



# Human authentication based on fusion of thermal and visible face images

Ayan Seal<sup>1</sup>  · Chinmaya Panigrahy<sup>1</sup>

Received: 22 October 2018 / Revised: 29 March 2019 / Accepted: 26 April 2019 /  
Published online: 20 May 2019  
© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

In recent past, considerable amount of research has been done to increase the performance of a face authentication system in uncontrolled environment such as illumination. However, the performance has not been improved significantly since visible face images are dependent of illumination. To overcome the limitation of visible face images, researchers are using infrared (IR) face images. However, it also does not completely independent of illumination. Image fusion of visible and thermal face images is an alternative in research community nowadays. In this work, a fusion method is introduced to fuse visible and IR images for face authentication. The proposed fusion method relies on translation invariant  $\hat{A}$ -trous wavelet transform and fractal dimension using differential box counting method. Five popular fusion metrics namely, ratio of spatial frequency error, normalized mutual information, edge information, universal image quality index, extended frequency comparison index are considered to measure the effectiveness of the proposed fusion algorithm quantitatively over four state-of-the-art methods. A new similarity measure is also proposed to check how close a fused face image from others are. All the experiments are performed on three databases namely, IRIS benchmark face database, UGC-JU face database and SCface face database. All the results depict that the proposed fusion method along with similarity measure for face authentication outperforms all the four state-of-the-art methods in terms of accuracy, precision and recall.

**Keywords** Visible face · Thermal face · Fractal dimension · Maximum bipartite matching · Image fusion

## 1 Introduction

Automatic human authentication based on face biometric has been extensively studied in computer vision with the knowledge of human physiology. Despite extensive studies, many problems in face authentication still persist due to the inherent difficulties of extracting face

---

✉ Ayan Seal  
ayan.seal@gmail.com

<sup>1</sup> PDPM Indian Institute of Information Technology, Design and Manufacturing,  
Jabalpur, 482005, India

biometric [55]. A wide variety of problems such as lighting and shadows plague the attempt for unconstrained face authentication. These problems are known as illumination problems. In addition to illumination problem, face authentication is equally challenging due to variations in facial expression, pose and disguise [3, 6, 30, 40, 42, 52]. Availability of several sensors has fuelled the interest of research community in this field. Optical sensor, infrared (IR) sensor and 3D sensor are few examples of sensors used in face authentication. These sensors are utilized for extracting different information, which are usually complementary in nature. For instance, an optical sensor captures visible band whereas an IR sensor acquires information about IR band of the electromagnetic spectrum. It is a challenging task to perform face authentication using visible images in an uncontrolled environment. IR sensor traps the emitted heat energy from face region and generates thermogram which is more robust to illumination disparities. This virtue of IR sensor is exploited by researchers nowadays, and they are focusing on IR imaging based face authentication system. An IR sensor focuses on electromagnetic spectrum with wavelength in the range of 8 to 14 micro meter / micron [41]. Thus, we can see that even under varying illumination conditions, the usage of IR image in face authentication has significant benefits over visible image. However, an IR face image is quite sensitive towards the ambient temperature changes. The heat pattern emitted from the body also fluctuates with its temperature, thus leading to inaccurate classifications. In other words, face authentication using visible images works effectively in such circumstances [39]. Few examples of visible and infrared (IR) face images from UGC-JU [41], IRIS [12] and SCface face databases have been shown in Figs. 1, 2 and 3 respectively.

Here, one can conclude easily that these modalities are not able to convey significant information in one single image. Therefore, multimodal fusion is required to extract all possible essential information in a fused image. Image fusion is a process to combine complementary information in addition to the redundant information present in source images in such a manner that the fused image carries more information for human or machine perception in contrast to any of the individual images. The redundant information increases the reliability and accuracy of a face authentication system. At the same time, integration of complementary information enhances the understandability of the fused image [19, 24, 37]. Fusion techniques could be classified into data, feature, score, rank, and decision levels [22] based on the stage at which it operates. Data and feature level fusions are known as low-level fusion whereas score, rank, and decision level fusions are called high-level fusion. The low-level fusion is superior to high-level fusion due to the following reasons: Firstly, an information loss is involved in each stage of succession from data to decision level of image



Fig. 1 The UGC-JU visible and their corresponding IR face images



**Fig. 2** The IRIS visible and their corresponding IR face images

fusion in a biometric security system. Hence, the sooner the fusion is performed, the wealthier is the data carried to the subsequent stages. Secondly, most of the researchers use fusion at lower levels in their work for its simplicity. Lower level fusion does not affect the design of the classifiers so it is easy. However, the statistical dependences between classifiers badly affect the performance of higher level fusion.

Normally, Mallat or pyramidal wavelet transforms, for instance ‘Haar’ and ‘Daubechies’, are considered for multi-resolution image fusion. Recognition of fused face images, by utilizing state-of-the-art methods, is shown in Section 2. However, Mallat wavelet transform (MWT) suffers from translation, which may reduce the accuracy of a biometric face authentication system significantly. It motivates to choose a translation invariant or stationary wavelet transform called as À-trous wavelet transform (AWT) in this work. The translation-invariance is accomplished by removing the down-samplers and up-samplers in the MWT and up-sampling the filter coefficients by a factor of  $2^{(l-1)}$  in the  $l^{th}$  level of AWT [45, 51]. AWT provides an output image after each level of decomposition, which is same as original image in size. It means AWT is an inherently redundant approach by default. In this study,



**Fig. 3** The SCface visible and their corresponding IR face images

a fusion rule is formulated using the approximation and wavelet coefficients of the AWT of input face images along with differential box counting (DBC) based fractal dimension (FD) method [38]. After obtaining the fused face images using the proposed fusion method, which would be discussed in Section 4, the superiority of the fused images has been estimated quantitatively using some of existing fusion metrics such as mutual information [35], sobel edge detection [56], spatial frequency [14] and universal image quality index [54], which illustrate the usefulness of the proposed method over existing fusion methods. In order to authenticate humans, instead of considering conventional machine learning algorithms for estimating the similarity between two face images, a novel similarity measure is proposed in this work. The proposed similarity measure is relied on maximum matching of a bipartite graph, which would be formed using superpixels of two fused face images and the weight of each edge is defined by a cost function which is described in Section 4.

The rest of this paper is ordered as follows. Section 2 presents a short literature review on fused face recognition. Different methods that help to define fusion rule and similarity measure between two fused face images are described in Section 3. The proposed fusion rule is illustrated in Section 4. Section 4.2 describes some of the existing fusion metrics for evaluating of the proposed fusion algorithm in addition to existing fusion algorithms. The experimental results and discussion are presented in Section 5. A comparative study of the proposed method with the state-of-the-art methods is also discussed in Section 5. Finally, the conclusion is drawn in Section 6.

## 2 Related work

The method proposed by S. G. Kong et al. in [25] is one of the pioneers in face recognition, discussed different aspects of visible and IR images and presented a fusion method for combining these two modalities images. Prior to image fusion, images were registered using some software. S. G. Kong et al. performed their experiments on NIST/Equinox and UTK-IRIS databases and the obtained recognition accuracies are 85% and 67% respectively. In [9], M. K. Bhowmik et al. presented a pixel-level fusion method using IR and visible face images by integrating 70% visible and 30% IR information at each pixel position. Then the fused images were mapped into an eigenspace, which was fed further into radial basis function and multi-layer perceptron based Artificial Neural Network (ANN) classifiers for recognition and the noted recognition accuracies are 96% and 95.07% respectively on Object Tracking and Classification Beyond Visible Spectrum data set (OTCBVS). The critical comment about this work is that there is no adequate reasoning behind integrating 70% visible and 30% IR face information to obtain the fused image. In [5], G. Bebis et al. introduced pixel-level and feature-level fusion techniques. ‘Haar’ wavelet was employed on both IR and visible face images. Then, a mask was created, which consists of the same number of pixels as the original image based on Genetic Algorithm (GA). This mask assisted to choose the wavelet coefficients from both the IR and visible face images. Two modalities face images were mapped separately into eigenspace for obtaining the features in their second experiment. Next, GA was considered to choose the eigen features from both the images. All the experiments were done on Equinox face database. The obtained maximum recognition rate is about 97% when eyeglasses were not present both in the gallery and probe sets contain multiple illuminations. In [47], R. Singh et al. presented a combined image fusion and match score fusion of multi-modality face images. For fusion of visible and IR face images, 2v-granular SVM (2v-GSVM) was used. The 2v-GSVM employed multiple SVMs to learn both the local and the global characteristics

of the multi-modality face images at different granularity levels and resolutions. The 2v-GSVM determines some weighted factors, which help to form the fused image. Then local and global features from the fused images were extracted using 2D log-polar Gabor transform [48] and local binary pattern [4, 7] algorithms respectively. The corresponding match scores were mixed by Dezert-Smarandache theory of fusion [47], which uses plausible and paradoxical reasoning. The presented method was tested on Notre Dame and Equinox databases and was assessed with existing statistical learning, and evidence theory based fusion methods. The verification accuracies are 95.85%, 94.80% using 2D log-polar Gabor, local binary pattern and 94.98%, 94.71% using 2D log-polar Gabor, local binary pattern on Norte Dame and Equinox databases respectively. A two-level hierarchical DWT based data level image fusion using IR and visible face images were presented by R. Singh et al. in [46]. Then 2D log-polar Gabor wavelet was considered to extract amplitude and phase features from the fused image. An adaptive SVM learning algorithm intelligently chooses either amplitude or phase features to produce a fused feature set. R. Singh et al. did their experiments on Equinox face database and shown that the integration of visible light and short-wave IR spectrum face images produced the best recognition rate with an equal error rate of 2.86%. In [8], M. K. Bhowmik et al. presented relative studies on fusion of visual and IR images by ‘Haar’ and ‘Daubechies’ wavelet transformations. The decomposition operation up to level 5 had been done on both types of images independently using ‘Haar’ and ‘Daubechies’. The wavelet coefficient for the fused image was formed by choosing higher magnitude coefficients from approximation coefficients of both IR and visible face images and smaller magnitude coefficients from the detail coefficients of both the modalities of images. After applying the inverse transform on the fused coefficient, the fused image was generated. PCA was considered to extract the features from the fused images. Here, multilayer perceptron (MLP) had been adopted for the classification purpose. Experiments had been done on IRIS dataset to validate the presented methods. The average recognition rates are 87% and 91.5% for ‘Haar’ wavelet and ‘Daubechies’ wavelet based methods respectively. In [53], N. Wang et al. presented a complex fusion strategy at both pixel-level and feature-level and different classification methods like two-dimensional PCA (2DPCA), two-dimensional LDA (2DLDA), two-directional two-dimensional PCA ((2D)<sup>2</sup>PCA), two-directional two-dimensional LDA ((2D)<sup>2</sup>LDA) and two-directional two-dimensional Fisher PCA ((2D)<sup>2</sup>FPCA). The NVIE visible and T-IR face database were used for their experiments. The obtained maximum recognition rates for pixel-based complex algorithm using ((2D)<sup>2</sup>LDA) and feature-based complex algorithm using ((2D)<sup>2</sup>FPCA) were 97.09% and 97.97% respectively. In [20], G. Hermosilla et al. presented a fusion method using GA. The presented method merged the most relevant information from IR and visible face images with image descriptors. GA searches weights for the IR and visible face images in the form of a genetic code in order to maximize the recognition rate as an objective function. G. Hermosilla et al. [20] used Equinox and PUCV-VTF databases for their experiments and reported 97% and 99% recognition for these two databases respectively.

### 3 Background

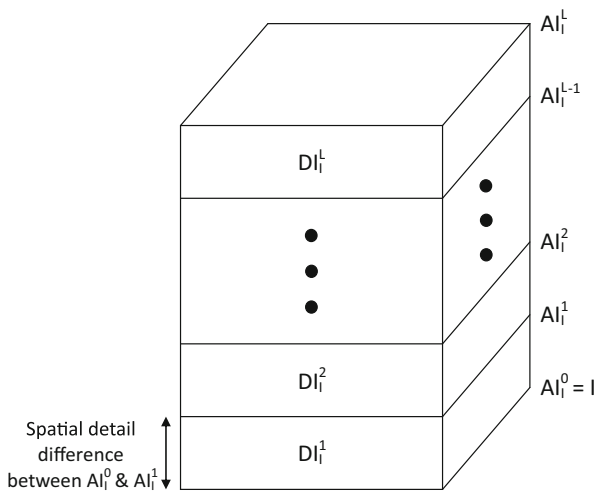
In this section, different methods namely, AWT, FD, local ternary pattern (LTP), superpixels, maximum matching would be discussed, which help to design a fusion rule to fuse IR and visible face images followed by similarity measure between two fused images for face authentication.

### 3.1 À-trous wavelet transform

The transform domain based methods to accomplish image fusion are gaining importance due to its good fusion performance [34]. However, in most of the cases, this type of methods based on DWT use pyramidal decomposition hence are translation-variant. Dutilleux [13] introduced a translation-invariance wavelet transform known as AWT, where down-sampling and up-sampling in DWT are removed to achieve translation-invariance. In AWT, À-trous is derived from the French word *trous* meaning holes in English. AWT is also known as stationary wavelet transform, undecimated DWT and redundant DWT [15]. AWT is also a non-orthogonal, shift invariant, symmetric, dyadic wavelet transform. At each level of decomposition, AWT decomposes an approximation image (AI) of an input image  $I$ , into a coarser AI and a detail image (DI), where AI and DI contain the approximation/low-pass and detail/high-pass coefficients respectively. The image decomposition scheme used by AWT is inherently redundant [49], where exactly same number of pixels are there in input and output images after each level of decomposition. Hence, AWT decomposition scheme can be depicted using a parallelepiped as shown in Fig. 4. Moreover, the successive AIs while going up through the resolution levels of the parallelepiped, have a coarser spatial resolution. The AI of  $I$  at level  $l$ ,  $AI_l^l$ , can be computed using (1).

$$AI_l^l(i, j) = AI_l^{(l-1)}(i, j) \otimes h_l \forall l = 1, 2, \dots, L, \tag{1}$$

where  $h_l$  is the b3-spline scaling function for level  $l$ ,  $L$  is the maximum level of decomposition and  $\otimes$  represents the convolution operator. Practically, AI of  $I$  at level  $l$  is obtained by convolving the AI of  $I$  at level  $(l - 1)$  with the low-pass scaling function for level  $l$ . The  $0^{th}$  level AI of  $I$  is the image  $I$  itself i.e.  $AI_l^0(i, j) = I(i, j)$ . Moreover, the AI of  $I$  at the maximum level  $L$ ,  $AI_l^L(i, j)$  is sometimes simply represented as AI of  $I$  i.e.  $AI_l(I, j)$ . The AI of  $I$  at different levels of decomposition represents its low frequency information at that level respectively.



$I$ : Input image,  $L$ : Maximum level of decomposition,  $AI_l^l$  Approximation Image of  $I$  at level  $l$ ,  $DI_l^l$ : Detail Image of  $I$  at level  $l$ .

**Fig. 4** Illustration of parallelepiped AWT

The b3-spline scaling function for level 1,  $h_1$  is obtained using (2) [16].

$$h_1 = \frac{1}{256} \begin{pmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix} \tag{2}$$

The scaling function for a subsequent level is obtained from the scaling function of its previous level by interlacing zeros in between the rows and columns respectively. For example, scaling function for level 2,  $h_2$ , is obtained from  $h_1$  by placing zeros in between the rows and columns which is represented in (3).

$$h_2 = \frac{1}{256} \begin{pmatrix} 1 & 0 & 4 & 0 & 6 & 0 & 4 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 16 & 0 & 24 & 0 & 16 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 6 & 0 & 24 & 0 & 36 & 0 & 24 & 0 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 16 & 0 & 24 & 0 & 16 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 4 & 0 & 6 & 0 & 4 & 0 & 1 \end{pmatrix} \tag{3}$$

The spatial information lost between two successive AIs of the parallelepiped are collected in a single DI [17] which is shown in Fig. 4. The DI of  $I$  comprises its high frequency information. The DI of  $I$  at level  $l$ ,  $DI_I^l$  is computed as the difference between the AI of  $I$  at level  $(l - 1)$  and  $l$  respectively which is represented by (4).

$$DI_I^l(i, j) = AI_I^{(l-1)}(i, j) - AI_I^l(i, j) \forall l = 1, 2, \dots, L \tag{4}$$

The AI of  $I$  at a level  $l$  can be reconstructed by adding the DIs of all successive decomposition levels with the AI of last level  $L$ , which is depicted in (5). Hence, the original input image,  $I$ , is restored by adding the DIs of all the decomposition levels with the AI of the maximum decomposition level  $L$ .

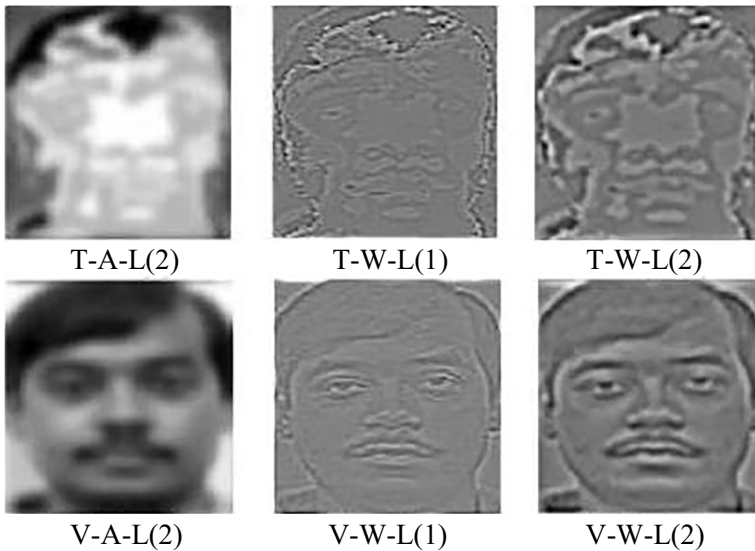
$$AI_I^l(i, j) = AI_I^L(i, j) + \sum_{k=1}^{L-l} DI_I^{l+k}(i, j) \tag{5}$$

AWT would be considered to devise a fusion rule for fusing IR and visible face images and the detail discussion would be there in Section 4. The approximation and wavelet coefficients after second level decomposition of both visible and IR face images of a person shown in Fig. 1 (second from the left) are shown in Fig. 5.

### 3.2 Fractal dimension using differential box counting method

FD is a measure of roughness or irregularities in the form of self-similar elements present in an image. Mandelbrot [28] introduced the term fractal. It is a Latin word and the meaning of this word is *fractus* in English, which denotes irregular segments. Several approaches are there in literature to calculate FD of an image [23, 31, 32]. These methods are quite expensive [38]. Sarkar and Chaudhuri [38] presented an efficient approach, called DBC to compute FD of a gray-scale image. Generally, DBC is used to measure texture of an image [29]. Consider an image of size  $M \times M$  pixels as a 3-Dimensional space where  $x$  and  $y$  coordinates represent the length and breadth of the 2-dimensional image plane respectively.





**Fig. 5** AI and DI of the second left images of Fig. 1 (T-IR,V-visible, A-AI, W-DI, and L-level of decomposition)

The  $z$  coordinate denotes the height i.e. intensity value of a particular pixel located at  $(x, y)$ . The  $xy$ -plane is partitioned into non-overlapping grids of size  $s \times s$  pixels, where  $s$  is an integer and varies from 2 to  $M/2$ . The gray-level of incomplete grids outside the image boundary is treated as zero. The scale of a grid of size  $s \times s$  is  $r$ , where  $r = s/M$ . On each grid there is a number of boxes of size  $s \times s \times h$ , where  $h$  is the height of box as shown in Fig. 6,  $(G/h) = (M/s)$  and  $G$  is the total number of gray-levels.

Let the maximum and minimum gray-levels on the  $(i, j)^{th}$  grid be  $g_{max}$  and  $g_{min}$  respectively. Suppose  $n_r(i, j)$  is the number of boxes required to fill  $(i, j)^{th}$  grid at scale  $r$ , which can be computed by (6).

$$n_r(i, j) = \left\lceil \frac{g_{max}}{h} \right\rceil - \left\lceil \frac{g_{min}}{h} \right\rceil + 1 \tag{6}$$

Total number of boxes,  $N_r$ , are required to fill the whole image at scale  $r$  is calculated by (7).

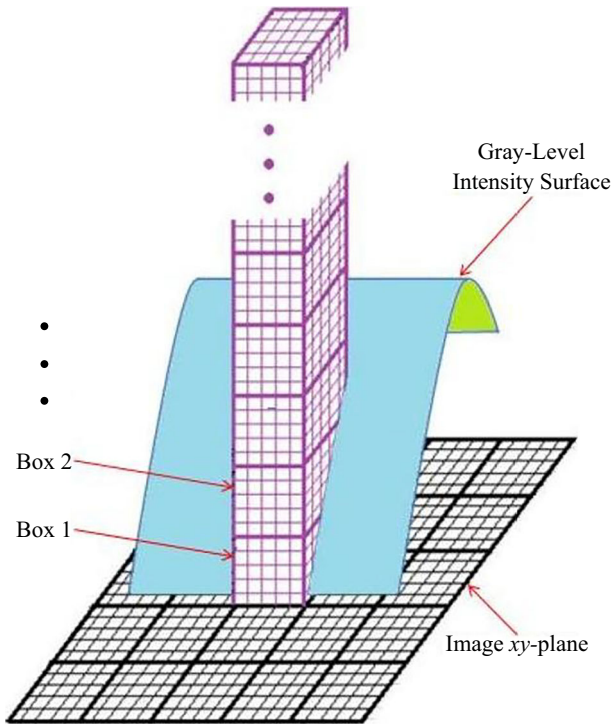
$$N_r = \sum_{i,j} n_r(i, j) \tag{7}$$

FD of the image is the slope of a line generated by least square linear fit method after plotting some points, where  $x$ -axis and  $y$ -axis represent  $\log(1/r)$  and  $\log(N_r)$  respectively. DBC also helps to construct the proposed fusion rule and further be discussed in Section 4.

### 3.3 Local ternary pattern

The local binary pattern (LBP) is used by many researchers to extract facial features due to its computational efficiency and discriminative power. In order to improve the performance of single resolution LBP, Chen et al. [10] have used multi-resolution LBP to extract face features. In case of LBP, all the neighboring pixels are equally treated but different neighboring pixels may contribute differently to the facial description. In order to solve the above





**Fig. 6** The process of finding out the number of boxes ( $n_r$ ) by DBC method. (Here,  $M=25$ ,  $s=5$  and  $n_r$  is 3 for the grid)

issue, Lei et al. [26] have proposed an optimal neighborhood sampling strategy which multiplies a pixel difference vector with an optimal learned soft sampling matrix for generating the facial feature vector. Tan et al. [50] have proposed local ternary pattern (LTP) to extract face features at different lightening conditions, which is an extension of LBP. It is normally used for texture extraction in uniform and near-uniform regions, which would be treated as features. Unlike LBP, LTP creates a ternary pattern based on (8).

$$s(I_k, I_c) = \begin{cases} -1, & \text{if } I_k < I_c - t \\ 0, & \text{if } I_k > I_c - t \text{ and } I_k < I_c + t \\ 1, & \text{if } I_k > I_c + t, \end{cases} \quad (8)$$

where  $I_k$  are neighboring pixels of a  $3 \times 3$  window surrounded by a centre pixel,  $I_c$  and  $t$  is threshold value. The value of  $t$  is 7 and is inherited from [50]. Thus,  $k$  varies from 0 to 7 as there are 8 neighboring pixels around  $I_c$ . Figure 7 demonstrates basic LTP process for a given  $3 \times 3$  image patch. The value of a particular pixel of upper pattern,  $U_k$  and lower pattern,  $L_k$ , would be 0 if the value of the  $I_k$  of the original image patch is between  $(I_c - t)$  and  $(I_c + t)$  respectively. Any values which are greater than  $I_c + t$  are assigned value 1 into upper pattern. Similarly for forming the lower pattern, any values that are lesser than  $(I_c - t)$  are assigned value 1 into lower pattern. Then ternary pattern would be produced by combining upper pattern and lower pattern. When the value of  $L_k$  is 1 then the respective position in ternary pattern would be -1 whereas in the position where the value of  $U_k$  is 1 then the respective position in the ternary pattern would be 1. All other positions of ternary

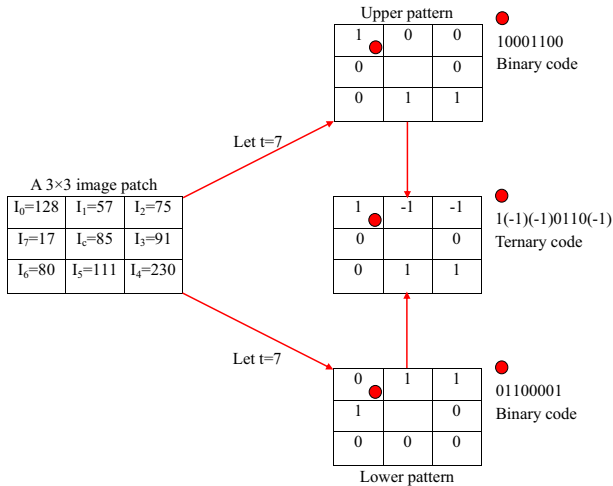


Fig. 7 Illustration of LTP operator

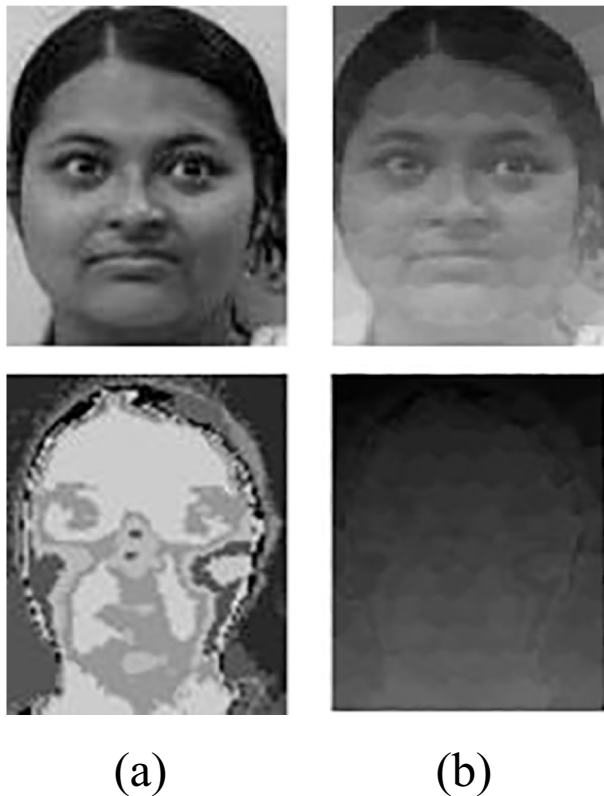
pattern would be filled by 0 value. The final binary code of upper pattern & lower pattern and ternary code while reading the bit pattern starting from the east location are marked by red color in Fig. 7 with respect to the centre, then going around counter-clockwise.

### 3.4 Simple linear iterative clustering superpixels

The concept of superpixel was first coined by X. Ren et al. in 2003 [36]. Superpixel is a group of connected pixels perceptually meaningful atomic regions with similar colors or gray-levels. Several approaches are there in literature to generate superpixels of an image. However, simple linear iterative clustering (SLIC) method is adopted in this work because it is faster and more memory efficient than others [2]. Normally, SLIC takes an image as input for clustering along with  $K$  as a parameter, which represents desired number of approximately equally sized superpixels. The color and position of each pixel of the input image would be denoted in the CIELAB color space,  $[lab]^T$  and spatial space  $[xy]^T$  respectively. Then a weighted distance measure integrates color and spatial proximity and groups pixels using  $k$ -means algorithm. The detail discussion of SLIC superpixels method is there in [1]. Figure 8 illustrates a visible and a IR face images from UGC-JU face database and their segmented images using SLIC of 100 number of superpixels.

### 3.5 Maximum bipartite matching

A bipartite graph is an undirected graph  $G = (V, E)$ , where each vertex belongs to one of two disjoint vertex sets,  $X$  or  $Y$ , and every edge connects a vertex in  $X$  to a vertex in  $Y$  [11]. A matching is the one-to-one pairing of some or all of the edges of one vertex set,  $X$ , with the edges of a second vertex set,  $Y$ , otherwise, unmatched. Figure 9 shows the notion of a matching in a bipartite graph. A complete matching is when every edge of  $X$  is paired with one edge of  $Y$ . A maximum weighted bipartite matching is the sum of the largest subset of edges in a bipartite graph such that no two selected edges share a common vertex. A new similarity measure between two fused face images has been proposed based on maximum



**Fig. 8** **a** Example of IR and visible face images of UGC-JU face database **b** segmented face images using SLIC of 100 number of superpixels

bipartite matching algorithm in this work. The detail formulation would be discussed in Section 4. The steps of maximum bipartite matching algorithm are shown in Algorithm 1.

---

**Algorithm 1** Maximum bipartite matching algorithm.

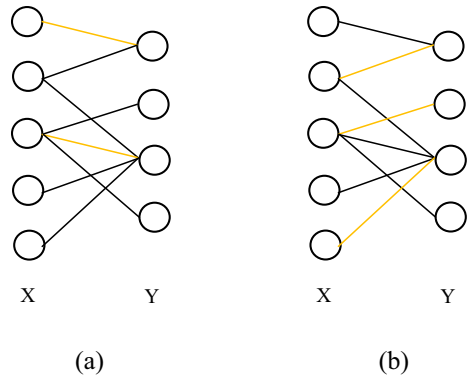
---

- 1: Start with a non-trivial matching.
  - 2: Find an *alternating path*, which starts at an unmatched initial matching of  $X$ , then goes along a path of edges that are *not in* (but possible) *and in* the initial matching and ends at an unmatched initial matching of  $Y$ .
  - 3: Note and change the status of all edges if an alternating path can be found. If no alternating path can be seen, stop and the obtained matching is a maximum matching.
  - 4: List the new matching with the new edges from the alternating path and the edges unmatched from the initial matching that are unchanged by the alternating path.
  - 5: When the matching is complete, stop, otherwise go to Step 2.
- 

## 4 Proposed method

In this section, a pixel-level fusion method using AWT and FD is proposed to fuse IR and visible face images for authentication purpose, which is followed by similarity measures based on superpixels and maximum bipartite matching algorithm between two fused face images.

**Fig. 9** A bipartite graph  $G=(V, E)$  with vertex partition. **a** A matching with number 2, marked by orange color edges. **b** A maximum matching with number 3, marked by orange color edges



#### 4.1 Fusion algorithm

In image fusion scheme, the source images namely, visible face,  $I_{vi}$ , and IR face,  $I_{th}$ , called as  $AI^0$ s are decomposed independently using (1) with the help of a scaling function,  $h_1$ , at desired level that could be obtained by (2) to form the lower-order approximation coefficients of  $I_{vi}$  and  $I_{th}$ . However, the higher-order detail coefficients of both the source images could be obtained using (4). The size of both the approximation and detail coefficients would be same as original image. The approximation coefficients provide the coarse or blurred form an image, which contains the base information. On the other hand, detail coefficients consist of edge and corner points present in an image. Most of the researchers selected the level of decomposition either randomly or experimentally for their works. In this study, a mathematical formula ( $\log_2 \min(S, T)$ ), has been devised to find the level of decomposition, where  $S$  and  $T$  are the number of rows and columns of an input image respectively. Fusion rule depends on the information present in the source images and the information required in fused image too. The objectives of the proposed fusion strategy are to preserve the temperature distributions of the IR face images along with sharpness and contrast of the visible face images in fused face images. The fusion rule for the approximation coefficient would be devised in such a way, that it can preserve all the basic information from both the source face images without any loss of generality. Moreover, it can retain sharpness and contrast of the source face images. After realizing these objectives, two fusion rules have been proposed one for combining the approximation coefficients and the other is for the detail coefficients of both visible and IR face images. A pseudocode for the proposed fusion algorithm is given by Algorithm 2.

In order to merge the approximation coefficients of both  $I_{vi}$  and  $I_{th}$ , a fusion rule is defined in line number 20 of Algorithm 2, which combines both the approximation coefficients of the source face images by taking average of them in order to form approximation coefficients of the fused face image. However, FD using DBC method is considered to formulate another fusion rule for combining the detail coefficients of  $I_{vi}$  and  $I_{th}$  upto level-number of decomposition that preserves higher frequency features like temperature distributions of the IR face images in the detail coefficients of the fused image. Normally, FD using DBC method returns a real number for an image, which is a measure of image roughness or texture. However, DBC method does not directly apply on each level of decomposed detail coefficients of both  $I_{vi}$  and  $I_{th}$  because DBC method returns a single real value, which does not help in fusion process. So, each level of decomposed detail coefficients of both  $I_{vi}$  and  $I_{th}$  are considered separately. Then a  $3 \times 3$  grid is moved from top-left corner

to bottom-right corner of a particular detail coefficients of both  $I_{vi}$  and  $I_{th}$ . Here, FD using DBC method helps to estimate the number of boxes is required to represent the roughness of a particular grid. Thus, a simple if-else rule is proposed to form a decision map based on the number of boxes of a grid for each level of decomposed detail coefficients of both the  $I_{vi}$  and  $I_{th}$ . The decision map is a two dimensional binary array of same size as the input images. If the number of boxes of a particular grid for a sub-band of the detail coefficients of the  $I_{th}$  is greater than visible counterpart,  $I_{vi}$  then the respective position of the decision map corresponding to the above sub-band would be 1 otherwise 0 is stored.

---

**Algorithm 2** Proposed fusion algorithm.
 

---

**Input:**
 $[I_{th}]_{S \times T}$  and  $[I_{vi}]_{S \times T}$ 
**Output:**
 $[I_f]_{S \times T}$ 
**Initialize:**
 $L$  to  $\log_2 \min(S, T)$ 
 $[I_f^d]_{S \times T \times L}$ ,  $[I_f]_{S \times T \times L}$  and  $[DP]_{S \times T}$  to zero matrix
**Required Functions:**

*atrousWavelet()* //Decomposes an image upto  $L$  levels using AWT and returns the corresponding AI and DIs

*diffBox()* //Returns the number of boxes required to represent the rough surface of the image using a  $3 \times 3$  window

*consVarification()* //Applies consistency verification to the decision map while fusing DIs

```

1:  $[I_{vi}^a, I_{vi}^d] = \text{atrousWavelet}(I_{vi}, L)$ ;
2:  $[I_{th}^a, I_{th}^d] = \text{atrousWavelet}(I_{th}, L)$ ;
3: for  $i = 0 \rightarrow L$  with a step size 1 do
4:    $tmp1 \leftarrow I_{vi}^d(:, :, i)$ ;
5:    $tmp2 \leftarrow I_{th}^d(:, :, i)$ ;
6:    $nR_{tmp1} \leftarrow \text{diffBox}(tmp1)$ ;
7:    $nR_{tmp2} \leftarrow \text{diffBox}(tmp2)$ ;
8:   for  $j = 1 \rightarrow S$  with step size 1 do
9:     for  $k = 1 \rightarrow T$  with step size 1 do
10:      if  $nR_{tmp1}(j, k) > nR_{tmp2}(j, k)$  then
11:         $DP(j, k) \leftarrow 1$ ;
12:      else
13:         $DP(j, k) \leftarrow 0$ ;
14:      end if
15:    end for
16:  end for
17:   $cv \leftarrow \text{consVarification}(DP)$ ;
18:   $I_f^d \leftarrow I_f^d + cv$ ;
19: end for
20:  $I_f^a \leftarrow (I_{th}^a + I_{vi}^a)/2$ ;
21:  $I_f \leftarrow I_f^a + I_f^d$ ;

```

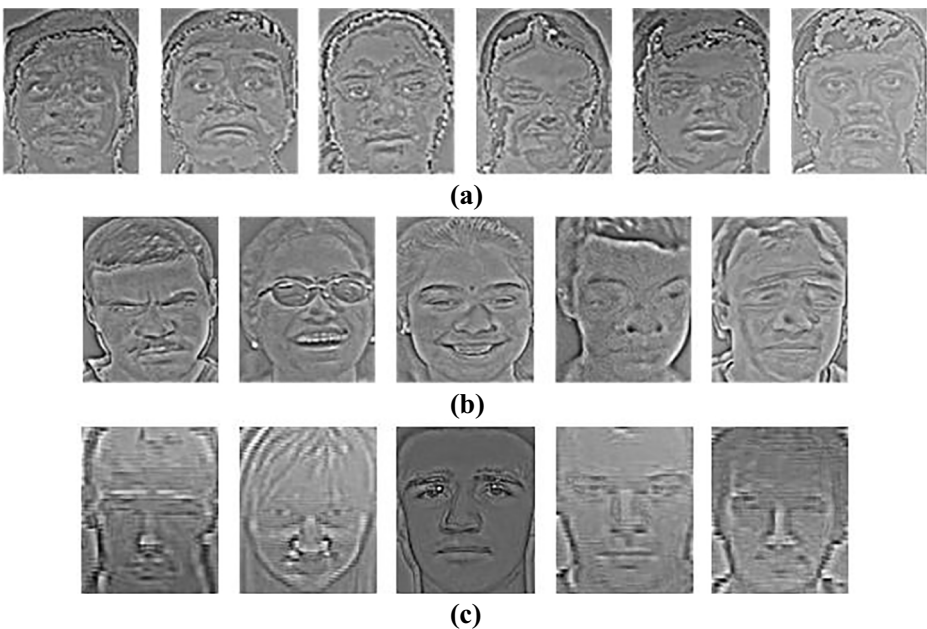
---

Each binary value of the decision map (DP) designates the type of image (IR or visible) contributing to the formation of fused image. A consistency verification of the DP is done for elimination of noise. Noise is characterized by the presence of an element value not

consistent with the others in a neighborhood. For handling this type of situation, a  $3 \times 3$  window is scrolled over the output that is obtained after executing line number 16 in order to eliminate the presence of noise and the value of the element coinciding with the centre of the window is made equal to that of the majority in the  $3 \times 3$  neighborhood. After noise removal, if the element  $(j, k)$  of the DP contains 1, coefficients for a particular decomposed details coefficients of the fused image at this pixel location will come from  $I_{th}^d$  and for the element value equal to 0,  $I_{vi}^d$  will be considered in line number 17. Then all the decomposed detail coefficients are added together in line number 18 to form the detail coefficients of the fused face image. Finally, the fused image is reconstructed by performing the inverse AWT using (5) in line 21. Figure 10 shows the fused face images obtained by Algorithm 2 of Figs. 1, 2 and 3.

## 4.2 Similarity measure

Final step of a face authentication system is matching. Instead of considering conventional machine learning algorithms, a novel similarity measure is proposed, which checks how close two fused face images are. The advantage of the proposed similarity measure is that it does not require training set separately. It focuses on one-to-one corresponds between one query image and the stored images. This similarity measure problem could be thought of as maximum bipartite matching problem. The detail discussion of maximum bipartite matching problem is there in Section 3.5. The first step of the proposed measure is to generate superpixels of both a query and one stored images. Section 3.4 presents the superpixels generation method. The centroid of all the superpixels of a query image are stored in set  $X$  whereas  $Y$  set consists of centroids of all the superpixels of one stored image. These two disjoint sets are used to form a bipartite graph, where the centroid of all the superpixels in



**Fig. 10** Some examples of fused images. **a** Fused face images for face images in Fig. 1. **b** Fused face images for face images in Fig. 2. **c** Fused face images for face images in Fig. 3

$X$  and  $Y$  sets are thought as a vertex. Edges are formed between all the centroid vertices in  $X$  to all the centroid vertices in  $Y$ . However, there would be no connection between two vertices in either  $X$  or  $Y$  since  $X$  and  $Y$  form a bipartite graph. All the edges are associated with a cost function, which would be estimated by combining spatial and gray proximity along with the distance between LTP codes. The cost function of an edge between  $i^{th}$  superpixel in an image  $p$  to  $j^{th}$  superpixel in an image  $q$  is calculated using (9).

$$D_{p_iq_j} = \frac{m}{R} \times \underset{\text{spatial distance}}{d_{p_iq_j}^{xy}} + \underset{\text{gray proximity}}{d_{p_iq_j}^g} + \underset{\text{ltp distance}}{d_{p_iq_j}^{ltp}}, \tag{9}$$

where cost function,  $D_{p_iq_j}$ , is the sum of gray proximity and ltp distance along with normalized spatial distance by grid interval  $R$ , which could be obtained by (10).

$$R = \frac{\sqrt{S \times T}}{K}, \tag{10}$$

where  $S$  and  $T$  are the row and column numbers of a face image and  $K$  is the total number preferred roughly equally-sized superpixels. The value of  $K$  is assumed to be 100 for this work as per the method proposed in [1]. The spatial distance,  $d_{p_iq_j}^{xy}$ , between  $i^{th}$  superpixel in an image  $p$  to  $j^{th}$  superpixel in an image  $q$  is computed by (11).

$$d_{p_iq_j}^{xy} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{11}$$

where  $(x_i, y_i)$  and  $(x_j, y_j)$  are the spatial coordinates of  $i^{th}$  superpixel in an image  $p$  and  $j^{th}$  superpixel in an image  $q$  respectively. A variable  $m$  is used in  $d_{p_iq_j}^{xy}$ , which permit to regulate the compactness of a superpixel. When the value of  $m$  is large, the clusters are more compact with more spatial proximity. The range of  $m$  is [1 : 20] and the value of  $m$  is considered as 20 for this work. The gray proximity,  $d_{p_iq_j}^g$ , between the same two points would be computed using (12), which is an absolute difference of the intensity values of the centroid of  $i^{th}$  superpixel in an image  $p$ ,  $g_i$  and the centroid of  $i^{th}$  superpixel in an image  $q$ ,  $g_j$ .

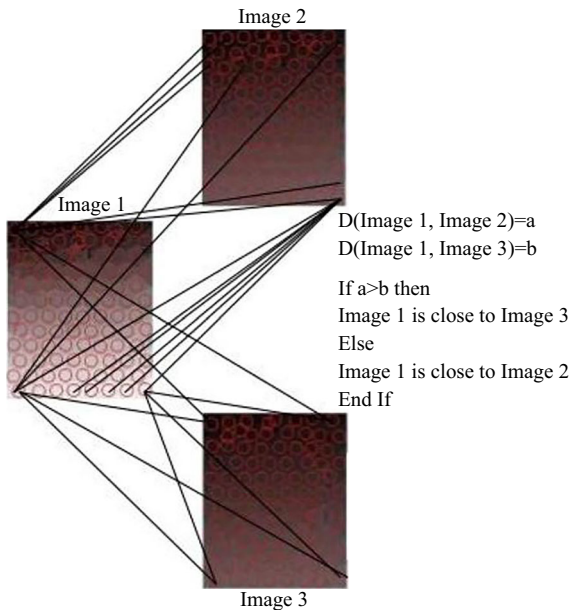
$$d_{p_iq_j}^g = (g_i - g_j) \tag{12}$$

In order to calculate  $d_{p_iq_j}^{ltp}$ , first two LTP codes of  $i^{th}$  superpixel in an image  $p$  and  $j^{th}$  superpixel in an image  $q$  would be computed using (8). Then bit-wise matching of these two ternary codes has to be performed using (13).

$$d_{p_iq_j}^{ltp} = \sum_{l=1}^8 (C_i(l) - C_j(l))^2, \tag{13}$$

where  $C_i$  and  $C_j$  are two ternary codes of  $i^{th}$  superpixel in an image  $p$  and  $j^{th}$  superpixel in an image  $q$  respectively. The  $l$  is used as an index variable and it varies from 0 to 8 since 8-bits are present in a ternary code. The reason behind using (13) is as follows: if a ternary bit changes from 1 to -1 or vice versa then  $d_{p_iq_j}^{ltp}$  contributes a value 4 to (9) as a high penalty. It means these ternary codes are not same. On the other hand, if a ternary bit changes from 0 to  $\pm 1$  or vice versa  $d_{p_iq_j}^{ltp}$  contributes comparatively a low-penalty value 1 to (9) because this changes might be due to noise. After designing the cost function for each and every edge in the bipartite graph using two images  $p$  and  $q$ , Algorithm 1 is used to find maximum matching. Here, a complete matching would be obtained since both the images have same number of superpixels i.e.  $K=100$ . After finding out a complete matching, similarity measure between two images has to be estimated using (14). Similarity measure





**Fig. 11** Illustration of the proposed similarity measure

between two images is the sum of all edges that help to form complete matching. Figure 11 illustrates to find the similarity between images.

$$D(p, q) = \sum_{i=1}^K \sum_{j=1}^K D_{p_i q_j} \quad (14)$$

## 5 Experimental results and discussion

All the experiments have been conducted on newly created UGC-JU face database [41], IRIS benchmark face database [12] and SCface face database [18] to evaluate the proposed system. In case of UGC-JU database, two types of sensors namely FLIR 7 sensor and Sony DSC-W350 digital sensor are used to capture IR and visible face images. Eighty four volunteers gave their concern to take images. Thirty nine different face images of a person with different pose changes (i.e. rotations about  $x$ -axis,  $y$ -axis, and  $z$ -axis) and different facial expressions (i.e. sad, angry, happy, fearful, disgusted, neutral, and surprised). All the captured images were 24-bit color images with resolutions of  $320 \times 240$  pixels. On the other hand, IRIS face database is a benchmark face database, which consists of visible and IR face images of 30 persons. This database also comprises of various facial expressions, pose and illumination changes. Moreover, SCface face database consists of visible and IR images of 130 subjects captured by different surveillance cameras at various distances. Twelve frontal face images of thirty volunteers are used from the UGC-JU face databases in this work. The frontal face images comprise of variation in facial expressions (UGC-JU face database), facial expressions and illumination changes (IRIS face database and SCface face database). The 12 frontal face images of 30 persons are used, which are selected randomly since IRIS

**Table 1** The ratio of spatial frequency error [14, 57] performance on UGC-JU face database, IRIS face database and SCface database

Database	UGC-JU face database	IRIS face database	SCface database
Minimum value	0.005	0.027	0.033
Average value	0.136	0.1140	0.1342
Maximum value	0.299	0.308	0.0371
Remark	Close to 0		

face database contains 30 images. On the other hand, twelve frontal face images of each subject are there in UGC-JU face database. Simultaneously, thirty subjects are selected from the SCface face database with twelve frontal face images, where the database contains twenty three frontal images of each person. Hence, to make uniformity of the proposed system, 12 frontal face images of 30 persons are used from these databases separately in this work. Some of the visible images and their corresponding IR images of three face databases are shown in Figs. 1, 2 and 3. Prior to the fusion process, some preprocessing steps have been considered to keep face part only by removing rest of the body parts since this work focuses in human authentication problem. The pre-processing steps involve binarization of an image, finding the largest component as a face skin region [39], and scaling of the face region to the size of  $112 \times 92$  pixels [39]. Then above mentioned image fusion method is applied on scaled visible and IR face images in order to get the fused face image. Some of the fused face images obtained by combining visible and IR face images of Figs. 1, 2 and 3 are shown in Fig. 10a, b and c respectively. Every time our bare eyes cannot assess the quality progress of the fused face images over the source images. Various fusion metrics are available in literature to judge automatically the improvement of the fused images with respect to one or more source image(s), which is called as reference image(s). Some of the popular fusion metrics, ratio of spatial frequency error (rSfE) [35, 57], normalized mutual information [27, 35], edge information [56], universal image quality index [21, 33], extended frequency comparison index [44] are used to prove the usefulness of the proposed fusion method. The detail discussion of each of the above mentioned fusion metric is there in [44]. Tables 1, 2, 3, 4 and 5 show the quantitative results of using rSfE, NMI, EI, UIQI and EFCI metrics respectively for the fused images obtained by combining visible and IR face images of UGC-JU, IRIS and SCface face databases. The minimum, average and maximum values are reported in all the above mentioned tables. The last row of each of the above tables show the ideal value of these metrics. It helps to judge how good the proposed fusion algorithm is. All the results depict the usefulness of the proposed fusion algorithm since these values are closed to ideal values mentioned in the last row of Tables 1, 2, 3, 4 and 5.

Previous experiment helps to know how effective the proposed fusion method is. Now, its time to know whether the fused face images assist in authentication process or not. First, one out of the twelve fused face images of each person is chosen randomly and is considered

**Table 2** The normalized mutual information [27, 35] performance on UGC-JU face database, IRIS face database and SCface database

Database	UGC-JU face database	IRIS face database	SCface database
Minimum value	2.06	2.01	1.87
Average value	2.76	3.02	2.35
Maximum value	3.51	3.87	2.73
Remark	Larger value corresponds to more information		

**Table 3** The edge information [56] performance on UGC-JU face database, IRIS face database and SCface database

Database	UGC-JU face database	IRIS face database	SCface database
Minimum value	0.41	0.46	0.37
Average value	0.59	0.59	0.61
Maximum value	0.81	0.78	0.76
Remark	0 to 1 and 1 is ideal		

as gallery image and rest of the fused face images are used as probe images. These gallery images will act as representative of their classes. Then a probe fused face image is selected at a time and tries to find out class belongingness of any of the gallery images using the proposed similarity measure. This process is continued till the probe images are exhausted. In this way, we can create an authentication model. Now, its time to evaluate the performance of this model by accuracy, precision and recall, because evaluating a model is an important task in biometric authentication system which delineates how good our predictions are. All these metrics rely on four terms namely, true positive (TP), true negative (TN), false positive (FP) and false negative (FN). These measures are stored in a 2 dimensional matrix, which is called as confusion matrix. Accuracy is one of the most intuitive performance metrics and it could be computed by (15).

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{15}$$

So, it is a ratio of the correctly predicted observation ( $TP + TN$ ) to the total observations ( $TP + FN + FP + TN$ ). Accuracy is good enough only when dataset is symmetric in nature, where false negatives and false positives are almost equal. It is not always true in practice. In that case even if accuracy is too high, the other metrics have to check in order to evaluate the performance of an authentication model. Precision is a ratio of the correctly predicted positive observations to the total predicted positive observations. It would be found by (16). High precision is always desired.

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

Recall is a ratio of the correctly predicted positive observations to all observations in actual class, which can be calculated using (17). The authentication model is good if the value of recall is greater than 0.5.

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

A comparative study has been presented in this section among the four state-of-the-art methods [5, 25, 47, 48] and the proposed method on IRIS benchmark face database,

**Table 4** The universal image quality index [21, 33] performance on UGC-JU face database, IRIS face database and SCface database

Database	UGC-JU face database	IRIS face database	SCface database
Minimum value	0.53	0.42	0.32
Average value	0.65	0.54	0.48
Maximum value	0.73	0.68	0.72
Remark	0 to 1 and 1 is ideal		

**Table 5** The extended frequency comparison index [44] performance on UGC-JU face database, IRIS face database and SCface database

Database	UGC-JU face database	IRIS face database	SCface database
Minimum value	0.62	0.61	0.46
Average value	0.70	0.71	0.63
Maximum value	0.82	0.79	0.78
Remark	-1 represents fused image is far from source images whereas 1 denotes fused image is very close to source images		

UGC-JU face database and SCface face database separately. Thirty experiments have been performed as 30 face images are there in each class. Every time 12 fused face images are selected randomly and used as gallery image and rest of the fused face images will act as probe image. Then the proposed similarity measure is used to estimate the class belongingness of each probe image. In each experiment, accuracy, precision and recall have been calculated and averages of these metrics are shown in Table 6. However, only pixel level fusion has been implemented because this work focuses on pixel level fusion only.

From Table 6, it is clear that the proposed method outperforms the state-of-the-art methods on IRIS benchmark face database, UGC-JU face database and SCface face database in terms of performance metrics namely accuracy, precision and recall. Moreover, in case of the proposed method, for each experiment, only one fused face image is misclassified for the UGC-JU face database, six fused face images are misclassified for the SCface face database and there is no misclassification event occurs for IRIS face database. The obtained results depict the usefulness of the proposed fusion scheme and similarity measure.

**Table 6** Performance comparison of the state-of-the-art methods on IRIS benchmark face database, UGC-JU face database and SCface database

		UGC-JU face database	IRIS face database	SCface database
S. G. Kong et al. [25]	Accuracy	81.51	82.42	80.56
	Precision	85.83	87.01	81.76
	Recall	87.71	87.77	82.34
G. Bebis et al. [5]	Accuracy	86.36	87.87	83.33
	Precision	89.13	89.36	85.83
	Recall	91.11	93.33	84.74
R. Singh et al. [47]	Accuracy	90.90	92.12	88.89
	Precision	91.86	90.83	87.17
	Recall	95.74	95.79	88.23
R. Singh et al. [48]	Accuracy	93.33	95.45	91.67
	Precision	98.21	97.89	91.11
	Recall	97.51	97.55	90.34
Proposed Method	Accuracy	<b>99.70</b>	100	98.46
	Precision	<b>99.24</b>	100	<b>98.05</b>
	Recall	99.24	<b>100</b>	97.69

## 6 Conclusion

In this work, a fusion method has been presented for fusing visible and IR face images based on AWT and FD using DBC method. The proposed fusion scheme preserves the temperature distribution of IR face images in fused face images along with contrast information of visible face images. The use of AWT is more beneficial over mallat wavelet transform as AWT is translation invariant in nature. On the other hand, FD measures image roughness in the form of image texture. Simple if-else rule has been defined to combine useful information of IR and visible face images in fused face images. Some of the popular fusion metrics have been considered to judge the usefulness of the fused face images quantitatively over source images namely, visible and IR face images. A similarity measure is also proposed based on superpixels and maximum matching bipartite algorithm. A distance measure is also introduced as a cost function of each edge in the bipartite graph. Then a complete matching has been obtained from the bipartite graph that would be considered as similarity measure between two fused face images. All the experiments have been performed on IRIS benchmark face database, UGC-JU face database and SCface face database. A comparative study has also been done using four state-of-the-art methods and the proposed method. Some of the available performance metrics such as accuracy, precision and recall have been adopted to evaluate the proposed model over existing systems. All the results depict that the proposed face authentication system outperforms four state-of-the-art methods in terms of used performance metrics namely, accuracy, precision and recall.

**Acknowledgments** Ayan Seal thank to Media Lab Asia, Ministry of Electronics and Information Technology, Government of India for providing young faculty research fellowship. Portions of the research in this paper use the SCface database of facial images. Credit is hereby given to the University of Zagreb, Faculty of Electrical Engineering and Computing for providing the database of facial images. We thank the anonymous reviewers for their many insightful comments and suggestions.

## References

1. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S (2010) SLIC Superpixels, EPFL technical Report 149300
2. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süssstrunk S (2012) SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell* 34(11):2274–2282
3. Adini Y, Moses Y, Ullman S (1997) Face recognition: the problem of compensating for changes in illumination direction. *IEEE Trans Pattern Anal Mach Intell* 19(7):721–732
4. Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: Application to face recognition. *IEEE Trans Pattern Anal Mach Intell* 28(12):2037–2041
5. Bebis G, Gyaourova A, Singh S, Pavlidis I (2006) Face recognition by fusing thermal infrared and visible imagery. *Image Vis Comput* 24(7):727–742
6. Ben-Arie J, Nandy D (1998) A volumetric/iconic frequency domain representation for objects with application for pose invariant face recognition. *IEEE Trans Pattern Anal Mach Intell* 20(5):449–457
7. Bhattacharjee D, Seal A, Ganguly S, Nasipuri M, Basu DK (2012) Comparative study of human thermal face recognition based on Haar wavelet transform and local binary pattern. *Comput Intell Neurosci* 2012:6
8. Bhowmik MK, Bhattacharjee D, Nasipuri M, Basu DK, Kundu M (2010) Fusion of wavelet coefficients from visual and thermal face images for human face recognition—a comparative study. arXiv:1007.0626
9. Bhowmik MK, Bhattacharjee D, Nasipuri M, Basu DK, Kundu M (2010) Image pixel fusion for human face recognition. arXiv:1007.0628
10. Chen J, Yi D, Yang J, Zhao G, Li SZ, Pietikainen M (2009) Learning mappings for face synthesis from near infrared to visual light images. *IEEE Conference on Computer Vision and Pattern Recognition*, pp 156–163

11. Cormen TH, Leiserson CE, Rivest RL, Stein C (2009) Introduction to algorithms, 3rd edit
12. Davis JW, Sharma V (2005) Fusion-based background-subtraction using contour saliency, IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops, pp 11
13. Dutilleux Pierre (1990) An implementation of the “algorithme à trous” to compute the wavelet transform, Wavelets, pp 298–304
14. Eskicioglu AM, Fisher PS (1995) Image quality measures and their performance. IEEE Trans Commun 43(12):2959–2965
15. Fowler JE (2005) The redundant discrete wavelet transform and additive noise. IEEE Signal Process Lett 12(9):629–632
16. Gonzalo C, Lillo-Saavedra M (2008) A directed search algorithm for setting the spectral–spatial quality trade-off of fused images by the wavelet à trous method. Can J Remote Sens 34(4):367–375
17. González-Audifána M, Otazu X, Fors O, Seco A (2005) Comparison between Mallat’s and the ‘à trous’ discrete wavelet transform based algorithms for the fusion of multispectral and panchromatic images. Int J Remote Sens 26(3):595–614
18. Grgic M, Delac K, Grgic S (2011) SCface–surveillance cameras face database. Multimed Tools Appl 51(3):863–879
19. Hall DL, Llinas J (1997) An introduction to multisensor data fusion. Proc IEEE 85(1):6–23
20. Hermosilla G, Gallardo F, Farias G, Martin CS (2015) Fusion of visible and thermal descriptors using genetic algorithms for face recognition systems. Sensors 15(8):17944–17962
21. Hossny M, Nahavandi S, Creighton D (2008) Comments on ‘Information measure for performance of image fusion’. Electron Lett 44(18):1066–1067
22. Jain AK, Ross A, Prabhakar S (2004) An introduction to biometric recognition. IEEE Trans Circuits Syst Video Technol 14(1):4–20
23. Keller JM, Chen S, Crownover RM (1989) Texture description and segmentation through fractal geometry. Computer Vision, Graphics, and image processing 45(2):150–166
24. Kokar M, Kim KH (1993) Review of multisensor data fusion architectures and techniques, IEEE International Symposium on Intelligent Control, pp 261–266
25. Kong SG, Heo J, Boughorbel F, Zheng Y, Abidi BR, Koschan A, Yi M, Abidi MA (2007) Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition. Int J Comput Vis 71(2):215–233
26. Lei Z, Pietikäinen M, Li SZ (2014) Learning discriminant face descriptor. IEEE Trans Pattern Anal Mach Intell 36(2):289–302
27. Li S, Kang X, Hu J (2013) Image fusion with guided filtering. IEEE Trans Image Process 22(7):2864–2875
28. Mandelbrot BB (1983) The fractal geometry of nature/Revised and enlarged edition. WH Freeman and Co., New York, p 495
29. Panigrahy C, Garcia-Pedrero A, Seal A, Rodríguez-Esparragón D, Mahato NK, Gonzalo-Martín C (2017) An Approximated Box Height for Differential-Box-Counting Method to Estimate Fractal Dimensions of Gray-Scale Images. Entropy 19(10):534
30. Pavlidis I, Symosek P (2000) The imaging issue in an automatic face/disguise detection system, Proceedings. In: IEEE workshop on computer vision beyond the visible spectrum: Methods and applications, pp 15–24
31. Peleg S, Naor J, Hartley R, Avnir D (1984) Multiple resolution texture analysis and classification. IEEE Trans Pattern Anal Mach Intell 6(4):518–523
32. Pentland AP (1984) Fractal-based description of natural scenes. IEEE Trans Pattern Anal Mach Intell 6(6):661–674
33. Piella G, Heijmans H (2003) A new quality metric for image fusion. International Conference on Image Processing 3:111–173
34. Qin X, Zheng J, Hu G, Wang J (2017) Multi-focus image fusion based on window empirical mode decomposition. Infrared Phys Technol 85:251–260
35. Qu G, Zhang D, Yan P (2002) Information measure for performance of image fusion. Electron Lett 38(7):313–315
36. Ren X, Malik J (2003) Learning a classification model for segmentation, pp 10
37. Rockinger O, Fechner T (1998) Pixel-level image fusion: the case of image sequences. Signal Processing, Sensor Fusion, and Target Recognition VII 3374:378–389
38. Sarkar N, Chaudhuri BB (1994) An efficient differential box-counting approach to compute fractal dimension of image. IEEE Trans Syst Man Cybern 24(1):115–120
39. Seal A, Bhattacharjee D, Nasipuri M (2016) Human face recognition using random forest based fusion of à-trous wavelet transform coefficients from thermal and visible images. AEU Int J Electron Commun 70(8):1041–1049

40. Seal A, Bhattacharjee D, Nasipuri M, Gonzalo-Martin C (2014) Robust thermal face recognition using region classifiers. *Int J Pattern Recognit Artif Intell* 28(5):1456008
41. Seal A, Bhattacharjee D, Nasipuri M, Basu DKR (2015) UGC-JU face database and its benchmarking using linear regression classifier. *Multimed Tools Appl* 74(9):2913–2937
42. Seal A, Bhattacharjee D, Nasipuri M, Gonzalo-Martin C, Menasalvas E (2014) Histogram of bunched intensity values based thermal face recognition. *Rough Sets and Intelligent Systems Paradigms*, Springer, pp 367–374
43. Seal A, Bhattacharjee D, Nasipuri M, Gonzalo-Martin C, Menasalvas E (2017) Fusion of visible and thermal images using a directed search method for face recognition. *Int J Pattern Recognit Artif Intell* 31(4):1756005
44. Seal A, Bhattacharjee D, Nasipuri M, Rodríguez-Esparragón D, Menasalvas E, Gonzalo-Martin C (2018) PET-CT image fusion using random forest and à-trous wavelet transform. *Int J Numer Methods Biomed Eng* 34(3):2933
45. Shensa MJ (1992) The discrete wavelet transform: wedding the a trous and Mallat algorithms. *IEEE Trans Signal Process* 40(10):2464–2482
46. Singh R, Vatsa M, Noore A (2008) Hierarchical fusion of multi-spectral face images for improved recognition performance. *Information Fusion* 9(2):200–210
47. Singh R, Vatsa M, Noore A (2008) Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition. *Pattern Recogn* 41(3):880–893
48. Singh R, Vatsa M, Noore A (2009) Face recognition with disguise and single gallery images. *Image Vis Comput* 27(3):245–257
49. Stramaglia S, Bassez I, Faes L, Marinazzo D (2017) Multiscale Granger causality analysis by à trous wavelet transform, 7th IEEE international workshop on advances in sensors and interfaces, pp 25–28
50. Tan X, Triggs B (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans Image Process* 19(6):1635–1650
51. Tazebay MV, Akansu AN (1995) Adaptive subband transforms in time-frequency excisers for DSSS communications systems. *IEEE Trans Signal Process* 43(11):2776–2782
52. Tian Y-I, Kanade T, Cohn JF (2001) Recognizing action units for facial expression analysis. *IEEE Trans Pattern Anal Mach Intell* 23(2):97–115
53. Wang N, Li Q, El-Latif AAAbd, Peng J, Niu X (2013) Multibiometric complex fusion for visible and thermal face images, *International Journal of Signal Processing. Image Processing and Pattern Recognition* 6(3):1–16
54. Wang Z, Bovik AC (2002) A universal image quality index, Wang, Zhou and Bovik, Alan C. *IEEE Signal Process Lett* 9(3):81–84
55. Xu Y, Li Z, Yang J, Zhang D (2017) A survey of dictionary learning algorithms for face recognition. *IEEE Access* 5:8502–8514
56. Xydeas CS, Petrovic V (2000) Objective image fusion performance measure. *Electron Lett* 36(4):308–309
57. Zheng Y, Elmaghraby AS, Frigui H (2006) Three-band MRI image fusion utilizing the wavelet-based method optimized with two quantitative fusion metrics, *Medical Imaging 2006: Image Processing*, vol 61440

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.





**Ayan Seal** received the PhD degree in engineering from Jadavpur University, West Bengal, India, in 2014. He is currently an Assistant Professor with the Computer Science and Engineering Department, PDPM Indian Institute of Information Technology, Design and Manufacturing Jabalpur, Madhya Pradesh, India. He has visited the Universidad Politecnica de Madrid, Spain as a visiting research scholar. He is the recipient of several awards. Recently, he has received Sir Visvesvaraya Young Faculty Research Fellowship from Media Lab Asia, Ministry of Electronics and Information Technology, Government of India. He has authored or co-authored of several journals, conferences and book chapters in the area of biometric and medical image processing. His current research interests include image processing and pattern recognition.



**Chinmaya Panigrahy** received the B.Tech and M.Tech degree in Computer Science and Engineering from B.P.U.T, Rourkela, Odisha, India and V.S.S.U.T., Burla, Odisha, India, in 2011 and 2015 respectively. He is currently an institute PhD research scholar with the Computer Science and Engineering Department, PDPM Indian Institute of Information Technology, Design and Manufacturing Jabalpur, Madhya Pradesh, India. He has visited the Universidad Politecnica de Madrid, Spain as a visiting research scholar. His current research interests include image processing and pattern recognition.