Gradient directional pattern: a robust feature descriptor for facial expression recognition

F. Ahmed

Presented is a novel local texture pattern, the gradient directional pattern (GDP), and an effective feature descriptor constructed with the GDP codes for facial expression recognition. The GDP operator encodes the texture information of a local region by quantising the gradient directional angles to form a binary pattern. The location and occurrence information of the GDP micro-patterns is then used as the facial feature descriptor. As the gradient operator can effectively enhance the edge information of an image, the resultant GDP features retain more information than grey-level based methods and describe the local image primitives in a more stable manner. Experiments with prototypic expression images from the Cohn-Kanade database shows the superiority of the GDP descriptor against some well-known appearance-based methods.

Introduction: Automated facial expression analysis is an interesting task that has attracted much attention in recent years owing to its potential applicability in various areas, such as human–computer interaction, data-driven animation, and customised applications for consumer products [1]. Deriving an efficient and effective feature representation that can minimise the within-class variations while maximising the between-class variations is the fundamental component for any successful facial expression recognition system [1]. Although much work has been done, there are still open challenges such as recognition in an uncontrolled environment. Therefore, the aim of the ongoing research in automated expression recognition is to increase the robustness of the underlying feature representation against different factors like variations in illumination, pose, alignment, and occlusions [2].

Among the different techniques available for representing human expression, appearance-based methods are widely-investigated owing to their superior performances in uncontrolled environments. These methods generally employ an image filter or filter bank on the whole face or some specific face regions in order to extract changes in facial appearance [1]. Principal component analysis (PCA) and independent component analysis (ICA) are the commonly-used appearance-based methods. Recently, local appearance descriptors based on Gabor wavelets, local binary pattern (LBP) [3], and its variants have focused more on the local region, and have shown better performances. Although LBP features are invariant to monotonic illumination changes, it performs weakly under the presence of random noise and large illumination variation [2]. To address this issue, local ternary pattern (LTP) [2] has been proposed with one additional discrimination level than LBP, which improves the robustness of LTP features under the presence of noise. More recently, Sobel-LBP [4] has been proposed to improve the performance of LBP by applying the Sobel operator to enhance the edge information prior to feature extraction. Local directional pattern (LDP) [1] took a further step by utilising the edge response values instead of grey levels for texture encoding. However, in uniform and smooth regions, LDP still generates unstable codes and is heavily dependent on the number of prominent edge directions.

This Letter introduces a robust appearance-based feature descriptor constructed with the proposed gradient directional pattern (GDP). Unlike the LBP or the LTP operator that exploits grey levels in order to encode the local texture, the proposed GDP operator utilises the more stable gradient direction values instead. The encoding scheme quantises the gradient direction angles in a local neighbourhood with respect to the centre pixel gradient angle and a threshold, which functions as templates for micro-structures such as spot, edge, or corner. Experiments with the Cohn-Kanade (CK) facial expression database [5] demonstrate that the proposed GDP operator is more robust in extracting facial information and provides a higher classification rate compared to some state-of-the-art feature representation techniques.

Local binary pattern: Local binary pattern (LBP) is a grey-scale and rotation invariant texture primitive that describes the spatial structure of the local texture of an image. LBP was originally introduced for texture analysis and later this method has been successfully applied in face recognition and facial expression analysis [3]. The LBP operator selects a local neighbourhood around each pixel of an image, thresholds the P neighbour grey values with respect to the centre pixel and concatenates the result binomially. The resulting binary value is then assigned to the centre pixel. Formally, the LBP operator can be described as:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(i_p - i_c)2^p, \text{ where } s(v) = \begin{cases} 1, & v \geq 0 \\ 0, & v < 0 \end{cases} \tag{1}$$

Here, i_c is the grey value of the centre pixel (x_c, y_c), i_p is the grey value of its neighbours, P is the number of neighbours and R is the radius of the neighbourhood. If any neighbour does not fall exactly on a pixel position, the value of that neighbour is estimated using bilinear interpolation. The histogram of the encoded image block obtained by applying the LBP operator is then used as a texture descriptor for that block.

Gradient directional pattern (GDP): The LBP operator thresholds at exactly the value of the centre pixel. Therefore, LBP codes are sensitive to noise and illumination variation since a little change can cause its value to alter with respect to the centre pixel. Although some recent approaches exploited more stable gradient magnitudes [4] and edge response values [1] instead of grey levels for forming binary patterns, they tend to produce inconsistent codes in uniform and near-uniform facial regions. This Letter presents a robust local texture pattern that is produced by quantising the angles of the gradient direction of the pixels, which retains more information of the local image content. The proposed method introduces a threshold t, which facilitates the generation of consistent texture patterns in smooth facial regions. In this approach, the direction of the gradient vector of each pixel in an image is computed first using the following equation:

$$\alpha(x, y) = \tan^{-1}(G_y/G_x) \tag{2}$$

Here, $\alpha(x,y)$ represents the gradient direction angle of the pixel (x,y), and G_x and G_y are the two elements of the gradient vector which can be obtained by applying the Sobel operator on the source image. The Sobel operator convolves the image with a horizontal and a vertical kernel to obtain the values of G_x and G_y, respectively, as shown in Fig. 1.

–1	–2	–1
0	0	0
1	2	1

a

–1	0	1
–2	0	2
–1	0	1

b

Fig. 1 *Sobel kernels*

a Horizontal kernel
b Vertical kernel

After computing the gradient direction angles, the GDP operator selects a 3 × 3 neighbourhood around each pixel of the image and quantises the neighbour gradient direction angles with respect to the direction angle of the centre pixel using a threshold t. Here, neighbours with a gradient angle $\pm t$ about the centre angle are quantised to 1, and the others are quantised to 0. The resulting binary pattern value is then assigned to the centre pixel.

$$GDP(x_c, y_c) = \sum_{p=0}^{P-1} s(GD_p, GD_c)2^p \tag{3}$$

$$s(GD_p, GD_c) = \begin{cases} 1, & GD_c - t \leq GD_p \leq GD_c + t \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

Here, GD_c is the gradient direction angle of the centre pixel (x_c, y_c), GD_p is the angles of its neighbours, and t is a user-specified threshold. Fig. 2 illustrates the basic GDP encoding method.

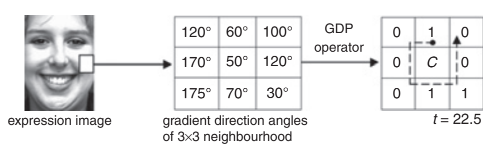

Fig. 2 *Illustration of basic GDP encoding method*

GDP code = 10001100 for C

Facial feature representation using GDP: Applying the GDP operator on all the pixels of an image will result in an encoded GDP image.

The distribution information of the GDP micro-patterns are then represented as a spatial histogram, the GDP histogram, which functions as the facial feature descriptor:

$$H_{GDP}(i) = \sum_{x=1}^{M} \sum_{y=1}^{N} f(GDP(x, y), i), \text{where} f(a, i)$$

$$= \begin{cases} 1, & a = i \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Here, H_{GDP} is the GDP histogram of an $M \times N$ encoded image and i is the GDP code value.

The histogram generated from the whole encoded image contains no location information of the GDP micro-patterns. It merely expresses their occurrence frequencies. However, the presence of location and spatial relationship information provides a better feature representation and describes the image content more accurately [1]. Therefore, in order to incorporate some degree of location information of the micro-patterns, the GDP histogram is modified to an extended histogram. This is done by partitioning the encoded image into a number of regions and generating individual histograms from each of the regions. Finally, the histograms of all the regions are spatially concatenated to obtain the extended GDP histogram. For the expression recognition process, this histogram collection is used as the facial feature vector, as shown in Fig. 3

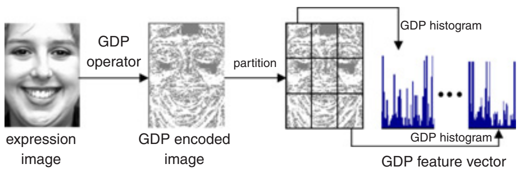

Fig. 3 *Each encoded image is partitioned into a number of regions and GDP histograms of each of the regions are concatenated to form the feature vector*

Experiments and results: The recognition ability of the proposed method is evaluated based on a set of prototypic emotional expressions that includes anger, disgust, fear, joy, sadness, and surprise. This six-class recognition problem can be further extended to a seven-class problem by adding neutral expression images. A well-known image database, namely the Cohn-Kanade facial expression database [5], was used for the experiments. The six-expression dataset comprises 1224 face images and the seven-expression dataset includes additional 408 neutral faces. The selected images were cropped from the original ones and normalised to 150 × 110 pixels. No attempt was made to remove illumination changes as GDP is robust in the presence of illumination variations. The only pre-processing step used was applying a 5 × 5 averaging filter to suppress any noise. A support vector machine with a radial basis function (RBF) kernel was used for classifying feature vectors representing facial expressions. Tenfold cross-validation was carried out to measure the recognition rate.

The performance of the GDP descriptor can be influenced by adjusting the number of regions into which the expression images are to be partitioned. Three different cases were considered in the experiments, where images were divided into 3 × 3, 5 × 5, and 7 × 6 regions. The threshold value t was empirically set to 22.5. The performance of GDP is compared with two state-of-the art facial feature descriptors, namely LBP [3] and LDP [1]. Tables 1 and 2 show the recognition rates of the six-class and the seven-class expression datasets, respectively, using different feature descriptors. From the recognition accuracy, it can be concluded that, facial feature representation based on the gradient directional pattern (GDP) is more robust than some well-known local pattern-based feature descriptors. The superiority of the GDP descriptor lies in thresholding the more stable gradient direction information instead of grey levels for encoding the local texture primitives, which makes it more robust in uncontrolled situations.

Table 1: Recognition rate (%) using the six-class expression dataset

Feature descriptor	Number of regions		
	3 × 3	5 × 5	7 × 6
LBP	79.3	89.7	90.1
LDP	80.2	91.9	93.7
GDP	86.8	95.4	95.9

Table 2: Recognition rate (%) using the seven-class expression dataset

Feature descriptor	Number of regions		
	3 × 3	5 × 5	7 × 6
LBP	73.8	80.9	83.3
LDP	75.7	86.3	88.4
GDP	81.9	90.4	91.6

Conclusion: This Letter presents the gradient directional pattern (GDP) and evaluates its robustness in representing and recognising facial expressions. The novelty of GDP lies in exploiting the gradient direction information for texture encoding, which is more stable and retains more information of the local primitives. Experimental results with prototypic expressions show the superiority of the GDP feature descriptor against some state-of-the-art feature representation methods.

25 May 2012
doi: 10.1049/el.2012.1841
One or more of the Figures in this Letter are available in colour online.
F. Ahmed (*Islamic University of Technology, Gazipur, Bangladesh*)
E-mail: fahmed@iut-dhaka.edu

References

1 Jabeed, T., Kabir, M.H., and Chae, O.: 'Robust facial expression recognition based on local directional pattern', *ETRI J.*, 2010, **32**, pp. 784–794
2 Tan, X., and Triggs, B.: 'Enhanced local texture feature sets for face recognition under difficult lighting conditions'. IEEE Int. Workshop on Analysis and Modeling Faces Gestures, 2007, *LNCS*, **4778**, pp. 168–182
3 Shan, C., Gong, S., and McOwan, P.W.: 'Facial expression recognition based on local binary patterns: a comprehensive study', *Image Vis. Comput.*, 2009, **27**, pp. 803–816
4 Zhao, S., Gao, Y., and Zhang, B.: 'Sobel-LBP'. IEEE Int. Conf. Image Processing., San Diego, CA, USA, 2008, pp. 2144–2147
5 Kanade, T., Cohn, J., and Tian, Y.: 'Comprehensive database for facial expression analysis'. IEEE Int. Conf. Automatic Face Gesture Recognition, 2000, pp. 46–53