust image descriptor for object recognition. In: 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 482–487 (2010)
5. Jabid, T., Kabir, M.H., Chae, O.: Local Directional Pattern (LDP) for face recognition. *Int. J. Innov. Comput. Inf. Control* **8**(4), 2423–2437 (2012)
6. Jaunjale, A.K., Baru, V.B.: Local number pattern: an efficient descriptor for face analysis. *Int. J. Innov. Res. Comput. Commun. Eng.* **6**(4), 11664–11672 (2016)

7. Kotsia, I., Ioannis, P.: Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Trans. Image Process.* **16**(1), 172–187 (2007)
8. Lopes, A.T., de Aguiar, E., De Souza, A.F., Oliveira-Santos, T.: Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. *Pattern Recogn.* **61**, 610–628 (2017)
9. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended cohn-kande dataset (CK+): a complete facial expression dataset for action unit and emotion specied expression. In: *CVPRW*, pp. 94–101, July 2010
10. Lyons, M., Akamatsu, S.: Coding facial expressions with gabor wavelets. In: *Third IEEE Conference on Automatic Face and Gesture Recognition*, pp. 200–205 (1998)
11. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on Local Binary Patterns: a comprehensive study. *Image Vis. Comput.* **27**(6), 803–816 (2009)
12. Tuzel, O., Porikli, F., Meer, P.: Region covariance: a fast descriptor for detection and classification. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3952, pp. 589–600. Springer, Heidelberg (2006). https://doi.org/10.1007/11744047_45
13. Wang, Z., Xie, L., Lu, T.: Research progress of artificial psychology and artificial emotion in china. *CAAI Trans. Intell. Technol.* **1**(4), 355–365 (2016)
14. Tian, Y.: 2004 Conference on computer vision and pattern recognition workshop. In: *2004 Conference on Computer Vision and Pattern Recognition Workshop CVPRW*, pp. 82 (2004)