Int. J. Mach. Learn. & Cyber. (2017) 8:1457–1469
DOI 10.1007/s13042-016-0516-0

ORIGINAL ARTICLE

Face sketch-photo recognition using local gradient checksum: LGCS

Hiranmoy Roy[1] · **Debotosh Bhattacharjee**[2]

Received: 13 August 2015 / Accepted: 22 February 2016 / Published online: 14 March 2016

Abstract A new approach for matching of face sketch images with face photo images and vice versa has been presented here. For the extraction of local edge features from both the sketch and photo images, a new local feature called local gradient checksum (LGCS) has been developed. LGCS is a modality reduction local edge feature on gradient domain. It is calculated as the summation of four pairs of gradient differences between two local pixels that are at 180° with each other. The Euclidean distance between query sketch and gallery of photos are measured depending on extracted LGCS features. To further improve the result, a multi-scale LGCS is proposed. A rank-1 accuracy of 100 % is achieved in a gallery of 606 photos consisting of CUHK, AR, and XM2VTS face dataset. The proposed face sketch-photo recognition system requires neither learning procedures nor training data. Further, the experiment is extended to test the robustness of the proposed algorithm on blurred, noisy and disguised sketches, as well as photos. Under those situations also, LGCS has outperformed center-symmetric local binary pattern, directional local extrema pattern and weber local descriptor feature extraction techniques.

Keywords Gradient-face · Checksum · LGCS · MLGCS · FSPR-system

✉ Hiranmoy Roy
hiranmoy.roy@rcciit.org

Debotosh Bhattacharjee
debotosh@ieee.org

[1] RCC Institute of Information Technology, Kolkata, West Bengal 700015, India

[2] Jadavpur University, Kolkata, West Bengal 70032, India

1 Introduction

Identification of criminals is generally accomplished either using DNA samples found at the location of crime or from the surveillance cameras, eyewitnesses, etc. Amongst the above resources, the evidences captured by eyewitnesses are usually and readily available in most of the cases.

As a result, for last two decades, human face recognition has become a burning and at the same time very challenging area of research. However, differences in illuminations, postures and angular rotations pose challenges for automated recognition of faces. The challenges are further escalated when the match is required to be executed between two or more heterogeneous face instances such as a face sketch or forensic sketch [1] and a gallery of images of suspected criminals (face photo) or mug-shot photo [1] available with police and administration. This difference is due to the fact that a face sketch is not just a replica of a face photo or mug-shot photo. They may highly contain (major or minor) deviations in the parameters mentioned earlier. A Third type of instance is the viewed sketch [1], which drawn from a photograph of the subject. In the rest of the paper we shall use the terminologies face sketch with viewed sketch and face photo with mug-shot photo interchangeably.

A hand sketch is an artist's impression of a face, based on the description provided by one or more eye witnesses. However, photo images recovered from digital cameras are likely to have difference in modalities with that of a hand sketch. This difference may be referred to as the modality gap between them.

Consequently, the two crucial challenges in matching a hand sketch with a digital photo happen to be:

– Reduction of the modality gap.
– Difference between viewed sketch and forensic sketch.

Although, the way a human brain recognizes a face, is still an unsolved research area, but it has already been established [2] that the human brain tries to focus on "the portions that are very similar" and tend to pose less attention on "the areas that are of high dissimilarity". Human eye mainly focus on edges and the shape of the face. Similarly, the artist emphasizes the edge and shape using different pencil strokes and patches. The missing part in the sketch is flat skin areas.

Therefore, the current work also avoid considering the dissimilar parts i.e. flat skin areas and at the same time put more emphasis on the similar parts i.e. edges and shapes.

Visual inspection of sketch and photo images reveals that, although there are huge differences in modalities, the edges of the vital facial components are very similar i.e. the eyes, nose and mouth etc. The positions of these vital components are highly critical as well. Minor deviations in their positions can generate a totally different face instance and hence a new face. Cognitive psychology [3, 4] revealed that human brain can recognize line drawings as quickly and accurately as gray images. Therefore, to compare a photo and a sketch, a person needs to compare the edges of a photo with that of a sketch. Edges can easily be detected from both photo and sketch using some edge detection method. But, the problem is the modality difference, which will create difference between sketch and photo. So, before attempting to detect the edge, both the sketch and the photo must be converted to one common domain with lesser difference in their modality values. Visual inspection also reveals that the modality difference between sketch and photo is due to its illumination. If somehow the illumination of a face instance can be manipulated by using some illumination-normalization mechanism, then it will be possible to reduce the modality gap between a sketch and a photo easily.

In the current paper Gradient-face [5] is chosen as our modality reduction domain. At the same time, Gradient-face is robust to image noise and image disguise (such as facial expression, mustache, and glasses) and it provides 3D like effect. Followed by that, a new local feature namely the local gradient checksum (LGCS) is introduced to capture the local edge features. Finally, Euclidean distance is measured to get the best matched sketch and photo.

This paper is organized as follows: Sect. 2 presents an overview of related work up to recent times. In Sect. 3, the proposed method is described in detail. Experimental results and comparisons are presented in Sect. 4, and finally the paper concludes at Sect. 5.

2 Related work

In 1996, Lobo et al. [6] introduced a matching technique between police sketches with mug-shot photos. As the police sketches or forensic sketches are not publicly available, most of the research works have been done on viewed-sketch.

Basically, sketch and photo matching techniques are divided into two major categories:

1. Synthesis technique.
2. Modality reduction based feature representation technique.

In the first category, a pseudo photo or pseudo sketch is created to bring one image to another domain and then matched them using some face recognition algorithms. Wang and Tang [7] is the pioneer in this category. After that, many different synthesis based algorithms [8–12] have been proposed. Gao et al. [8] proposed a synthesis mechanism using Hidden Markov Model and selective ensemble (E-HMM). Wang and Tang [9] again proposed a synthesis mechanism using local face patch Multiscale Markov Random Field (M-MRF) model. Followed by, Random Sampling LDA was used for recognition (MS MRF + LDA). Gao et al. [13] proposed a sketch-photo synthesis using sparse neighbor selection (SNS) and sparse-representation-based enhancement (SRE). Wang et al. [14] proposed one transductive face sketch-photo synthesis method (TSPS). Algorithms in this synthesis based category are computationally expensive and may need lots of training, cropping and alignment of the images.

In the second category of modality reduction based features representation, both photo and sketch are transformed into common feature and compared them by different measure. Bhatt et al. [15] proposed a genetic algorithm based extended uniform circular LBP (EUCLBP + GA). Klare et al. [1] combined SIFT and MLBP features (SIFT + MLBP) to get the key points and texture information. A coupled information theoretic encoding (CITE) based feature was suggested by Zhang et al. [16]. Gong et al. [31] combined HOG and MLBP with canonical correlation analysis mechanism (MCCA). Algorithms, in this category, apply some filter operation to reduce the modality and then extract the feature to compare sketch and photo.

Recently Han et al. [17] proposed matching algorithm for composite sketches, which is a new kind of system generated sketches, using facial component wise MLBP feature. Geometric feature based face-sketch recognition is proposed by Pramanik et al. [18]. Rahman et al. [19] proposed a sketch face recognition using fuzzy logic based natural language processing method, called sketching with words (SWW).

Our proposed method belongs to the category of Modality reduction based feature representation. Few significant contributions are done in this paper:

1. Gradient domain is used to reduce the modality difference.
2. A new LGCS local descriptor selection for similarity matching.
3. Effects like blurring, noise, and disguise are added with both sketch and photo for checking the robustness of the algorithm.
4. Query sketch against photo gallery and query photo against sketch gallery are matched, which is the basic property of heterogeneous face recognition.

3 Proposed method

A pictorial diagram describes overview of the proposed method in Fig. 1. Both sketch and photo images are manually aligned according to the eye coordinates, and then cropped and resized. The query sketch image is provided to the system. The system will convert the query sketch into gradient image and from the gradient image to LGCS image step by step. Then the LGCS query sketch will be matched with all the gallery photo database images by converting them to the gradient image and LGCS image respectively. Finally, the best four matched gallery photo database images, according to their ascending order of Euclidean distance, will be displayed. This proposed methodology is called as face sketch-photo recognition (FSPR) system.

3.1 Gradient face generation

For modality deduction we have chosen Gradient-face [5], because it is insensitive to illumination change (proved by Zhang et al.) and on the other hand, it is robust to image noise and image disguise (such as facial expression, mustache, and glasses) and it provides 3D like effect.

Zhang et al. [5] proposed gradient-face in a different way. First the image (I) is convolved with Gaussian kernel function for smoothing (As shown in Eq. 1).

$$I_f = I * G(x, y, \sigma). \tag{1}$$

G is the Gaussian kernel with standard deviation σ. The smooth image I_f is again convolved with derivatives of Gaussian kernel in x and y directions. $G_x(x, y, \sigma)$ and $G_y(x, y, \sigma)$ are the derivatives of Gaussian kernel in x and y direction respectively (As shown in Eqs. 2 and 3).

$$I_x = I_f * G_x(x, y, \sigma) \tag{2}$$

$$I_y = I_f * G_y(x, y, \sigma) \tag{3}$$

$$GF = \arctan\left(\frac{I_y}{I_x}\right) \quad \varepsilon \ [0, 2\pi) \tag{4}$$

Finally, Gradient-face (GF) is calculated using Eq. 4.

The gradient-face is not only able to detect the underlying relationships between adjacent pixels, but also able to capture the shape and facial texture. It is illumination insensitive and at the same time position of key facial features are not changed [5].

3.2 Local edge feature

Although, gradient-face can capture the global feature for matching photo and sketch, a local feature is also needed. Best and well known local feature capturing method is LBP (local binary pattern). The LBP operator was introduced by Ojala et al. [20] for texture classification. It is illumination-invariant but sensitive to noise. A modified version of LBP, called center-symmetric LBP (CS-LBP) [21], is proposed to solve the problem of LBP, which has a long histogram. Murala et al. [22] proposed a directional local extrema pattern (DLEP) at four different directions (0°, 45°, 90°, 135°). Instead of comparing the center-symmetric pairs, they have extracted directional edge information at each and every pixel in a 3 × 3 neighbourhood. The problems with DLEP are its long histogram and sensitivity towards noise.

Another very powerful and robust local descriptor is weber local descriptor (WLD) proposed by Chen et al. in [23]. Wang et al. [24] applied WLD for illumination normalized face recognition. Though WLD is robust to noise and able to reduce the effects of illumination change, but it is sensitive to image blurring effect.

Proposed local edge feature is inspired from CS-LBP. To implement the proposed local edge feature, a 3 × 3 window is considered around a central pixel g_c at a radius (R) one (as shown in the Fig. 2). The differences between the opposite pixels i.e. the pixels those are at 180° with each other with respect to the centre pixel g_c, are considered.

Differences of four different directions (0°, 45°, 90°, and 135°) are measured as:

$$\begin{aligned} D_{f1(0^\circ)} &= |g_1 - g_5| \\ D_{f2(45^\circ)} &= |g_2 - g_6| \\ D_{f3(90^\circ)} &= |g_3 - g_7| \\ D_{f4(135^\circ)} &= |g_4 - g_8| \end{aligned} \tag{5}$$

The local difference edge pattern in four different directions are described using these $D^o_{f1(0)}$, $D^o_{f2(45)}$, $D^o_{f3(90)}$ and $D^o_{f4(135)}$. Out of these, the maximum difference will represent the edge direction.

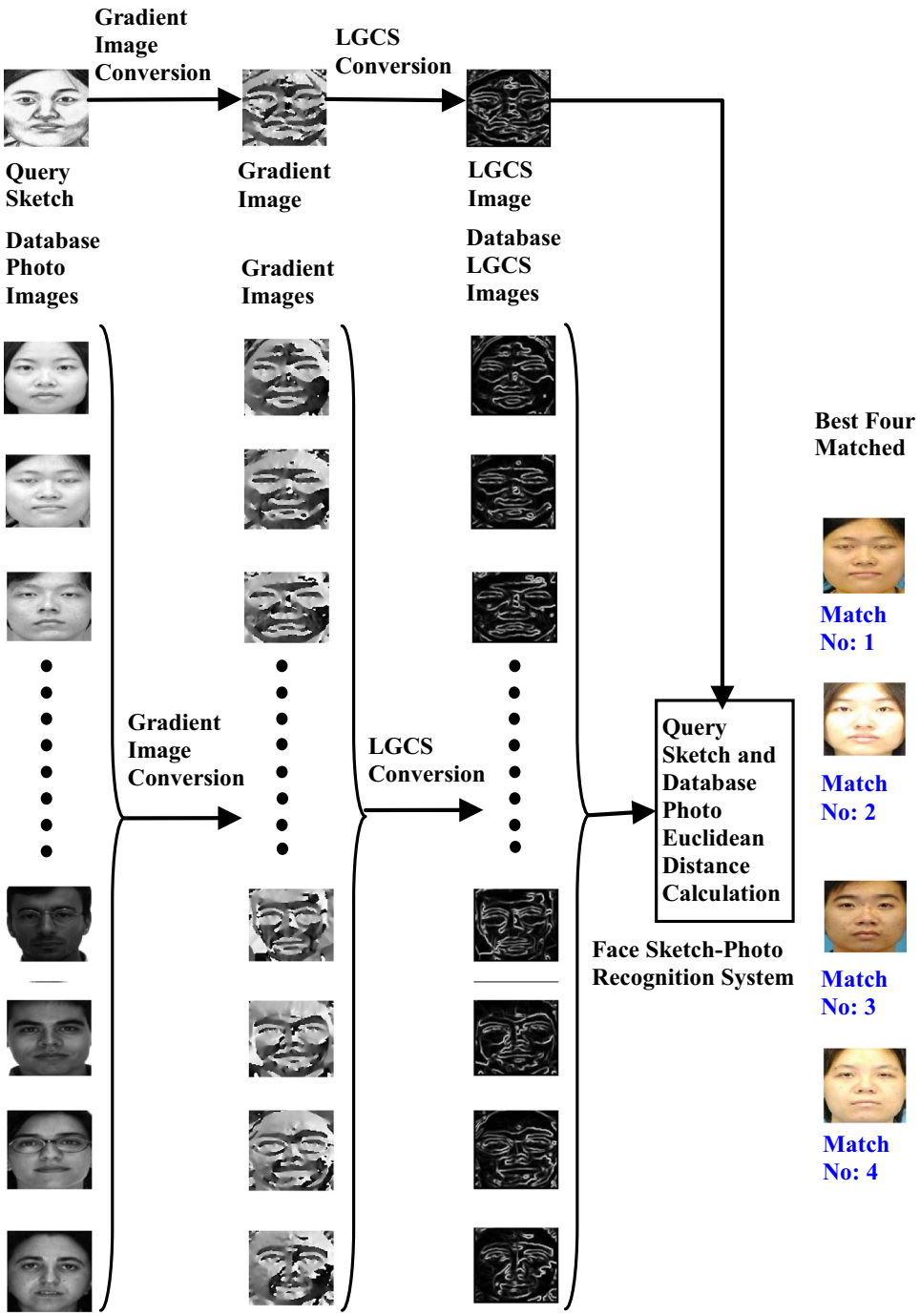

Fig. 1 Overview of the proposed FSPR-system

3.3 Checksum (error detection method)

It is a well known error detecting protocol in data communications [25]. In this mechanism, to send a set of numbers to the destination, the sum of the set of numbers is sent in addition, and the sum is called a checksum. In the receiver end, again the sum of the set of numbers is calculated and matched with the checksum. If they are same then data has no error, so the data set is accepted. Otherwise, it is discarded.

Let S = {7,11,12,0,6} be a set of five integer numbers.
The *Checksum* (*CS*) = (7 + 11 + 12 + 0+6) = 36.
The sender will send: S_d = {7, 11, 12, 0, 6, 36}.

The receiver after receiving, it will again calculate sum of the data excluding the checksum.

R_{dSum} = (7 + 11 + 12 + 0 + 6) = 36.

If difference (D) between *CS* and R_{dSum}: D = $|R_{dSum} - CS|$ is '0' then, receive the data, otherwise discard it.

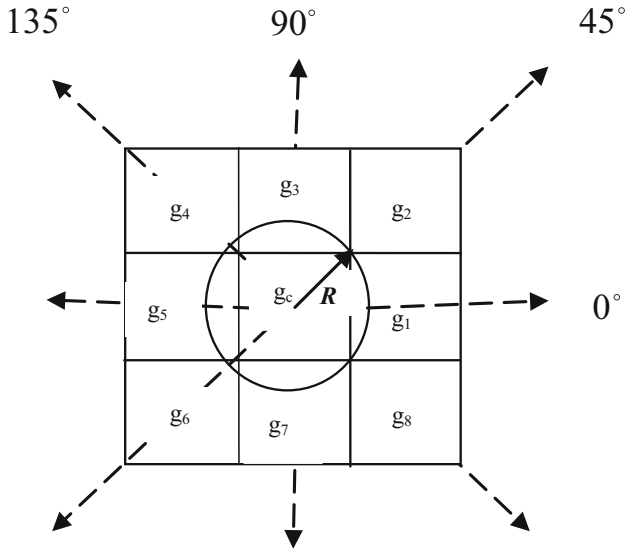

Fig. 2 Local edge direction around the central pixel g_c at a radius (R) one

The checksum is representing the overall information about the set. As 'D' increases, the error or the dissimilarity also increases between the sender and receiver set, on the other hand as 'D' becomes '0', the error decreases and the similarity between the two set increases and become equal.

Four different local edge features are generated to compare sketch and photo faces. Inspired from the concept of checksum error detection mechanism, instead of four difference edge features ($D^o_{f1(0)}$, $D^o_{f2(45)}$, $D^o_{f3(90)}$ and $D^o_{f4(135)}$) their sum (CheckSum) is used to compare face sketch and photo. The local gradient checksum (LGCS) for P (8 pixels) neighbourhood pixels g_1, g_2, g_3, g_4, g_5,g_6, g_7, g_8 (Fig. 2) of the center pixel g_c, is calculated and set at g_c as follows:

$$LGCS_{g_c} = \sum_{i=1}^{P/2} |g_i - g_{(i+P/2)}| \tag{6}$$

In general, at a central pixel, the LGCS can be represented as follows:

$$LGCS_{(P,R)} = \sum_{i=1}^{P/2} |g_i - g_{(i+P/2)}| \tag{7}$$

where, g_i and $g_{(i+P/2)}$ correspond to the intensities of the pairs of pixels which are at 180° to each other on a circle of radius R having P numbers of neighbors.

3.4 Multi-scale LGCS representation

To further improve the matching result, we have used a multi-scale representation of the LGCS feature. Using different radius (R) and numbers of neighbouring pixels (P), we are able to generate different scales LGCS feature. In this paper, (8,1), (16,2) and (24,3) combinations are considered for experimentation. We have performed a score level fusion on the multi-scale LGCS features. Let, for different (P, R) the matching scores are $SM_{(8,1)}$, $SM_{(16,2)}$,$SM_{(24,3)}$. These scores are normalized using min–max normalization and then combined using weighted sum rule-based fusion [26]. The final result (SM_{MLGCS}) is given in Eq. 8.

$$SM_{MLGCS} = w_1 * SM_{(8,1)} + w_2 * SM_{(16,2)} + w_3 * SM_{(24,3)} \tag{8}$$

where, w_1, w_2, w_3 are the weights and $w_1 + w_2 + w_3 = 1$. The different contributions of different scale are done heuristically by computing the recognition rate (r) for each scale separately. Then, we set each weight $w_i = \frac{r_i}{\sum_{i=1}^{3} r_i}$ where r_i denotes recognition rate at i-th scale and $\sum_{i=1}^{3} r_i$ denotes the total recognition rate in three different R-values. The ckecksum mechanism is summing multiple values and producing one single value. Therefore, we can say that checksum mechanism is reducing the number of features. In case of (8,1), the four different edge differences surrounded in each pixel are summed up to one LGCS feature, similarly in case of (16,2) and (24,3) the number of different edge differences are converted to one LGCS feature, respectively.

Though proposed LGCS is inspired from CS-LBP, but it is more advantageous than CS-LBP. The main advantages of LGCS over CS-LBP are:

1. In CS-LBP, the binary pattern is obtained by thresholding the gray value differences according to their sign, the magnitude part is not considered, whereas proposed LGCS not only retain the difference values but also represent the total checksum to measure the edge feature.
2. The proposed LGCS is robust to noise; because a Gaussian smoothing is applied in GF creation (Eq. 1) and again a summation of local edge differences are measured (similar to local smoothing) for final LGCS generation (Eq. 7). In both the two steps, we are using local smoothing and it makes LGCS more robust to noise. On the other hand CS-LBP is sensitive to noise.
3. LGCS does not need any histogram calculation, whereas in CS-LBP a histogram is must and which is time consuming.

Total energy is used to measure the information present in the image. We have chosen entropy for this energy calculation [27]. Experimental results show that, the proposed edge feature (LGCS) has higher entropy value than CS-LBP, DLEP and WLD, which means that the proposed edge feature, captures more information for face recognition than CS-LBP, DLEP and WLD. Table 1 and Fig. 4

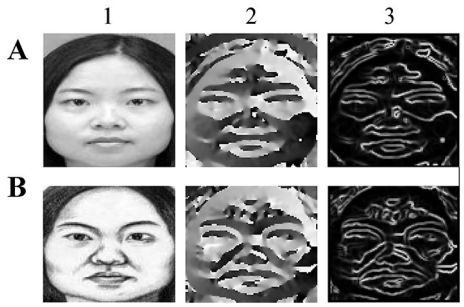

Fig. 3 (**A**, **1**)-Photo image, (**A**, **2**)-gradient-face image of the corresponding photo, (**A**, **3**)-LGCS Image of the corresponding photo, (**B**, **1**)-Sketch Image, (**B**, **2**)-Gradient-face image of the corresponding sketch, (**B**, **3**)-LGCS Image of the corresponding sketch

show experimental results, where LGCS outperforms CS-LBP, DLEP and WLD.

Figure 3 shows the conversion result of photo and sketch into its corresponding gradient-face and LGCS image. Here, A and B represents rows, and 1, 2, 3 represents columns. (A,1) stands for A-th row and 1st column.

3.5 Similarity measure

For a query image of dimension: ($rw \times cl$), where rw is the row-size and cl is the column-size of the image, the feature vector has a size $(L) = rw \times cl$. Using different (P, R) for LGCS we are able to generate different scale LGCS features. We have considered (8,1), (16,2) and (24,3) to generate the multi-scale features. Therefore, the final feature vector size becomes $(L) = 3 \times rw \times cl$. LGCS is a local feature and we are calculating the checksum. Therefore, Euclidean distance between the corresponding points of query image and gallery image will give the difference (similarity or dissimilarity) between the points. The more is the difference between them means more dissimilar they are (As explained in Sect. 3.3). Let the feature vector for the query sketch image QSK is represented as $f_{QSK} = (f_{Q_1}, f_{Q_2}, f_{Q_3}, \ldots, f_{Q_L})$. Similarly, each image in the photo gallery database is represented with the feature vector $f_{DB_j} = (f_{DB_{j1}}, f_{DB_{j2}}, f_{DB_{j3}}, \ldots, f_{DB_{jL}})$, $j = 1, 2, 3, \ldots, |DB|$. The goal is to choose the n best images that resemble the query sketch image. In order to match the pictures, we measure simple Euclidean distance between the query image and images in photo gallery database having length |DB|.

$$E_d(QSK, DB) = \sqrt{\sum_{i=1}^{L} (f_{DB_{ji}} - f_{Q_i})^2} \tag{7}$$

here, $f_{DB_{ji}}$ denotes i-th feature of the j-th image of the photo gallery database $|DB|$.

Algorithm: **The FSPR-LGCS Algorithm.**

Input: Gallery of cropped face photo images of C_N classes; query image I_{QSK}. 'N' is the size of the gallery. Each image has a dimension $rw \times cl$.

Output: Top 'M' ranked matched gallery images C_M. 'M' is the size of the rank.

1: ***Feature space Conversion***:
Convert all gallery photo images into gradient domain images and those gradient images into LGCS's images.
Build the Database:

$DB = (LGCS_1, LGCS_2, \ldots\ldots\ldots\ldots, LGCS_N) \in R^{rw \times cl}$

2: ***Matching***:
Convert I_{CSKQ} image into gradient domain image and then into LGCS image (I_{LGCS}).
3: **For** i=1 To N do
Calculate Euclidean distance (E_i) between I_{LGCS} and $LGCS_i$ image.
4: **end**
5: Sort the C_N class images in ascending order of their E_i ($E_1 < E_2 < \ldots\ldots\ldots < E_N$).
6: Print the top 'M' ranked images (C_M).

4 Results and discussions

In the experimentation, a face photo-sketch database of 606 pairs has been built for the experimental study, which includes 188 faces from the CUHK student data set [28],

Table 1 Comparison results of the proposed LGCS against CS-LBP, DLEP and WLD

Query images ↓	Methods			
	CS-LBP	DLEP	WLD	$LGCS_{(8,1)}$
Rank-1 accuracy (%) for query sketches against photo gallery of size: 606				
Blur sketches of size: 100	79	91	86	97
Gaussian noise added sketches of size: 100	67	81	88	98
Salt & pepper noise added sketches of size: 100	69	89	88	98
Normal test sketches of size: 606	88.42	95.71	91.31	98.67
Average entropy of 606 sketches after conversion to corresponding feature domain (10^5)	1.984	2.124	2.367	3.234

123 faces from the AR data set [29], and 295 pairs from XM2VTS data set [30]. For each of the photo image, there is one hand drawn sketch present.

For testing more accuracy and robustness of the proposed algorithm, some new modified, manually created face photo images are included in the face photo database after adding effects (noise, blur and disguise). As face photo galleries are matched with a query face sketch image, sketches are also added with same effects. Thus, the face photo database as well as sketches are enriched with modified images. Finally, photos are matched against gallery of sketches for testing the performance of modality reduction based property of LGCS. For MLGCS, the weight values are $w_1 = 0.55$, $w_2 = 0.15$, $w_3 = 0.30$.

4.1 Comparison of LGCS with CS-LBP, DLEP and WLD

At first, we have tested the superiority of proposed feature against CS-LBP, DLEP and WLD in recognizing sketches from gallery of photos. Table 1 shows the comparison result of the proposed feature with CS-LBP, DLEP and WLD for recognizing sketches in different noisy conditions and also representing total information present (using average entropy calculation of the database). The result is tested on combined CUHK student data set, AR data set and XM2VTS dataset.

Figure 4 shows how the effects of noise and blur degrade the quality of the CS-LBP and WLD images. Whereas proposed method is still able to represent the facial structures and edges. Here, A,B,…,D represent rows and 1,2,..5 represent columns. (A,1) stands for A-th row and 1st column. Column-1 represents the original photo images and column-2 represents (top to bottom) the gray image, Gaussian noise added image, salt and pepper noise added image and blur added image respectively. Column-3 represents the corresponding output of the Column-2 images after performing CS-LBP operation. Column-4 represents the corresponding output of the Column-2 images after performing WLD operation. Column-5 represents the corresponding output of the Column-2 images after performing proposed LGCS operation. From visual inspection of the output results, we can say that in case of noise CS-LBP and WLD transformed images are almost imperceptible and look like chaos. On the other hand, LGCS transformed images are still perceptible and almost capturing the shape and important edge information.

Gradient-face (GF) is used to reduce the modality gap between sketch and photo. Therefore, we can consider that GF is a kind of pre-processing stage. To show that the proposed LGCS feature is superior to other local feature descriptors even after, they are pre-processed by GF, the following test is done. Other local descriptors like CS-LBP, DLEP and WLD are pre-processed by applying Gradient-face (GF). Table 2 shows that proposed MLGCS outperforms other methods.

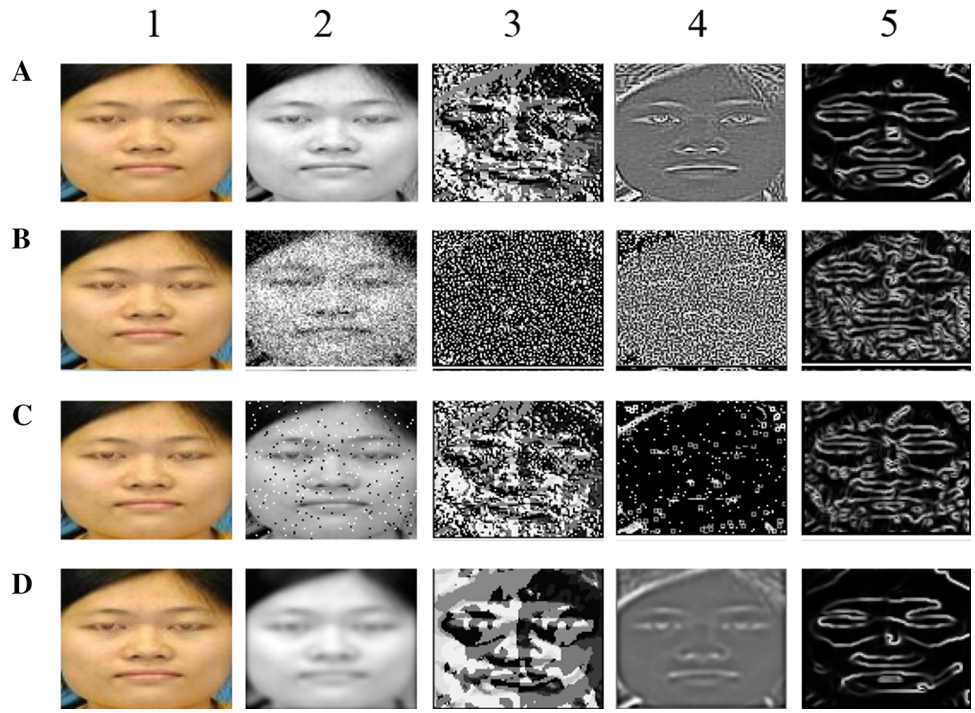

Fig. 4 *Column-1* Original images, (**A**, **2**) simple *gray image*, (**B**, **2**) image corrupted by Gaussian noise ($\mu = 0.01$, $\sigma = 0.03$), (**C**, **2**) Image corrupted by salt & pepper noise (density = 0.04), (**D**, **2**) image corrupted by Gaussian blur (15×15 window, $\sigma = 1.5$), *Column-3* CS-LBP transformed images, *Column-4* WLD transformed images, *Column-5* LGCS transformed images

Table 2 Comparison result of the proposed MLGCS against CS-LBP, DLEP and WLD descriptors, where CS-LBP, DLEP and WLD are pre-processed with Gradient-face (GF)

Query images ↓	Methods				
	GF	GF + CS-LBP	GF + DLEP	GF + WLD	MLGCS
Rank-1 accuracy (%) for query sketches against photo gallery of size: 606					
Blur sketches of size: 100	60	93	97	95	98
Gaussian noise added sketches of size: 100	71	82	91	94	100
Salt & pepper noise added sketches of size: 100	68	89	94	97	100
Normal test sketches of size: 606	68.32	94.42	97.71	96.31	100

Table 3 Comparison of rank-1 recognition accuracy of the methods mentioned above in CUHK dataset (combination of CUHK student, AR dataset, and XM2VTS dataset)

Methods	Training samples	Testing samples	Rank-1accuracy (%)
MS MRF + LDA	306	300	96.3
E-HMM	306	300	95.24
SNS-SRE	306	300	96.5
TSPS	306	300	97.7
EUCLBP + GA	78	233	94.12
CITE	306	300	99.87
SIFT + MLBP	306	300	98.47
MCCA	306	300	99.67
GF	0	606	68.32
Proposed $LGCS_{(8,1)}$	0	606	98.67
Proposed MLGCS	0	606	100

4.2 Query sketch against photo gallery

To evaluate the performance of the proposed system, the query face sketches are tested against the gallery of face photos. Subsequently, we have tested the robustness of the proposed FSPR system on application of different external effects like noise, blurring and disguise. Also, we compare the results of the proposed approach with different face sketch-photo matching algorithms. These include:

- Sketch or photo synthesis algorithms:
  1. MS MRF + LDA [10]
  2. E-HMM [9]
  3. SNS-SRE [13]
  4. TSPS [14]
- Modality reduction based feature algorithms:
  1. EUCLBP + GA [15]
  2. CITE [16]
  3. SIFT + MLBP [1]
  4. MCCA [31]

All the face sketch database and face photo gallery database are converted to gray tone image and then manually aligned according to eye coordinates, and then cropped and resized to 110 × 110 dimension.

Cropped and resized face photo images are stored in the gallery of face photo database and face sketch images are stored in query face sketch database. Then selecting each face sketch as query image accuracy of the FSPR-system is tested.

$$\text{Accuracy rate} = \frac{\text{Number of correct matches at rank}-1}{\text{Total number of test images}} \times 100\%$$

The Rank-1 accuracy result of all these algorithms mentioned above is found from their respective papers. Our proposed method is better than other methods. Most significantly, our proposed method does not require any training. We have tested the accuracy by setting each and every sketch as one query image, matched against the gallery of 606 photos. $LGCS_{(8,1)}$ gives a rank-1 accuracy of 98.67 % means 598 sketches are successfully recognized out of 606. MLGCS gives 100 % accuracy result at rank-1.

Our proposed feature belongs to modality reduction based feature representation category. Therefore, we have tested the robustness of the proposed LGCS feature with other modality reduction feature representation methods. (EUCLBP + GA, CITE, SIFT + MLBP, and MCCA). These four algorithms have been implemented according to their reference papers and trained according to the

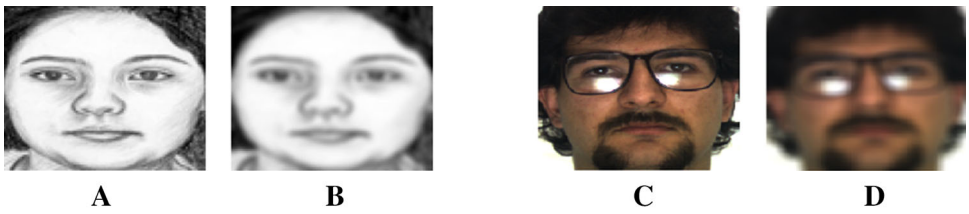

Fig. 5 **A** Sketch image, **B** blur effect (15 × 15 windows & $\sigma = 1.5$) added to the image (**A**), **C** photo image, **D** blur effect (15 × 15 windows & $\sigma = 1.5$) added image of image (**C**)

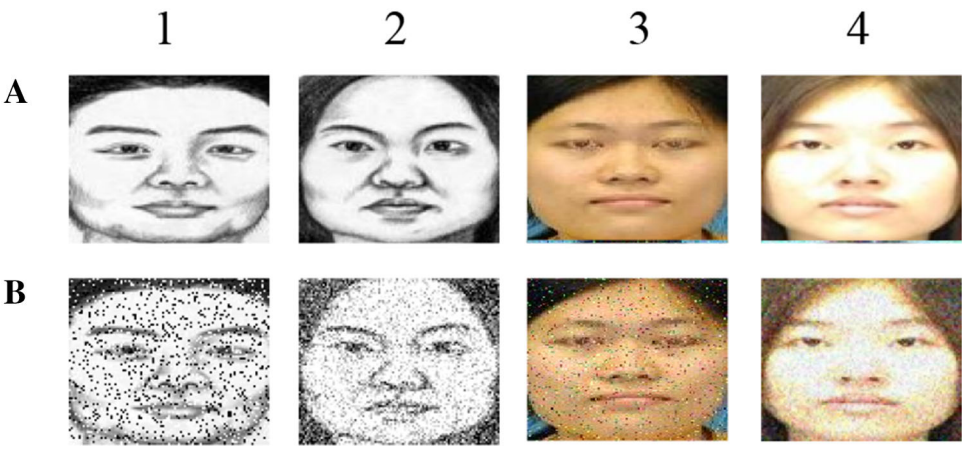

Fig. 6 (**A**, **1**) Sketch image, (**B**, **1**)-salt & pepper noise (density = 0.04) added image, (**A**, **2**)-sketch image, (**B**, **2**)-Gaussian white noise ($\mu = 0.01$, $\sigma = 0.03$) added image, (**A**, **3**)-Photo Image, (**B**, **3**) salt & pepper noise (density = 0.2) added image, (**A**, **4**) photo Image, (**B**, **4**) Gaussian white noise ($\mu = 0.01$, $\sigma = 0.03$) added image

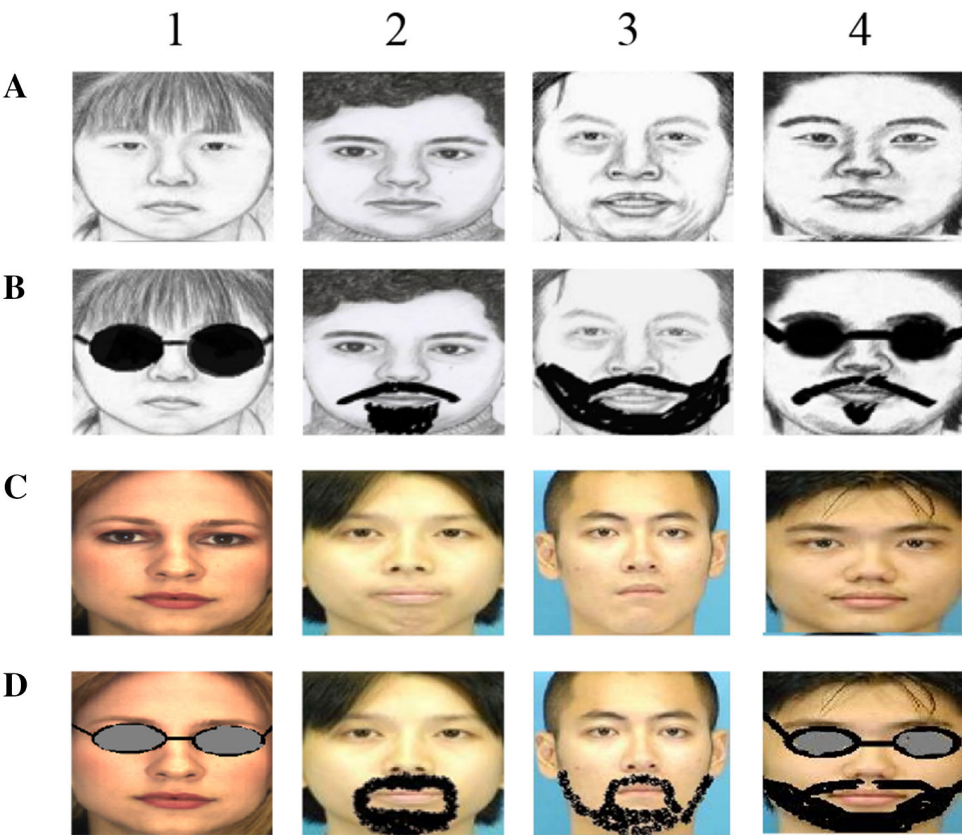

Fig. 7 *Row*-**A** Sketch images, *Row*-**B** (*left* to *right*) Glasses disguise added sketch, mustache disguise added sketch, beard disguise added sketch, and hybrid disguise (*glasses* & *mustache*) added sketch, *Row*-**C** photo images, *Row*-**D** (*left* to *right*) Glasses disguise added photo, mustache disguise added photo, beard disguise added photo, and hybrid disguise (*glasses* & *mustache*) added photo

Fig. 8 (**A**, **1**) Sketch image from CUHK dataset, (**B**, **1**) sketch image from AR dataset, (**C**, **1**) sketch image from XM2VTS dataset, (**D**, **1**) photo Image from CUHK dataset, (**E**, **1**) photo Image from AR dataset, (**F**, **1**) photo image from XM2VTS dataset, (**G**, **1**) Photos after added with disguise, (**H**, **1**) sketches after added with disguise, (*Col-2*)- Rank-1 matched images, (*Col-3*)-Rank-2 matched images, (*Col-4*)-Rank-3 matched images, (*Col-5*)-Rank-4 matched images

mentioned number of samples, as shown in Table 3. The Same set of training samples is used for the following algorithms.

EUCLBP + GA [15]: 3 Laplacian Pyramid of block size 8 × 8, 7 × 7, 6 × 6 is created respectively. For each Laplacian pyramid, CLBP (circular local binary pattern) of radius = 2 having points = 8 is calculated. Scores for each level of the Laplacian pyramid are fused using weighted sum rule fusion with weights ($w_0 = 0.2$, $w_1 = 0.3$, $w_2 = 0.5$). Finally, GA is used to find the optimized Chi square distance.

CITE [16]: Five trees in CITE forest and 256 nodes for each tree are used. Pattern sampling is done by a single radius of size: 2. Finally, PCA-LDA classifier is used for feature reduction.

SIFT + MLBP [1]: Each image is subdivided into overlapping patches of size = 32. For each patch SIFT and LBP of radius R = {1, 3, 5, 7}, points P = 8 are measured. LFDA classifier is used to measure the distance.

MCCA [31]: Each image is subdivided into overlapping patches of size = 12 × 12. For each patch HOG of 8 orientations and MLBP of radius R = {1, 3, 5, 7} are

Table 4 Comparison of rank-1 recognition accuracy of the MLGCS on effects added query sketches

Query images ↓	Methods				
	EUCLBP + GA	CITE	SIFT + MLBP	MCCA	MLGCS
Rank-1 accuracy (%) for query Sketches (after adding effects) against modified photo gallery of size: 1006					
Blur sketches of size: 100	81	93	87	91	94
Gaussian noise added sketches of size: 100	78	88	83	89	96
Salt & pepper noise added sketches of size: 100	80	91	85	90	95
Disguised sketches of size: 100	75	80	86	81	90
Normal test sketches of size: 606	85.67	94.67	88	95.33	98.33

measured. Random sampling is used to fetch the features into 100 canonical correlation based multiple-sub-classifier system. Finally, the cosine distance from each sub-classifier is summed up.

For testing the robustness of the proposed feature some new modified, manually created face photo images are included in the face photo database. As face photo galleries are matched with a query face sketch image, sketches are also added with some effects. Thus the face photo database, as well as sketches, are enriched with modified face photos. Sketch and photo images, which are not used as training samples for the algorithms mentioned above, are used for adding effects. Figures 5, 6, 7 and 8 are in matrix form, where rows are represented using alphabets: A, B…, and columns are represented using integer numbers: 1, 2, 3,… (A, 1) stands for A-th row and 1st column.

Adding effects

- The Gaussian blur filter, with window size 15 × 15 and $\sigma = 1.5$. (As shown in Fig. 5): Randomly selected 100 sketches and 100 photos are added with blur effect.
- Salt and Pepper Noise with a noise density = 0.04. Gaussian White Noise with $\mu = 0.01$ and $\sigma = 0.03$.

(As shown in Fig. 6): Randomly selected 100 sketches and 100 photos are separately added with different noise effect.

- Different variations in face disguise like beard, mustache, glasses etc. (As shown in Fig. 7): Randomly selected 100 sketches and 100 photos are added with disguise.

$$\text{Modified sketch gallery size} : 606 + 100 + 100 + 100 + 100 = 1006$$

$$\text{Modified photo gallery size} : 606 + 100 + 100 + 100 + 100 = 1006.$$

Table 4 shows the rank-1 comparison result of different algorithms after adding effects with query sketches and compared against modified photo gallery. The proposed MLGCS is much more robust than other methods in case of effects like Gaussian, and Salt & Pepper noise. Result degrades in case of blurring and disguised sketches. However, MLGCS outperforms other algorithms in case of blurring and disguised sketches. Normal sketches are also matched against modified gallery, and the proposed MLGCS performs better than other methods.

4.3 Query photo against sketch gallery

To show that the proposed FSPR-system is a kind of heterogeneous face recognition system, we have tested photos as query images and sketches as gallery. The face photos are also matched against gallery of sketches after adding effects.

Table 5 shows that the proposed MLGCS performed much better than other methods in case of effects like blurring, noise, and disguise.

Sample of successfully recognised sketches and photos from three different databases are shown in Fig. 8A–F. One sample of successful recognition of disguise (hybrid) photo is shown in Fig. 8Row-G. One sample of successful recognition of disguise (hybrid) sketch is shown in Fig. 8Row-H.

Proposed method is implemented in MATLAB in a system having 2 GHz Quad Core Processor and 4 GB RAM. To compute the LGCS descriptor for whole photo gallery database (1006 images), the system needs 105 s. For a given query image, the proposed algorithm requires 0.08 s to compute the LGCS descriptor along with match it to face photo gallery images.

5 Conclusion

In this paper, a method is presented for matching of face sketch images and face photo images using a newly developed feature extraction technique which uses local gradient checksum (LGCS) features to capture the local edge information about face sketch images, as well as

Table 5 Comparison of rank-1 recognition accuracy of the MLGCS on effects added query photos

Query images ↓	Methods				
	EUCLBP + GA	CITE	SIFT + MLBP	MCCA	MLGCS
Rank-1 accuracy (%) for query photos (after adding effects) against modified sketch gallery of size: 1006					
Blur photos of size: 100	78	84	83	86	87
Gaussian noise added photos of size: 100	72	81	76	87	89
Salt & pepper noise added photos of size: 100	76	87	82	87	91
Disguised photos of size: 100	70	74	83	80	85
Normal test photos of size: 606	80.67	91.67	84.33	92.33	95.67

face photo images. By converting both sketch and face images into same gradient domain, the modality difference between sketch and photo is reduced. Then local edge information is captured, and checksum mechanism is used to compare the sketch and photo images. To improve the result a multi-scale LGCS (MLGCS) is also proposed.

In the age of CCTV (closed-circuit television), disguise faces and partial faces are the most important scenarios. So face photo galleries can have those types' of images. Sketches may also have disguise. We have tested our proposed approach for frontal disguised sketch and photo images.

LGCS may have promising prospect in other fields of image processing like texture recognition, and which worth further investigations.

References

1. Klare BF, Li Z, Jain AK (2011) Matching forensic sketches to mug shot photos. IEEE Trans Pattern Anal Mach Intell 33(3):639–646
2. Jacobs DW, Weinshall D, Gdalyahu Y (2000) Classification with nonmetric distances: image retrieval and class representation. IEEE Trans Pattern Anal Mach Intell 22(6):583–600
3. Biederman I, Ju G (1988) Surface versus edge-based determinants of visual recognition. J Cognit Psychol 20(1):38–64
4. Bruce V, Hanna E, Dench N, Healey P, Burton M (1992) The importance of 'mass' in line drawings of faces. J Appl Congnit Psychol 6(7):619–628
5. Zhang T, Yan Y, Fang B, Shang Z, Liu X (2009) Face recognition under varying illumination using gradientfaces. IEEE Trans Image Process 18(11):2599–2606
6. Uhl RG Jr, da Lobo NV (1996) A framework for recognizing a facial image from a police sketch. In: Proc Int Conf Comp Vision Pattern Recognit, pp 586–593
7. Tang X, Wang X (2004) Face Sketch Recognition. IEEE Trans Circt Syst Video Tech 14(1):50–57
8. Liu Q, Tang X, Jin H, Lu H, Ma S (2005) A nonlinear approach for face sketch synthesis and recognition. In: Proc Int Conf Comp Vision Pattern Recognit, pp 1005–1010
9. Gao X, Zhong J, Li J, Tian C (2008) Face sketch synthesis algorithm on E-HMM and selective ensemble. IEEE Trans Circt Syst Video Tech 18(4):487–496
10. Wang X, Tang X (2009) Face photo-sketch synthesis and recognition. IEEE Trans Pattern Anal Mach Intell 31(11):1955–1967
11. Chang L, Zhou M, Han Y, Deng X (2010) Face sketch synthesis via sparse representation. In: Proc Int Conf Image Process, pp 2146–2149
12. Wang N, Gao X, Tao D (2011) Face sketch-photo synthesis under multi-dictionary sparse representation framework. In: Proc Int Conf Image and Graphics, pp 82–87
13. Gao X, Wang N, Tao D, Li X (2012) Face sketch-photo synthesis and retrieval using sparse representation. IEEE Trans Circt Syst Video Tech 22(8):1213–1226
14. Wang N, Tao D, Gao X, Li X, Li J (2013) Transductive face sketch-photo synthesis. IEEE Trans Neural Net Learn Syst 24(9):1364–1376
15. Bhatt HS, Bharadwaj S, Singh R, Vatsa M (2010) On matching sketches with digital face images. In: Proc Int Conf Biom Theory Appl Syst, pp 1–7
16. Zhang W, Wang X, Tang X (2011) Coupled Information-theoretic encoding for face photo-sketch recognition. In: Proc Int Conf Comp Vision Pattern Recognit, pp 513–520
17. Han H, Klare BF, Bonnen K, Jain AK (2013) Matching composite sketches to face photos: a component-based approach. IEEE Trans Info Forensic Secur 8(1):191–204
18. Pramanik S, Bhattacharjee D (2012) Geometric Feature Based Face-Sketch Recognition. In: Proc Int Conf Pattern Recognit Info Med Engg, pp 409–415
19. Rahman A, Beg S (2015) Face sketch recognition using sketching with words. Int J Mach Learn Cyber 6(4):597–605
20. Ojala T, Pietikäinen M, Mäenpää T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987
21. Heikkila M, Pietikainen M, Schmid C (2009) Description of interest regions with local binary patterns. J Pattern Recognit 42(3):425–436
22. Murala S, Maheshari R, Balasubramaniun R (2012) Directional local extrema patterns:a new descriptor for content based image retrieval. J Multimed Info Retr 1:191–203
23. Chen J, Shan S, He C, Zhao G, Pietikäinen M, Chen X, Gao W (2010) WLD: a robust local image descriptor. IEEE Trans Pattern Anal Mach Intell 32(9):1705–1720
24. Wang B, Li W, Yang W, Liao Q (2011) Illumination normalization based on weber's law with application to face recognition. IEEE Signal Process Letters 18(8):462–465
25. Forouzan BA (2010) Data communications and networking. 4th ed. Tata McGraw-Hill, New Delhi, India **(Special Indian Edition)**
26. He M, Horng S, Fan P, Run R, Chen R, Lai J, Khan MK, Sentosa KO (2010) Performance evaluation of score level fusion in multimodal biometric systems. J. Pattern Recognit 43(5):1789–1800

27. Pun C, Lee M (2003) Log-polar wavelet energy signatures for rotation and scale invariant texture classification. IEEE Trans Pattern Anal Mach Intell 25(5):590–603
28. CUHK Face Sketch database. [Online]. http://mmlab.ie.cuhk.edu.hk/facesketch.html
29. Martinez AM, Benavente R (1998) The AR face database. CVC Tech Report. p 24
30. Messer K, Matas J, Kittler J, Luettin J, Maitre G (1999) XM2VTSDB: The Extended of M2VTS Database. In: Proc Int Conf Audio Video Person Auth, pp 72–77
31. Gong D, Li Z, Liu J, Qiao Y (2013) Multi-feature canonical correlation analysis for face photo-sketch image retrieval. In: ACM International Conference on Multimedia, pp 617–620

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com