Local Force Pattern (LFP): Descriptor for Heterogeneous Face Recognition

Article *in* Pattern Recognition Letters · April 2019

DOI: 10.1016/j.patrec.2019.03.028

CITATIONS
3

READS
229

4 authors:

Shubhobrata Bhattacharya
Indian Institute of Technology Kharagpur
18 PUBLICATIONS **43** CITATIONS
SEE PROFILE

Gowtham Sandeep Nainala
2 PUBLICATIONS **30** CITATIONS
SEE PROFILE

Suparna Rooj
4 PUBLICATIONS **8** CITATIONS
SEE PROFILE

Aurobinda Routray
Indian Institute of Technology Kharagpur
240 PUBLICATIONS **3,519** CITATIONS
SEE PROFILE

Pattern Recognition Letters
journal homepage: www.elsevier.com

Local Force Pattern (LFP): Descriptor for Heterogeneous Face Recognition

Shubhobrata Bhattacharya[a,**], Gowtham Sandeep Nainala[b], Suparna Rooj[a], and Aurobinda Routray[b]

[a]*Advanced Technology Development Center, Indian Institute of Technology Kharagpur, 721306, India*
[b]*Department of Electrical Engineering, Indian Institute of Technology Kharagpur, 721306, India*

ABSTRACT

This paper proposes a novel local-appearance based feature descriptor namely Local Force Pattern (LFP), for robust heterogeneous face recognition. LFP features encode the directional information of the textures of the given face images compactly, producing a more discriminating code than the state of the art methods. We compute the structure of each micro-pattern with the aid of Van der Waal's force between a pair of pixels, which extracts the directional information, and encodes it using a 2-D vector. This directional information helps us to discriminate images having similar structural patterns with different modality based intensity transitions. We divide the face image into several small pixel boxes called 'cell' and extract the LFP features from them. Several cell forms 'block'. The resultant force obtained from the block is vector-summed to get the final feature vector. Then, we concatenate these features into a vector and use it as the face descriptor. Experiments show that the proposed descriptor consistently performs well over state-of-the-art methods for different heterogeneous face datasets like CUSF, IIIT-D and CASIA.

1. Introduction

Face recognition is an eminent domain of research among computer vision community. Face representation plays the most cardinal role in face recognition. The effectiveness of the representation is highly dependent on the efficacy of the descriptor. An efficient face descriptor should ideally satisfy three primary attributes: (1) maximise the margin between inter-class; (2) minimise the variance between intra-class; (3) can be extracted with low computational cost from the original image. Features obtained by descriptors can be grouped into two major types based on their approach: geometric-feature-based and appearance-based methods. The geometric feature of the face is encoded from the shape and locations of different facial components and combined into a feature vector that represents the face. The appearance-based methods use filters, either on the over-all face image, to create holistic features, or some specific region, to create local features, to extract the appearance wise changes in the face. Compared to the geometric feature-based approach, the acceptance of appearance-based methods has grown over the time for their simple representation, quick execution and comparatively better performance. These descriptors achieve satisfactory results in controlled environments but fail to attain competent results in unconstrained situations [1]. Hence in cases of two different modalities like visual & near infrared (NIR) or visual & sketches, face recognition still an open challenge in the arena of face recognition technology [2]. Hence modality independent descriptors are suitable for performing face recognition in heterogeneous domains. In recent years, research carried out on heterogeneous face recognition (HFR) can be broadly classed into three heads - synthesis [3],[4],[5], common subspace representation [6],[7], and modality invariant feature representation [8],[9] matching. Synthesis approaches try to narrow down the gap between VIS and NIR face images by bringing images of one or both the modality to same/ pseudo same domain. Common sub-space methods are meant to project the features from both the modalities into a subspace where intra-person representation points are compact than inter-person points. Both these methods are computationally expensive; requires extensive training data and time-consuming. On the other hand, direct matching approaches mostly rely on using domain invariant features. Due to their simplistic approach, they are more suitable for real-time applications with a minimal number of training samples. We discuss

**Corresponding author:
e-mail: emailshubho@gmail.com (Shubhobrata Bhattacharya)

here a few pioneering works done for HFR.

Tang and Wang [10] were among the pioneers to work on reduction of modality gap of sketch-photo by transforming a sketch into a pseudo photo or a photo into a pseudo sketch and later matching with photo or sketch respectively. Similarly, Liu et al. [11] used Locally Linear Embedding (LLE) to synthesise sketches. Xiao et al. [12] stated an approach based on Embedded Hidden Markov Model (E-HMM) for nonlinear transformation of sketches into pseudo-photo, however this approach is computationally expensive because it is an iterative method. In the field of feature representation, Yuen et al. [13] proposed an approach for representing local facial feature and geometrical relationship between facial features as a global representation of face sketch and photos. In this work [3], Wang et al. proposed a novel face photo-sketch synthesis and recognition method using a multiscale Markov Random Fields (MRF) model. These work synthesise sketch/photo images, by dividing the face region into overlapping patches for learning. The joint photo-sketch model learnt at multiple scales using a multiscale MRF model from a training set contains photo-sketch pairs. The idea of Bhatt et al. [14] was to resolve sketches and photos into multi-resolution Laplacian pyramid and then to each pyramid level, apply a variant of LBP descriptor with the weights of different local regions, obtained using genetic optimisation algorithm. Weights were assigned as per their importance for achieving better accuracy. To identify forensic sketches, Klare et al. [15] present a framework called local feature-based discriminant analysis (LFDA) in which sketches and photos are individually represented using SIFT feature descriptors and multiscale local binary patterns (MLBP). Multiple discriminant projections are then used on partitioned vectors of the feature-based representation for minimum distance matching. Nejati and Sim [16] examined the usage of facial component outlines and distinct facial marks (non-artistic face sketches) for matching to photographs. Zhang et al. [17] introduced theoretic encoding-based descriptor to encode facial structures which ensure similar structures in sketch and photo have the same codes while different structures have different codes. Peng et al. [4] propose a novel graphical representation based method (G-HFR) using Markov networks representation. A coupled representation similarity metric (CRSM) is designed to measure the similarity between obtained representations. The other schemes which were used for face sketch synthesis from photos include sparse representation [18], multi-dictionary sparse representation [6] and local regression models [19].

Yi et al. [20] used a modality invariant feature, canonical correlation analysis-based learning in linear discriminant analysis (LDA) subspace. Multi-scale block local binary with regularised LDA classifier was utilised by Liao et al. [21]. Klare et al. [22] obtained a local histogram of oriented gradients (HOG) and local binary pattern (LBP) features from both VIS and NIR, and matching is done incorporating random subspaces based ensemble of the classifier, nearest neighbour (NN) and sparse representation based matching. Lei et al. [23] represented every face using graph embedding framework, and then NIR and VIS images are projected with two different learned projections on a common discriminative subspace for classification, which is coupled spectral regression. Zhu et al. [2] presented a transductive framework based on learning a subspace such that modality differences are removed while intra-domain features remain preserved. Likewise, Maeng et al. [24] employed HOG features for cross-distance and cross-spectrum face matching.

In recent past descriptors are proposed which are applicable both for the scenario of sketch and NIR face recognition. The work of [25] proposes a generic HFR framework in which both probe and gallery images are represented in terms of nonlinear similarities to a collection of prototype face images. In [26] Lei et al. propose a method to learn a discriminant face descriptor (DFD) in a data-driven way. The descriptor learns the most discriminant local features that minimise the difference of the features between images of the same person and maximise that between images from different people. In [27], Lu et al. propose a compact binary face descriptor (CBFD) feature learning method for HFR. In [8] Roy et al. proposed a novel method called local-gravity-face (LG-face) for illumination-invariant and heterogeneous face recognition (HFR). It employs the concept of gravitational force angle to find the pixel to pixel relation. Recently Roy et al. presented a robust local image representation called local maximum quotient (LMQ) in [9] for capturing modality-invariant facial features. Deep learning methods are quite popular these days for different recognition purposes because of its learning capability of features that describe the inherent shape and appearance of the face. However, the use of deep learning to recognise modality invariant features is relatively new. A few attempts like [28], [29] has been carried out using deep learning, but the abrupt changes in illumination in HFR is challenging to encode using deep learning techniques. Fine-tuning a model requires an adequate number of samples from each class. Moreover, the training of a model with thousands of weights may not be suitable while using a few hundred samples. Therefore, the lack of a sufficient number of training data samples is a hindrance for deep learning based HFR. In this paper, we propose a face descriptor, Local Force (LFP), for robust face recognition that encodes the structural information of the face in a compact form. LFP encodes the structure of a local neighbourhood by analysing its directional information. It computes the edge response in the neighbourhood, in surrounding directions with a relationship equation (inspired from Van der Waals force equation) that relates the central pixel to its neighbours within a prescribed mask of varying size. For all the pixels we choose a vector depending on the variation of intensity of the neighbouring pixels and then, from all the directions, we choose the resultant direction within different sequential segmentation of the image termed as *cell* and *blocks* (Cells comprise of pixels and blocks are made of cells). The resultant direction gives histogram per block which are concatenated to produce the final micro-pattern descriptor for different faces. This approach helps us to encode the intensity changes in facial texture, which can be used as an effective face signature. Moreover, our descriptor uses the information of the entire neighborhood, instead of using selective surrounding pixels for its computation like existing micro-descriptors such as mentioned in [1], [30], [31], etc. Experiments have also been performed

with different variants of the proposed descriptor. We found that LFP is a better performing descriptor in comparison to state of the art. The evaluation of LFP has been detailed in the later section of the paper.

This paper is structured as follows: In Section 2 we discussed the motivating concepts behind the proposed work and its implementation ideas. In Section 3 we have discussed the conventions and assumption made during the implementation of the concept. Section 4 discuss with a brief analysis of the modality invariant reflectance model (IRM) and its relevance with LFP. In Section 5 the proposed LFP method is described in detail. In Section 6 the performance of the proposed descriptor is evaluated and discussed. Finally, we present concluding remarks in Section 7.

2. Van der Waal Forces

For macroscopic bodies with known volumes and numbers of atoms or molecules per unit volume, the total Van der Waals force is often computed based on the "**microscopic theory**" as the sum over all interacting pairs. It is necessary to integrate over the total volume of the object, which makes the calculation dependent on the objects' shapes. For example, [32] approximated the Van der Waals interaction energy between spherical bodies of radii R_1 and R_2 and with smooth surfaces as :

$$U(z;R_1,R_2) = -\frac{A}{6}\Big(\frac{2R_1R_2}{z^2-(R_1+R_2)^2} + \frac{2R_1R_2}{z^2-(R_1-R_2)^2} + \ln\Big[\frac{z^2-(R_1+R_2)^2}{z^2-(R_1-R_2)^2}\Big]\Big) \quad (1)$$

where A is the Hamaker coefficient, which is a constant that depends on the material properties (it can be positive or negative in sign depending on the intervening medium), and z is the center-to-center distance; i.e., the sum of R_1, R_2, and r (the distance between the surfaces). In the limit of close-approach, the spheres (molecules are assumed in spherical forms) are sufficiently large compared to the distance between them; i.e., $r \ll R_1$ or R_2, so that Eq. (1) for the potential energy function simplifies to:

$$U(r;R_1,R_2) = -\frac{AR_1R_2}{(R_1+R_2)6r} \quad (2)$$

The Van der Waals force between two spheres of constant radii (R_1 and R_2 treated as parameters) is then a function of separation since the force on an object is the negative of the derivative of the potential energy function

$$F_{VW}(r) = -\frac{d}{dr}U(r) \quad (3)$$

This yields:

$$F_{VW}(r) = -\frac{AR_1R_2}{(R_1+R_2)6r^2} \quad (4)$$

It is found from the above expression, that the Van der Waals force decreases with decreasing size of bodies (R). The Van der Waals forces become dominant for collections of tiny particles such as very fine-grained dry powders (where there are no capillary forces present) even though the force of attraction is smaller in magnitude than it is for larger particles of the same substance. The simplified version of the force equation acting between a pair of the microscopic body can be given as:

$$F_{VW}(r) = \frac{AR_1R_2}{(R_1+R_2)r^2} \quad (5)$$

We neglect the negative sign as that is decided by the conversion discussed in section 3.

3. Analogy, Convention and Considerations

The grid arrangement of molecules in a crystal has stunning similarity to the sequential arrangement of pixels in a standard image. The bonding force acting among the molecules cause them to hold together resulting in the formation of the crystal structure. This analogy intrigued us to consider the molecule-to-molecule force relation as a representative for the pixel-to-pixel interactive relationship where the volume (a function of radius) of the molecule is analogically similar to the intensity of the pixels. In the due metaphor, we also made certain assumptions to suit the formation of our theory which is as follows:

1. The force of attraction or repulsion is directed between two bodies in the opposite direction to each other. We choose the convention as follows: The force directed from neighbouring pixel to the central pixel is taken 'positive'. To illustrate the scenario we choose an example as shown in Fig.1. Here we have considered the central pixel has a bigger radius and it exercises the attraction force on the neighbouring pixels. In case if the central pixel has intensity lesser than the neighbouring pixel the force direction will be taken 'negative' to satisfy the direction convention. The magnitude, however, remains unaffected by the convention.
2. By the consideration of the Van der Waals force equation as the pixel to pixel relation, we may encounter a situation where the pixel intensity is zero, resulting in zero force. To avoid the null value calculation, we consider a small bias (at a range of 10^{-6}) to all the pixels in the image.
3. The microscopic theory assumes pairwise additivity and neglects the multi-body interactions and retardation [33]. Following the inference of this theory, in the present scenario, we consider that the force of the neighbouring pixel on the central pixel one at a time. We also consider that the central pixel is acted upon only by the assigned neighbourhood (such as (3×3) or (5×5)) and not the other adjacent pixels.
4. Finally, we consider the pixel intensity as (I), the distance between the pixels as delta (δ) and interactive force between a pair of the pixel as (F_{pp}). Fig. (1) shows the inter-active force diagram for a (3×3) pixel window.
5. The constant A of Eq. (5) is Hamakar constant whose value depends on the material property. In other words, the value of A depends on the size of the interacting particles [34]. In case of crystals, all the molecules are of the same shape and hence the value of A remains constant in equation (5). However, in case of a face image, the intensity of pixels are necessarily varying from 0 to 255 (considering the grey scale image). Hence the value of A

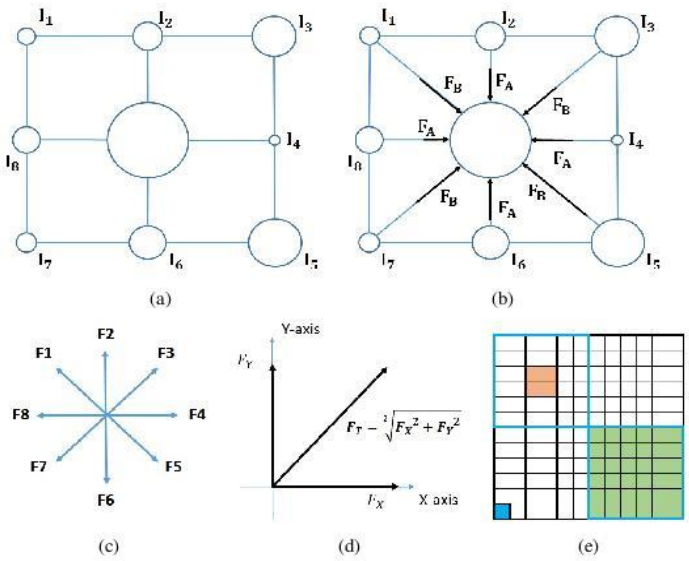


Fig. 1: Fig. (1a) shows a set of pixel where the circle size denotes the variation of intensity. Fig. (1b) show the forces acting towards the center (All forces are positive as per the convention taken in section 3); Fig. (1c): Central pixel to neighboring pixel force relation for a (3×3) window; Fig. (1d): Resultant force of the forces in Fig. (1c); Fig. (1e): The Blue square represents the *Pixel* (smallest component of the image). Four pixels together makes a *Cell*, which is shown in orange. Nine cells makes the formation of *Block*, shown in green shades. Several blocks makes an image together.

in Eq. (11) is taken as a function of pixel intensity for the particular window to match the analogy of considering the molecules collectively for a particular medium.

4. Modality Independent Feature Extraction

The purpose of the proposed method is to recognize the facial features, which are invariant in different modalities. From our visual inspection and after a thorough study of the literature, we conclude that edges are the most important modality-invariant feature. Psychological studies also say that we can recognize a face from its edges only [35]. In literature, there are different methods [36], [37] which used edge-based features for face recognition. Now, it is easy to understand that facial components in a face have maximum edges and they belong to high-frequency components of an image. In this work, we find the modality-insensitive representation of the image using Lambertian model [38]. An image can be represented as

$$I(x, y) = G(x, y).L(x, y) \tag{6}$$

where $I(x,y)$, $G(x,y)$ and $L(x,y)$ are pixel intensity, reflectance and luminance values at position (x, y) respectively. We intensionally choose $G(x, y)$ as reflectance to avoid confusion with radii R_1 and R_2 mentioned previously. $L(x, y)$ is the low frequency component of the image and experience slow variations, even when there are abrupt changes in $G(x, y)$. This smoothing property can be stated as $L(x + \Delta x) \simeq L(x, y)$ and $L(y + \Delta y) \simeq L(x, y)$. On the other hand the effect of the $G(x, y)$ being prominent can be given as

$$i(x + \Delta x, y) \simeq G(x + \Delta x, y).L(x, y) \tag{7}$$

$$i(x, y + \Delta y) \simeq G(x, y + \Delta y).L(x, y) \tag{8}$$

Now it can be concluded that the luminance variation in an image of a particular modality remains more or less consistent throughout the image. Let us consider a window of (3×3) pixels, where the central pixel is I_C, any of the neighboring pixel is I_i (where i ranges from 1 to 8), and the average intensity of the 9 pixel (including the central pixel) is I_{Avg}. Hence by the consideration of the Lambertian theory we get

$$I_C = G_C.L_C \qquad I_i = G_i.L_i \qquad I_{Avg} = G_{Avg}.L_{Avg} \tag{9}$$

By Eq. (7) and Eq. (8) we make the approximation of luminance as:

$$L_C \simeq L_i \simeq L_{Avg} \simeq L_l \tag{10}$$

We take constant as

$$A = \frac{1}{[I_C - I_{Avg}]} = \frac{1}{[G_C.L_C - G_{Avg}.L_{Avg}]} = \frac{1}{[G_C.L_l - G_{Avg}.L_l]} \tag{11}$$

From Eq. (5) and (11) the force expression is derived as:

$$F_{pp} = \frac{1}{(G_C - G_{Avg})L_l} \times \frac{G_C \times G_i}{(G_C + G_i)L_l} \times {L_l}^2 \times \frac{1}{\delta^2} \tag{12}$$

Simplifying Eq. (12) gives the force expression as:

$$F_{pp} = \frac{1}{(G_C - G_{Avg})} \times \frac{G_C \times G_i}{(G_C + G_i)} \times \frac{1}{\delta^2} \tag{13}$$

The force Eq. (13) does not contain the L component. Hence the micro-features, i.e., a force obtained is not affected by luminance and only preserve the changes due to reflectance is G.

5. Proposed Descriptor.

5.1. Force Calculation for each pixel.

The proposed LFP method is a micro-pattern descriptor which captures the pixel-to-pixel relation using the force equation as discussed in section (2). We rewrite the relationship between the pixels as follows:

$$F_{pp} = \frac{1}{(I_C - I_{Avg})} \times \frac{I_C \times I_i}{(I_C + I_i)} \times \frac{1}{\delta^2} \tag{14}$$

where G is replaced by intensity I and δ stands for the pixel to pixel distance (in our case, we chose δ = 1,2,3,4). The force vector acts among the central pixel to the neighbouring pixels in the operating filter size as shown in Fig. (7a). Once the force of individual pixels of a cell (in this case it is (3×3)) is obtained, the resultant force is calculated using force-component relation given as :

$$F_x = (F_4 - F_8) + [(F_3 + F_5)\frac{1}{\sqrt{2}}] - [(F_1 + F_7)\frac{1}{\sqrt{2}}] \tag{15}$$

$$F_y = (F_2 - F_6) + [(F_1 + F_3)\frac{1}{\sqrt{2}}] - [(F_7 + F_5)\frac{1}{\sqrt{2}}] \tag{16}$$

The final force magnitude and angle diagram is shown in fig. (1d).

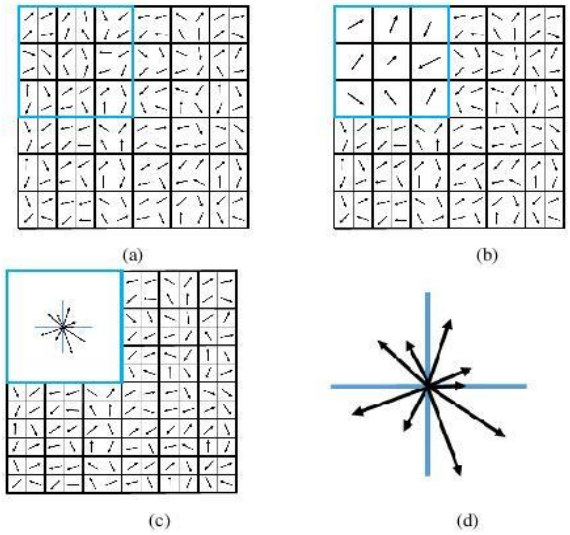


Fig. 2: The distribution of force in a pixel. Fig. (2a) shows the force experienced by each pixel. Fig. (2b) shows the resultant force experienced by each cell (Where *cell* size is (2×2)). Fig. (2c) shows the angular histogram of an individual block (Where Block size is (3×3)).

5.2. Pixel, Cell and Block relation of an Image.

An image is a matrix where each element serves the purpose of a pixel. We declare a set of pixel as *cell* (say 2×2, 4×4 or 8×8). We further declare the collection of cell(s) as *blocks*. Fig. (1e) vividly explains the concept of the cell and block in an image. Within a cell, we compute the force at each pixel as shown in Fig.(2a). The resultant of these individual cells are added vectorially to get the resultant for each cell as shown in Fig. (2b) and finally Fig. (2c) shows the angular histogram of the block. The choice of bins in the histogram has a detailing effect on the feature. In usual practice the bin is taken as nine, i.e., these nine bins represent 360°. In our case, we have also considered the "signed" and "unsigned" conditions for experimental reasons. The left to right and top to bottom move of the operating filter also includes the option of overlapping. In general practice, 50% is taken in the popular descriptor (e.g. Histogram of the gradient) of similar nature. Finally, block-wise histograms are concatenated to obtain the final feature.

5.3. Variants of LFP

Using (3×3) neighbourhood, we can capture extremely fine-grained details in the face image. To increase the capture details at various other scales, we propose variants to our method. We introduce the radius parameter 'δ' which represents the distance between the central pixel of the square window and its neighbouring pixels. These variants not only increase the size of the neighbourhood with an increase in 'δ' but also increase the number of angular force representation in the resulting LFP feature. In this paper, we experimented with four variants by varying the value of 'δ' from 2 to 5, i.e., (5×5) to (11×11).

When several overlapping patches are used, or the angular orientation is taken higher (usually taken as 9), the final feature representation becomes highly redundant, and if the classifier does not have any mechanism for feature selection, it might severely suffer from overfitting. Hence the selection of the descriptor parameters is highly subjected to the computational space complexity. Apart from these variants, the change of parameters like cell size or block size makes significant changes in the performance of LFP. While the smaller cell size increases the capturing details of the micro-features, the increase of the block size helps in better capturing of edge pattern. The feature-length also gets affected due to the variation in cell size, block size and histogram bins. Higher histogram bins capture more subtlety but also results in the introduction of sparse values.

5.4. How is LFP different from state-of-the-art descriptors like LBP, HOG, DoN and BSIF?

Local shape information often well described by the distribution of intensity gradients or edge directions even without precise information about the location of the edges themselves. As a dense feature, HOG [39] has shown great success in object detection and recognition using edge or local shape information. In [39] it is stated that HOG only considers the gradient of individual pixels in horizontal and vertical directions. The final gradient vector obtained is assigned to the concerned pixel as a replacement value. This mechanism is well to do for capturing the edge orientation of the object but fails to capture the more delicate details within the object.

Micro-patterns are on the other hand feature extraction techniques inspired from texture recognition techniques. The most popular micro-pattern descriptor is LBP [1]. Many more variants of this micro-pattern descriptors have been proposed in the literature in recent past. The central theme of these descriptors is based on the fact that the central pixel is related to the surrounding pixel with threshold comparison of value which results in the generation of an array of $'0's$ and $'1's$. This mechanism generalises the pixel-to-pixel relation (there can be multiple situations where the different set of pixels may lead to the same encoding). LFP as a descriptor is fundamentally different from these above-stated descriptors. As explained in section (5.1), LFP incorporates the concepts of both micro-feature descriptor like LBP [1] and orientation descriptor like HOG [39]. Unlike LBP and its variants, in LFP the magnitude force vector is a resultant of the intensity of both the interacting pixels and it avoids levelling it into $'0's$ and $'1's$. The angular gradient of LFP is more sensitive to the changes in pixel intensity, unlike HOG. Its difference has been shown with four illustrative images of various patterns (column-wise: Diagonal edge, Horizontal edge, blob and diagonal line) in Fig. 3. The force vectors of LFP are more distinctively directed from lower intensity pixels to higher intensity pixels.

Recently BSIF [40] has been proposed as a context-aware feature descriptor. The descriptor is designed to extract feature with pre-trained filters for the specific domain. However in the case of HFR, the domains are completely different, and hence the filters trained in respective domain textures are well functioning in corresponding domains only, i.e. homogeneous scenario.

Zheng et al. proposed a descriptor named DoN in [41] which is partly similar to LFP. The descriptor follows the derivation from the Lambertian model [38]. However, the context of DoN and its considerations (explained in Section 4) are different from LFP. In case of DoN, the palm surface is considered

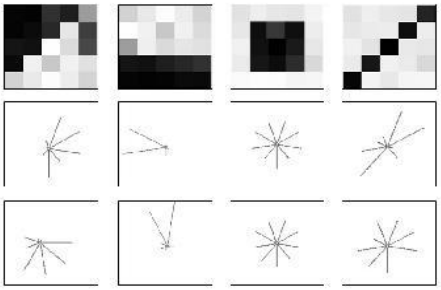
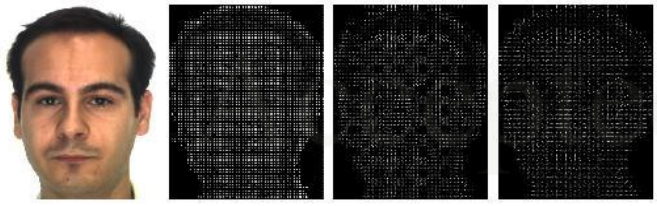


Fig. 3: Angular histogram for distribution of the force vectors for each of the pixels. First row shows the variety of pixel patches. Second row shows the gradient distribution due of HOG [39]. Third row shows the force distribution due to LFP (proposed).

homogeneous, with the same value of L and G throughout the image. However, in the case of heterogeneous face recognition (HFR), the change of G is the prominent signature which is captured by LFP. Moreover, DoN is essentially a 3D descriptor for capturing the pixel-to-pixel relation along the x, y and z-axis. LFP, on the other hand, is a $2D$ descriptor which considers the pixel-to-pixel relation in all directions (including diagonal pixels). The performance of LFP is computationally less expensive than DoN due to its simplistic formulation. The better performance of LFP in comparison to LBP, BSIF and DoN is shown in Fig. 5

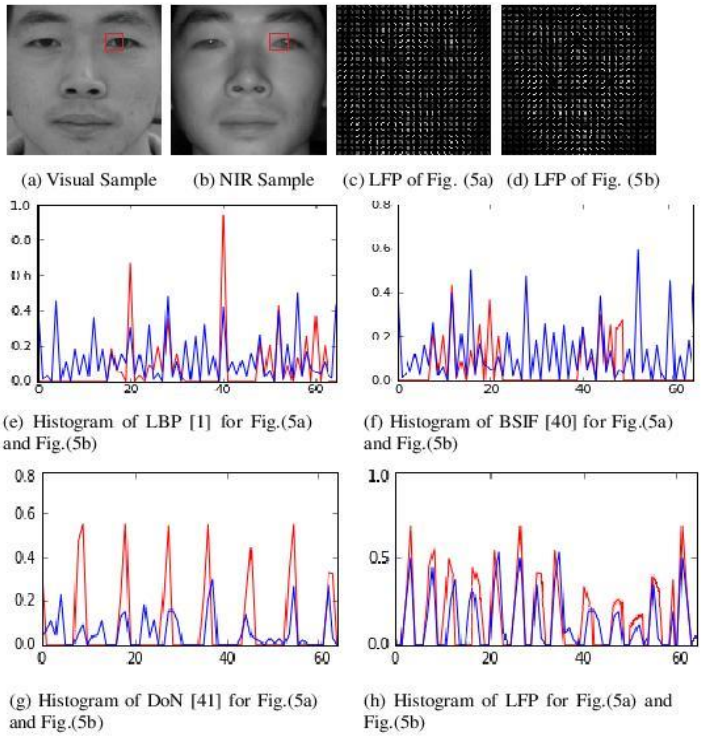


(a) (b) (c) (d)

Fig. 4: Visual samples of application of LFP on Visual image for different configuration of **cell** and **block**. Fig. (4a) : Original Image. Fig. (4b) : Cell 2×2, Block 16×16. Fig. (4c) : Cell 4×4, Block 16×16. Fig. (4d) : Cell 6×6, Block 16×16.

6. Experiment

We performed several experiments to evaluate the performance of the proposed coding scheme for HFR. The face images are geometrically, and several suitable variants of the proposed method are tried on the data-sets with prescribed protocols. We have conducted the experiments using five-fold technique and reported the mean results in respective tables. The classification of features was done using standard chi-square distance [1] owing to its better performance over other standard methods. We cropped and geometrically normalised all images to (128×128) pixels, based on the position of their eyes and mouth when available, or using state-of-the-art face detectors. We experimented to get the best performance for the combination of cell size and block size. The result is shown graphically in Fig. (6b). The cell size of (4×4) and block size of (8×8) are the best possible combination for a fixed radius of the filter. We maintained the same parameter for the experiments in all the cases. We extended the work by applying principal component analysis (PCA) for feature reduction. It increases the performance accuracy as shown in Fig. 7.



(a) Visual Sample (b) NIR Sample (c) LFP of Fig. (5a) (d) LFP of Fig. (5b)

(e) Histogram of LBP [1] for Fig.(5a) and Fig.(5b)

(f) Histogram of BSIF [40] for Fig.(5a) and Fig.(5b)

(g) Histogram of DoN [41] for Fig.(5a) and Fig.(5b)

(h) Histogram of LFP for Fig.(5a) and Fig.(5b)

Fig. 5: Fig. (5e), (5f), (5g) and (5h) compares the histogram of different descriptors for the highlighted region in Fig. (5a) and (5b)

6.1. CUHK Face Sketch (CUFS) Results

The CUFS [3] is a face sketch database built by combining 188 pairs from the Chinese University of Hong Kong (CUHK) student database, 123 pairs from the AR database, and 295 pairs from the XM2VTS database. Altogether forming 606 sketch-photo pairs database for face sketch synthesis and recognition. A comparison made between each sketch image and the corresponding photographic image. The rank 1 accuracy values for each of the state-of-the-art methods (FPSK [3], CITE [17], PLS [7], G-HFR [4], KPRS [25], LG-face [8] and QPLMQ [9]) mentioned above are taken from their corresponding publications. As shown in Table 1.(a) the proposed method is superior to the other methods with a Rank 1 performance accuracy of 98.98%. LFP face does not require any training, whereas some of the other cutting edge methods require numerous training samples.

6.2. IIIT-D (Viewed Sketch) Results

IIIT-D [42] is a testing database of viewed sketches of 238 subjects. For each subject, there is one shape exaggerated sketch drawn by an artist seeing his/her photo. The proposed method along with other state-of-the-art methods (MCWLD [43], LG face [8] and QPLMQ [9]) are tested on this database. As shown in Table 1.(b) the Rank-1 recognition result of proposed LFP on IIIT-D database is 93.29% and outperforms the other methods. Among the three experimented variants of LFP, descriptor with radius 2 performs better than the other variants for cell size $(4, 4)$ and block size $(8, 8)$.

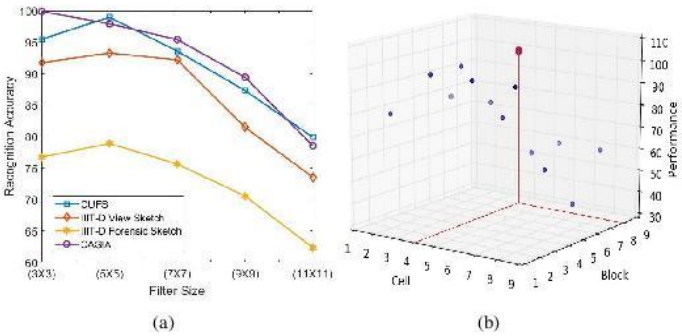


Fig. 6: Optimum parameters for best possible performance accuracy in different heterogeneous face datasets.; (6a) The performance evaluation for the variation of filter size for different heterogeneous face datasets.; (6b) The performance evaluation for the variation of *cell* and *block* size with fixed radius on CUFS dataset.

6.3. IIIT-D (Forensic Sketch) Results

A more challenging database is the IIIT-D Semi-forensic database [42]. This database was made by the forensic artist based on his/her memory after viewing the photo of the subjects and thus carrying less similarity with the photo. The protocol in [43] is followed to test the performance of LFP on this database consisting of 140 semi-forensic sketches. LFP shows 78.91% recognition on IIIT-D Semi-forensic database for Rank-1, which is better than other state-of-the-art methods (G-HFR [4], MCWLD [43], LG-face [8] and QPLMQ [9]). Table. 1.(c) shows the result. Among the three variants of LFP tried in the experiment, descriptor with radius 2 performs better than the other variants for cell size $(4, 4)$ and block size $(8, 8)$.

6.4. CASIA-HFB Results

We used the popular CASIA-HFB database for this experiment. This database contains probe images of 200 subjects captured in the NIR and corresponding gallery images captured at VIS wavelengths. Two NIR images and the two corresponding VIS images were selected from the database, at random for each subject, i.e., a total of 400 NIR and 400 VIS images. The protocol in [28] is followed to test the performance of LFP. The output results are tested against those of five other state-of-the-art methods (CSJDLR [5], KP-RS [25], G-HFR [4], THFM, LG-face [8] and QPLMQ [9]). Table 1.(d) shows the Rank 1 accuracy results for the proposed method and the other state-of-the-art methods. The recognition accuracy of the proposed method is superior to those of the other methods. Moreover, some of the other methods require a large number of training samples, whereas the proposed LFP method does not require any training.

7. Conclusion

In this paper, we presented a novel modality-invariant face descriptor inspired from the principle of Van der Waals force called Local Force pattern (LFP). The descriptor uses the pixel to pixel relation in light of microscopic force to automatically eliminate the luminance (L) component of an image using the normalisation method and simultaneously enhancing the local reflectance (G) component by preserving the facial information. Experiments were done and resulted obtained on the CUFS, IIIT-D, and CASIA database shows that LFP-face offers the best recognition accuracy among all compared methods. We also showed the proper choice of parameters for the cell, block and filter size can lead to the maximum possible accuracy for different datasets. In future, we would like to extend the work by applying in other face representation tasks.

References

[1] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: Application to face recognition, IEEE transactions on pattern analysis and machine intelligence 28 (12) (2006) 2037–2041.

[2] Z. Lei, S. Liao, A. K. Jain, S. Z. Li, Coupled discriminant analysis for heterogeneous face recognition, IEEE Transactions on Information Forensics and Security 7 (6) (2012) 1707–1716.

[3] X. Wang, X. Tang, Face photo-sketch synthesis and recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (11) (2009) 1955–1967.

[4] C. Peng, X. Gao, N. Wang, J. Li, Graphical representation for heterogeneous face recognition, IEEE transactions on pattern analysis and machine intelligence 39 (2) (2017) 301–312.

[5] F. Juefei-Xu, D. K. Pal, M. Savvides, Nir-vis heterogeneous face recognition via cross-spectral joint dictionary learning and reconstruction, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2015, pp. 141–150.

[6] N. Wang, X. Gao, D. Tao, X. Li, Face sketch-photo synthesis under multi-dictionary sparse representation framework, in: Image and Graphics (ICIG), 2011 Sixth International Conference on, IEEE, 2011, pp. 82–87.

[7] A. Sharma, D. W. Jacobs, Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch, in: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, 2011, pp. 593–600.

[8] H. Roy, D. Bhattacharjee, Local-gravity-face (lg-face) for illumination-invariant and heterogeneous face recognition, IEEE Transactions on Information Forensics and Security 11 (7) (2016) 1412–1424.

[9] H. Roy, D. Bhattacharjee, A novel quaternary pattern of local maximum quotient for heterogeneous face recognition, Pattern Recognition Letters.

[10] X. Tang, X. Wang, Face photo recognition using sketch, in: Image Processing. 2002. Proceedings. 2002 International Conference on, Vol. 1, IEEE, 2002, pp. I–I.

[11] Q. Liu, X. Tang, H. Jin, H. Lu, S. Ma, A nonlinear approach for face sketch synthesis and recognition, in: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 1, IEEE, 2005, pp. 1005–1010.

[12] B. Xiao, X. Gao, D. Tao, X. Li, A new approach for face recognition by sketches in photos, Signal Processing 89 (8) (2009) 1576–1588.

[13] P. C. Yuen, C. Man, Human face image searching system using sketches, IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 37 (4) (2007) 493–504.

[14] H. S. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, On matching sketches with digital face images, in: Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, IEEE, 2010, pp. 1–7.

[15] B. Klare, Z. Li, A. K. Jain, Matching forensic sketches to mug shot photos, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (3) (2011) 639–646.

[16] H. Nejati, T. Sim, A study on recognizing non-artistic face sketches, in: Applications of Computer Vision (WACV), 2011 IEEE Workshop on, IEEE, 2011, pp. 240–247.

[17] W. Zhang, X. Wang, X. Tang, Coupled information-theoretic encoding for face photo-sketch recognition, in: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, 2011, pp. 513–520.

[18] L. Chang, M. Zhou, Y. Han, X. Deng, Face sketch synthesis via sparse representation, in: Pattern Recognition (ICPR), 2010 20th International Conference on, IEEE, 2010, pp. 2146–2149.

[19] N. Ji, X. Chai, S. Shan, X. Chen, Local regression model for automatic face sketch generation, in: Image and Graphics (ICIG), 2011 Sixth International Conference on, IEEE, 2011, pp. 412–417.

Table 1: Comparison of Rank 1 recognition accuracy of LFP and other methods in the (a) CUFS (b) IIIT-D Viewed Sketch (c) IIIT-D Semi-forensic Sketch and (d) CASIA Database.

State of the Art	Rank 1
FPSK [3]	96.30
CITE [17]	89.24
PLS [7]	93.60
G-HFR [4]	96.04
KP-RS [25]	80.95
LG-face [8]	98.67
QPLMQ [9]	98.95
$LFP_{(R_1,C_4,B_8)}$	95.48
$LFP_{(R_2,C_4,B_8)}$	**98.98**
$LFP_{(R_3,C_4,B_8)}$	93.59

(a)

State of the Art	Rank 1
MCWLD [43]	84.24
LG-face [8]	89.67
QPLMQ [9]	92.88
$LFP_{(R_1,C_4,B_8)}$	91.75
$LFP_{(R_2,C_4,B_8)}$	**93.29**
$LFP_{(R_3,C_4,B_8)}$	92.19

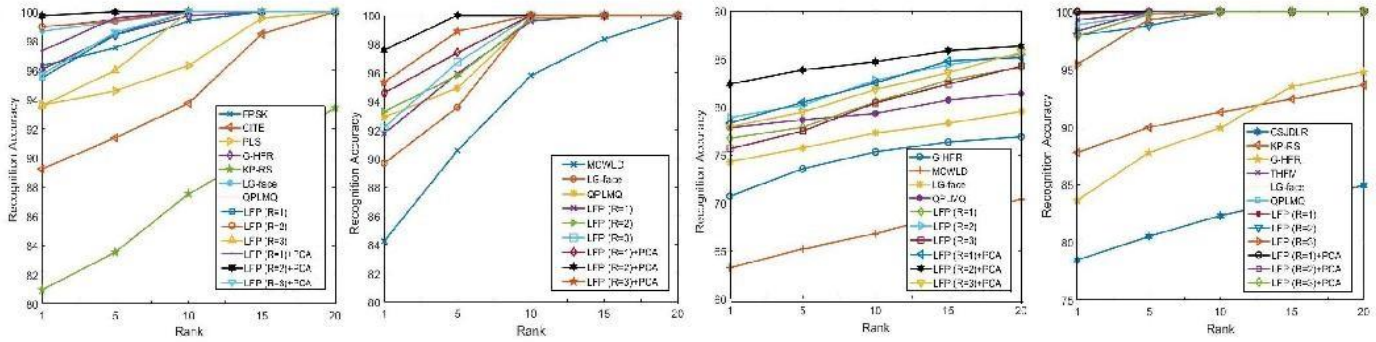
(b)

State of the Art	Rank 1
G-HFR [4]	70.71
MCWLD [43]	63.24
LG-face [8]	74.29
QPLMQ [9]	77.83
$LFP_{(R_1,C_4,B_8)}$	76.73
$LFP_{(R_2,C_4,B_8)}$	**78.91**
$LFP_{(R_3,C_4,B_8)}$	75.64

(c)

State of the Art	Rank 1
CSJDLR [5]	78.46
KP-RS [25]	87.80
G-HFR [4]	83.63
THFM [44]	99.28
LG-face [8]	99.78
QPLMQ [9]	99.84
$LFP_{(R_1,C_4,B_8)}$	**99.88**
$LFP_{(R_2,C_4,B_8)}$	97.94
$LFP_{(R_3,C_4,B_8)}$	95.42

(d)



(a) (b) (c) (d)

Fig. 7: Comparison of Rank 20 recognition accuracy of LFP and other methods in the (a) CUFS (b) IIIT-D Viewed Sketch (c) IIIT-D Semi-forensic Sketch and (d) CASIA Database.

[20] D. Yi, R. Liu, R. Chu, Z. Lei, S. Z. Li, Face matching between near infrared and visible light images, in: International Conference on Biometrics, Springer, 2007, pp. 523–530.

[21] S. Liao, D. Yi, Z. Lei, R. Qin, S. Z. Li, Heterogeneous face recognition from local structures of normalized appearance, in: International Conference on Biometrics, Springer, 2009, pp. 209–218.

[22] B. Klare, A. K. Jain, Heterogeneous face recognition: Matching nir to visible light images, in: Pattern Recognition (ICPR), 2010 20th International Conference on, IEEE, 2010, pp. 1513–1516.

[23] J.-Y. Zhu, W.-S. Zheng, J. Lai, Transductive vis-nir face matching, in: Image Processing (ICIP), 2012 19th IEEE International Conference on, IEEE, 2012, pp. 1437–1440.

[24] H. Maeng, S. Liao, D. Kang, S.-W. Lee, A. K. Jain, Nighttime face recognition at long distance: Cross-distance and cross-spectral matching, in: Asian Conference on Computer Vision, Springer, 2012, pp. 708–721.

[25] B. F. Klare, A. K. Jain, Heterogeneous face recognition using kernel prototype similarities, IEEE transactions on pattern analysis and machine intelligence 35 (6) (2013) 1410–1422.

[26] Z. Lei, M. Pietikäinen, S. Z. Li, Learning discriminant face descriptor, IEEE Transactions on Pattern Analysis and Machine Intelligence 36 (2) (2014) 289–302.

[27] J. Lu, V. E. Liong, X. Zhou, J. Zhou, Learning compact binary face descriptor for face recognition, IEEE transactions on pattern analysis and machine intelligence 37 (10) (2015) 2041–2056.

[28] X. Liu, L. Song, X. Wu, T. Tan, Transferring deep representation for nir-vis heterogeneous face recognition, in: Biometrics (ICB), 2016 International Conference on, IEEE, 2016, pp. 1–8.

[29] Y. Sun, Y. Chen, X. Wang, X. Tang, Deep learning face representation by joint identification-verification, in: Advances in neural information processing systems, 2014, pp. 1988–1996.

[30] B. Zhang, Y. Gao, S. Zhao, J. Liu, Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor, IEEE transactions on image processing 19 (2) (2010) 533–544.

[31] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, Analysis and modeling of faces and gestures (2007) 168–182.

[32] H. Hamaker, The london—van der waals attraction between spherical particles, physica 4 (10) (1937) 1058–1072.

[33] E. Lifshitz, The theory of molecular attractive forces between solids.

[34] C. Argento, R. French, Parametric tip model and force–distance relation for hamaker constant determination from atomic force microscopy, Journal of Applied Physics 80 (11) (1996) 6081–6090.

[35] P. Sinha, B. Balas, Y. Ostrovsky, R. Russell, Face recognition by humans: Nineteen results all computer vision researchers should know about, Proceedings of the IEEE 94 (11) (2006) 1948–1962.

[36] Y. Gao, M. K. Leung, Face recognition using line edge map, IEEE Transactions on Pattern Analysis & Machine Intelligence (6) (2002) 764–779.

[37] W. Chen, Y. Gao, Face recognition using ensemble string matching, IEEE Transactions on Image Processing 22 (12) (2013) 4798–4808.

[38] R. Basri, D. Jacobs, Lambertian reflectance and linear subspaces, in: Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, Vol. 2, IEEE, 2001, pp. 383–390.

[39] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 1, IEEE, 2005, pp. 886–893.

[40] J. Kannala, E. Rahtu, Bsif: Binarized statistical image features, in: Pattern Recognition (ICPR), 2012 21st International Conference on, IEEE, 2012, pp. 1363–1366.

[41] Q. Zheng, A. Kumar, G. Pan, A 3d feature descriptor recovered from a single 2d palmprint image, IEEE Transactions on Pattern Analysis & Machine Intelligence (6) (2016) 1272–1279.

[42] H. S. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, Memetic approach for matching sketches with digital face images, Tech. rep. (2012).

[43] H. S. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, Memetically optimized mcwld for matching sketches with digital face images, IEEE Transactions on Information Forensics and Security 7 (5) (2012) 1522–1535.

[44] J.-Y. Zhu, W.-S. Zheng, J.-H. Lai, S. Z. Li, Matching nir face to vis face using transduction, IEEE Transactions on Information Forensics and Security 9 (3) (2014) 501–514.