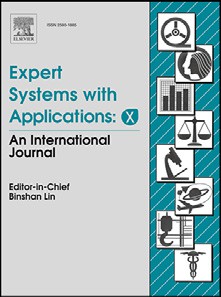
[Expert Systems with Applications: X 3 (2019) 100008](https://doi.org/10.1016/j.eswax.2019.100008)

Contents lists available at [ScienceDirect](http://www.ScienceDirect.com/)

Expert Systems with Applications: X

journal homepage: [www.elsevier.com/locate/eswax](http://www.elsevier.com/locate/eswax)

A geometric and fractional entropy-based method for family photo classification

Maryam Asadzadeh Kaljahi[a,](#_bookmark0) Palaiahnakote Shivakumara[a,](#_bookmark0)[∗,](#_bookmark3) Tianping Hu [b,](#_bookmark1) Hamid A. Jalab[a,](#_bookmark0)

Rabha W. Ibrahim[a,](#_bookmark0) Michael Blumenstein[c,](#_bookmark2) Tong Lu[b,](#_bookmark1) Mohamad Nizam Bin Ayub [a](#_bookmark0)

a *Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia*

b *National Key Lab for Novel Software Technology, Nanjing University, China*

c *Faculty of Engineering and IT, University of Technology Sydney, Australia*

a r t i c l e i n f o a b s t r a c t

*Article history:*

Received 11 November 2018

Revised 14 April 2019

Accepted 3 July 2019

Available online 5 July 2019

*Keywords:*

Face recognition Facial points

Facial geometric features Fractional entropy Convolutional neural networks Family photo classification

Due to the power and impact of social media, unsolved practical issues such as human traﬃcking, kinship recognition, and clustering family photos from large collections have recently received special attention from researchers. In this paper, we present a new idea for family and non-family photo classification. Unlike existing methods that explore face recognition and biometric features, the proposed method ex- plores the strengths of facial geometric features and texture given by a new fractional-entropy approach for classification. The geometric features include spatial and angle information of facial key points, which give spatial and directional coherence. The texture features extract regular patterns in images. The pro- posed method then combines the above properties in a new way for classifying family and non-family photos with the help of Convolutional Neural Networks (CNNs). Experimental results on our own as well as benchmark datasets show that the proposed approach outperforms the state-of-the-art methods in terms of classification rate.

© 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license. (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

### Introduction

The evolution of communication technologies, such as Face- book, Google+, Twitter, Instagram, Flicker and WhatsApp, help people to interconnect quickly ([Zhen, Caiming, & Caixian, 2018](#_bookmark56)). One such example is photo-sharing services for social networking. By taking advantage of the advancements in mobile digital cam- era technologies, people can easily take photos when they find something interesting and upload them to a social media plat- form to share exciting moments with their friends, families and colleagues ([Cai, Hio, Bermingham, Lee, & Lee, 2014](#_bookmark39)). As a result, one can expect large collections, which is evident, as the uploaded photo count was “about 4.5 million daily” according to the report in [Cai et al. (2014)](#_bookmark39). In addition, the development of multimedia technologies and cost effective CCTV cameras for surveillance ap- plications produce diversified images or videos at a larger scale.

∗ Corresponding author.

*E-mail addresses:* [asadzadeh@um.edu.my](mailto:asadzadeh@um.edu.my) (M.A. Kaljahi), [shiva@um.edu.my](mailto:shiva@um.edu.my) (P. Shivakumara), [htp@smail.nju.edu.cn](mailto:htp@smail.nju.edu.cn) (T. Hu), [hamidjalab@um.edu.my](mailto:hamidjalab@um.edu.my) (H.A. Jalab), [rabhaibrahim@um.edu.my](mailto:rabhaibrahim@um.edu.my) (R.W. Ibrahim), [michael.blumenstein@uts.edu.au](mailto:michael.blumenstein@uts.edu.au) (M. Blu- menstein), [lutong@nju.edu.cn](mailto:lutong@nju.edu.cn) (T. Lu), [nizam\_ayub@um.edu.my](mailto:nizam_ayub@um.edu.my) (M.N.B. Ayub).

This leads to a huge collection with a high degree of diversity and unstructured data ([Shen, Liu, Shih, & Hong, 2009](#_bookmark51)). For instance, some sample images of family and non-family photos chosen from our dataset are shown in [Fig. 1](#_bookmark4)(a) and (b), respectively, where we can see each image has its own variety of foreground (face regions) and background information. In this context, face recognition alone may be insuﬃcient to identify family or non-family photos. This is because the recognition methods developed might not work well for images which contain faces with multiple emotions, postures and actions. This makes the problem of finding photos that be- long to the same family complex and challenging. As a result, fam- ily photo classification/identification can play a vital role in find- ing a solution to unsolved issues such as human traﬃcking, kin- ship recognition, and the problem of identifying/locating refugees ([Robinson, Shao, Wu, Gillis, & Fu, 2018](#_bookmark52)). Hence, there is an urgent need for developing an intelligent expert system for tackling the above-mentioned challenges.

There are methods for identifying humans, facial expressions and emotions based on biometric features, which can be used for [family and non-family image identification (Haghighat, Zonouz, & Mottaleb, 2015; Mehta, Ramnani, & Singh, 2018). However, one ma-](#_bookmark44) jor challenge of biometric systems is the variability in character-

<https://doi.org/10.1016/j.eswax.2019.100008>

2590-1885/© 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license. (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)



**Fig. 1.** Samples images of family and non-family photos chosen from our dataset.

istics of the biometric of each individual. For example, the hu- man face is complex, with features that change over time. In addi- tion, facial features change due to variations in illumination, head pose, facial expression, cosmetics, aging, and occlusion because of beards or glasses ([Haghighat et al., 2015](#_bookmark44)). In addition, most of the methods require cropped face images for achieving better results ([Haghighat et al., 2015; Mehta et al., 2018](#_bookmark44)). Therefore, recognition- based systems may not be suitable for family and non-family photo classification because the images can have unconstrained back- grounds and multiple faces with numerous emotions or expres- sions ([Wang et al., 2017](#_bookmark55)). Hence, we can conclude that we need an expert and robust system that can cope with background com- plexities and issues of multiple faces with different emotions and expressions.

In this work, we propose to find a solution for family and non- family photo classification based on the characteristics defined be- [low for family and non-family images in Wang, Guo, Rohith, and Kambhamettu (2015, 2017).](#_bookmark57)

In the case of family photos, it is expected that

* Photos will have parents and their children either sitting or standing in a cascaded order. It should not contain persons of different families, namely, more than one family.
* The number of persons in an image should be more than 3, including parents and one child.
* Photos can be captured at both indoor and outdoor areas, such as houses, scenery, parks and tourist places with persons present. In other words, an image can have persons with any background.

In the case of non-family photos, it is expected that

* Photos must have persons with almost the same age, and it is expected persons of different families, for example, friends and colleagues might be present.
* The number of persons in an image should be 3 at a minimum.
* Images must have persons with different poses and any order with any background, which may include indoor and outdoor scenes.

### Related work

To overcome the limitations of recognition-based systems, methods which use unsupervised features such as clustering, grouping, and similarity between the parents and children’s faces, as well as personal traits such as age, race and gender ([Dandekar et al., 2014a](#_bookmark40)) have been developed.

[Ng, Zheng, Chan, and Yeung (2011)](#_bookmark49) proposed social relationship discovery and face annotations for personal photo collections. This method explores the combination of ensemble RBFNN with pair- wise social relationships as context for recognizing people. How- ever, the method requires face annotations and parameter tuning for social relationship identification. In addition, the focus of the method does not relate to family and non-family image classifica- tion; rather it explores general social relationships.

[Dandekar et al. (2014b)](#_bookmark41) proposed verification of family relation- ships from parents’ and children’s’ facial images. The method uses local binary pattern features and degree of similarity between the faces of children and parents. The method follows conventional feature extraction and classifiers for achieving results. However, the method is good for cropped face images but not those with multi- ple faces, emotions, expressions and complex backgrounds. In ad- dition, the main target of the method is to match childrens’ faces with parents’ faces but not finding group images.

[Xia, Pan, and Qin (2014)](#_bookmark58) proposed face clustering in a photo al- bum, where the method explores spectral features, similarity fea- tures, minimum cost flow and clustering. The proposed features are extracted from cropped face images. The main objective of the method is to find images which share the same faces. This idea is good for grouping personal collections but not family and non- family image classification.

[Qin, Tan, and Chen (2015)](#_bookmark50) proposed tri-subject kinship verifica- tion for understanding the core of a family. The method proposes a degree of similarity between children and parents, resulting in a triangular relationship. To achieve this, the method uses a rela- tive symmetric bilinear model for estimating similarity. To improve the results, the method takes spatial information into account. This method is good as long as the recognition approach provides suc- cessful results; however, recognition-based methods may not be

robust for the images affected by severe illumination, postures and actions.

[Dai, Carr, Sigal, and Hoiem (2015)](#_bookmark42) proposed family member identification from photo collections. The method explores an un- supervised EM joint inference algorithm with a probabilistic CRF. The proposed model identifies role assignments for all detected faces along with associated pairwise relationships between them. The performance of the proposed model depends on the success of face detection and recognition; however, the extracted biomet- ric features used to find relationships may not be suﬃciently ro- bust when images are exposed to an open environment. In addi- tion, the main target is to identify relationships between members of a family but the approach does not focus on family and non- family classification.

[Robinson et al. (2018)](#_bookmark52) proposed visual kinship recognition of families in the wild. This method explores deep learning for face verification, clustering and boosted baseline scores. The method involves multimodal labeling to optimize the annotation process. This includes information of faces and metadata collected from family photos. It is noted that although the method explores re- cent powerful deep learning approaches for kinship identification, it is still limited to family photos but not non-family photos.

[Wang et al. (2015, 2017)](#_bookmark57) proposed leveraging geometry and ap- pearance cues for recognizing family photos. The methods identify facial points for each face in an image. Based on facial points, the method constructs polygons to study geometric features of faces in the image. Due to the height difference of persons and the arrange- ment of faces in family and non-family images, the method gets different polygons to study geometric features. It estimates pair- wise relationships like kinship recognition, and generates a code- book using k-means clustering. Furthermore, the degree of similar- ity of each group is extracted for classifying family and non-family photos with the help of an SVM classifier. However, classification may not be accurate when the heights of persons in an image do not follow a hierarchical arrangement. In addition, one might ex- pect that non-family members could have the same arrangements and heights.

In light of the above discussions, we can assert that a few methods have addressed family and non-family photo classification or identification, but most of the methods focus on kinship recog- nition based on face detection and recognition. These methods may not work well for images where we can see faces with multi- ple emotions, postures and actions. The methods which addressed family and non-family classification explore only foreground in- formation (facial information) for achieving their results. This is good for images with simple backgrounds but not images that have complex backgrounds, where we can expect open scenes and out- door environments in the case of non-family photos. Therefore, we can conclude that there is a critical need for an accurate method to classify family and non-family photos.

Hence, we propose a novel method which explores the advan- tages of spatial and angle information of facial key points and frac- tional entropy features for classification of family and non-family images. As noted from related work, facial points and geometric features for faces play a vital role in identifying members of a fam- ily, including kinship/relationships ([Wang et al., 2017, 2015](#_bookmark55)). Moti- vated by this argument, we propose spatial and angle features in a new way to study geometric structures of faces, which captures the spatial and directional coherence of the face regions. Furthermore, to improve the discriminative power of the features, the propose method explores regular patterns in images. It is observed that in general, persons’ standing or sitting arrangements in family photos follow regular patterns such as particular orders, while non-family photos may not follow these. To extract such observations, we pro- pose a novel fractional entropy feature to study the texture of fa- cial regions as well as the background (other than facial region) of

images. The combination of spatial information, angles that extract the geometric structure of faces, and fractional entropy that ex- tracts the texture of facial and background regions, produces a fea- ture vector. Furthermore, the feature vector is passed to a Convo- lutional Neural Network (CNN) to overcome the above-mentioned challenges.

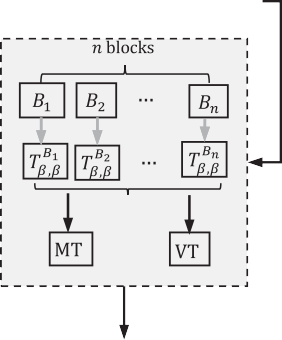
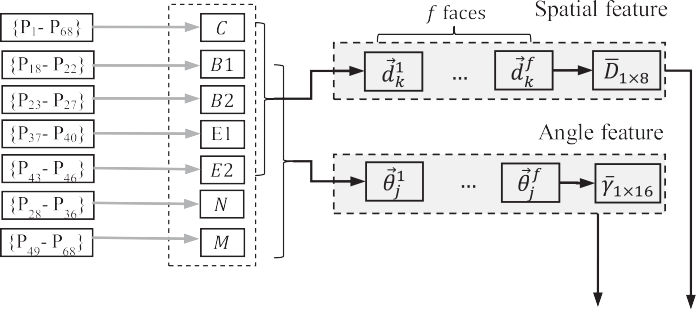
The contributions of this work are two-fold. (1) Exploring spa- tial and angle features for extracting the spatial and directional co- herence through the geometric structure of face regions. (2) Intro- ducing fractional entropy for extracting the texture of facial and background regions, which extracts regular patterns in the images.

### Proposed method

We noted from the Introduction and Related Work sections that facial features are important for discriminating between family and non-family persons. As a result, we propose to explore the same for finding facial key points (mouth, nose, left and right eyes and [eyebrows) for the input of family and non-family images (Ren, Cao, Wei, & Sun, 2014). The spatial relationship and angles between fa-](#_bookmark53) cial points provide unique cues for identifying a member of the same family or to distinguish between non-family members. Moti- vated by this observation, we propose to extract spatial and angle features for facial key points in a new way based on major and mi- nor axes. It is stated in [Wang et al. (2017)](#_bookmark55) that facial appearance in family images has a high degree of similarity with the unique pattern of spatial arrangement of persons (regular patterns), while in the case of non-family, one cannot expect such a high degree of similarity between faces and regular patterns in arranging persons (irregular patterns due to randomness in the ordering of persons). To extract such an observation, we propose to estimate the dis- tance between facial key points with respect to major and minor axes of the respective face images, which gives spatial coherence. In the same way, we also estimate the angle between facial key points of the respective face images, which gives directional coher- ence. Spatial and directional coherence together extracts geomet- ric properties of face images. However, the geometric features are limited to facial regions. In order to extract regular patterns from both the foreground and background (other than face regions), we further explore fractional entropy which extracts texture proper- ties in the regions. In this way, the proposed method combines the strengths of geometric features and fractional entropy for classify- ing family and non-family images successfully.

The proposed method extracts 8 distances and 24 angle fea-

tures using facial key points and two features from fractional en- tropy for face regions and background information (other than face regions). Therefore, for each input image, it gives a feature vec- tor containing 26 features (8 + 16 + 2). Furthermore, the feature vector is fed to a Convolutional Neural Network (CNN) for classi- fication ([McAllister, Zheng, Bond, & Moorhead, 2016](#_bookmark45)). The overall steps of the proposed method are shown in [Fig. 2](#_bookmark5). In [Fig. 2](#_bookmark5), P1 to [P68 are the points given by the face detection method (Ren et al., 2014), and based on those points, the same method detects five fa-](#_bookmark53) cial key points, namely, left Eyebrow (*B1*), right Eyebrow (*B2*), left Eye (*E1*), right Eye (*E2*), Nose (*N*), Mouth (*M*) and the centroid, us- ing all the 68 points. The distances are estimated between facial key points (d) for each face and finally the proposed method com- putes the mean of all the 8 features of all the faces (*f*) in the image (*D*), which gives a vector of 8 features. Similarly, angles (*θ* ) are es- timated between the facial points, and we compute the mean of all the angles of all the faces in image (*γ* ), which gives a vector of 16 features. For faces and background regions, which are other than face regions, the proposed method extracts fractional entropy for each non-overlapping block (*B*). The mean (*MT*) and variance (*VT*) of the fractional entropy of all the blocks are considered as a feature vector containing features.



**Fig. 2.** Flow of the proposed method.

The above observations are illustrated in [Fig. 3](#_bookmark8), where we draw line graphs for distance/angle features *vs.* variances of dis- tance/angle values for family and non-family images shown in [Fig. 3](#_bookmark8)(a). It is noticed from [Fig. 3](#_bookmark8)(b) and (c) that the line behavior which represents families is smoother than that representing non- family for both spatial and angle features. This confirms that the appearance of faces in a family image does not have many varia- tions, while non-family have high variations. The same conclusion can be drawn from the illustration of the angle feature shown in [Fig. 3](#_bookmark8)(c). This motivates us to use spatial and angle features for family and non-family image classification. Detailed explanations for each step of the spatial and angle feature extraction process is discussed in subsequent sections.

* 1. *Geometric features for facial key points*

For a given input image, the proposed method uses face align- ment via regression of local binary features for detecting facial key points, namely, mouth, noise, left and right eyes, and eyebrows ([Ren et al., 2014](#_bookmark53)). The method basically proposes a better learning-

in [Ren et al. (2014)](#_bookmark53). The reason to choose this method is that it is said to be generic, eﬃcient and accurate for finding facial key points. In addition, it can cope with issues of partial occlusion and distortion. This is justifiable because the proposed work considers family and non-family images with complex backgrounds and di- versified content. The sample results of the above method are il- lustrated in [Fig. 4](#_bookmark10), where (a) gives the results of candidate point detection for the input image, while [Fig. 4](#_bookmark10)(b) shows samples of fa- cial key points for family and non-family images. It is noted from [Fig. 4](#_bookmark10)(b) that although the images are affected by distortion and poor quality, the method finds key points successfully.

Let *B*1 , *B*2 , *E*1 , *E*2 , *N* and *M* be center points given by the method ([Ren et al., 2014](#_bookmark53)), which denote left and right eyebrows, eyes, nose and mouth, respectively. These points are marked manually in [Fig. 5](#_bookmark11) for the family and non-family faces chosen from the im- ages shown in [Fig. 4](#_bookmark10). To extract spatial features to study geometric characteristics, the proposed method finds the centroid using can- didate points of the face region as defined in [Eq. (1)](#_bookmark6), where *m* is [the number of candidate points given by the method (Ren et al.,](#_bookmark53)

[2014).](#_bookmark53)

based approach. It works based on learning with a “locality” prin- ciple. The principle is defined as: for locating a certain landmark at a given stage, the most discriminative texture information lies

*(XC , YC )* =

*i*=1 *, i*=1

(1)

µ Σ*m Xi*

*m*

Σ*m Yi* ¶

*m*

in a local region around the estimated landmark from the previ- ous stage. Shape context, which gives locations of other landmarks and local textures of this landmark, provides suﬃcient informa- tion. With these observations, the method first learns intrinsic fea- tures to encode local textures for each landmark independently; it then performs joint regression to incorporate shape context. The method first learns a local feature mapping function to generate local binary features for each landmark. Here, it uses a standard re- gression random forest to learn each local mapping function. Then it concatenates all the local features to obtain the mapping func- tions. It learns linear projections by linear regression. This learn- ing process is repeated stage by stage in a cascaded fashion. After that, a global feature mapping function, and a global linear projec- tion and objective function are used to incorporate shape context. This process can effectively enforce the global shape constraint to reduce local errors. In the case of the testing phase, shape incre- ment is directly predicted and applied to update the current esti- mated shape. More details regarding implementation can be found

With the help of the centroid (*XC* , *YC* ), the proposed method

draws an ellipse to find the major and minor axis as shown in [Fig. 6](#_bookmark19)(a) and (b) for family and non-family faces, respectively. The proposed method moves in a perpendicular direction to each key facial point (*B*1 , *B*2 , *E*1 , *E*2 ) of family and non-family images un- til it reaches pixels of the major axis as shown in the second il- lustration in [Fig. 6](#_bookmark19)(a) and (b), respectively. Similarly, the proposed method moves in a perpendicular direction to each key point of

family and non-family images until it reaches pixels of the mi- nor axis as shown in the third illustration in [Fig. 6](#_bookmark19)(a) and (b), re- spectively. Then the method finds the distance between facial key points *r* = *{B*1 , *B*2 , *E*1 , *E*2 } and respective pixels of the major and

minor axes in *r*r = {major, minor} defined in [Eq. (2)](#_bookmark7), which outputs

8 distances *dk, k* = {1,2,…,8}for each face *i*:

*d* = ,*(Xr* — *Xr*r *)* + *(Yr* — *Yr*r *)* (2)

*i* 2 2

*k*

The distance features are extracted with respect to the major and minor axes to make the features robust to different rotations.

|  |
| --- |
|  |
|  |
|  |

**Fig. 3.** Cues for extracting geometrical features (Spatial + Angle) for family and non-family photo classification.

In other words, if the input image is rotated in different directions, the feature still works well. For this step, we consider only four facial key points (that is, *B*1 , *B*2 , *E*1 and *E*2 ) for distance calcula- tion because Mouth (*M*) and Nose (*N*) do not contribute much to

as defined in [Eq. (3)](#_bookmark9), resulting in an average distance vector *D*¯ for each input image, where *f* is the number of faces:

Σ *f di*

classification because the *M* and *N* points lie on the minor axis. Note that the perpendicular distance is calculated by finding the

*D*¯  *i*=1 *k*

*f*

=

(3)

smallest distance between facial key points and the pixels of ma- jor/minor axes. The step finds many distances by considering a few left and right pixels of major and minor axes to the key points. Then it chooses the pixel which produces the smallest distance be- tween the pixels of the major/minor axis and key points. We be- lieve that the smallest distance is the same as the perpendicular distance. Since the input image contains many faces and the num- ber of faces is not predictable, the proposed method computes the mean of the 8 respective distances *d* of all faces in the input image

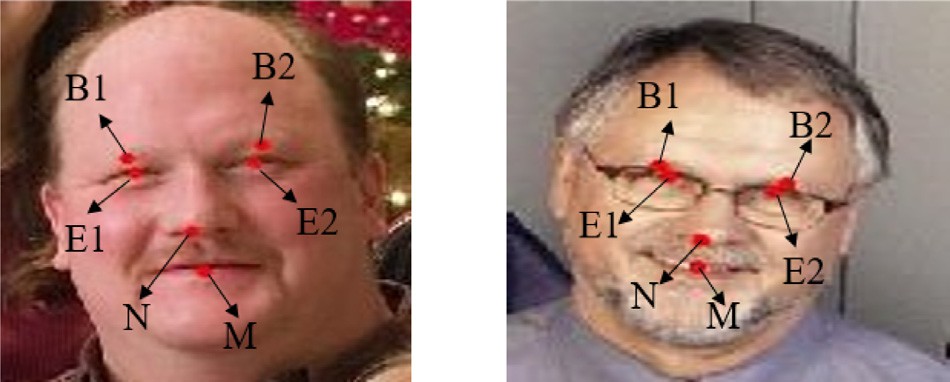


To make the geometric features robust, we also propose to cal-

culate the angles between facial key points to study the structure of the face region. This is because, as the face shape changes, the angle between facial key points also changes. To extract such ob- servations, we construct a rectangle using *B1 –B2 –E1 –E2* as shown in the first image in [Fig. 6](#_bookmark19)(c), which gives four angles. In the same way, the proposed method forms triangles using *B1–B2–N, B1 –B2 –M, E1 –E2 –N, E1 –E2 –M* as shown respectively in [Fig. 6](#_bookmark19)(c), which gives twelve angles. In total, the proposed method obtains 8



**Fig. 4.** Facial key point detection for family and non-family images ([Ren et al., 2014](#_bookmark53)).

cos *γj*

*xJ*

= *v ,* sin *γj*

*J*

*yj*

= *vJ*

(8)

*γ Arctan*  sin *γj* (9)

*j* cos *γ*

= µ  ¶

*j*

**Fig. 5.** Labelling six facial key points of the family and non-family images marked in [Fig. 4](#_bookmark10).

spatial + 16 angles = 24 geometric features for family and non- family image classification.

Let *A*(*xA, yA* ), *B*(*xB, yB* ), *C*(*xC, yC* ) be the coordinates of the ABC triangle. The inner angles *B*ˆ for the ABC triangle can be calculated

as defined in [Eqs. (4)](#_bookmark14) and [(5)](#_bookmark15). [Eq. (4)](#_bookmark14) computes a vector between B and A called —*A*→*B* and the vector between C and B similarly called *C*—→*B.* Angle *θ B* is driven by [Eq. (5)](#_bookmark15) by computing the four-quadrant

inverse tangent, where *xAB yAB* is determinant, while —*A*→*B.C*—→*B* is

| |

*xAB yAB*

the scalar dot product of the two vectors. Similarly, the proposed method estimates angles for the rectangle and the other triangles in this work.

—*A*→*B* = *B* — *A, C*—→*B* = *C* — *B* (4)

The proposed method computes the mean of distances to ex- tract spatial features and the mean of angles for extracting angle features for each image. The reason for computing the average is to widen the difference between family and non-family images. As discussed in the Introduction Section, family images have persons with almost the same facial appearance, while non-family images have persons with different facial appearances. This is valid be- cause one can expect a high degree of similarity between the ap- pearances of faces from the same family. It may not be true for non-family images. In addition, family and non-family images can have any number of faces, which should be more than 3 persons in the images. In this situation, the average features for a family does not make much difference, while for non-family, the average makes a vast difference. Since the appearance of faces in a family have a high degree of similarity compared to those in non-family images, it is expected that the average gives almost the same val- ues for family images while for non-family, we cannot predict the same values always. Besides, to make the spatial and angle features invariant to the number of faces, the proposed method considers the average for achieving better results.

*θB* = *Arctan*2

µ¯*xAB*

*yAB* ¯*,* —*A*→*B.C*—→*B*¶

# (5)

* 1. *Fractional entropy feature extraction*

*xAB yAB*

¯ ¯

Since we can expect many faces in a single input image, we propose to consider the average of the angles of the respective 16 angles. In order to average the respective angles of *f* faces, the cir- cular mean is computed. First, since the angles {*θ*1*, θ*2*, . . . , θj* }, *j* = 16 are defined on a circular coordinate system, the coordi- nate system should be changed to a rectangular one according to [Eq. (6)](#_bookmark16), where *θ i* is the *j*th angle *θ* of the *i*th face in the image.

*j*

Afterwards, *vj* the resultant vector and its direction are calculated as defined in [Eqs. (7)](#_bookmark17) and [(8)](#_bookmark12), respectively. Finally, *γj* mean of the *j*th angle for all the *f* faces is computed as defined in [Eq. (9)](#_bookmark13).

As mentioned in the Introduction Section, it is found that the other than face region also provides cues for discriminating fam- ily and non-family images. However, the previous step does not explore other than face region. Therefore, inspired by the method in [Ibrahim, Moghaddasi, Hamid, and Noor (2015)](#_bookmark46) where fractional calculus has been used for studying texture in splicing images, this section explores a new Tsallis fractional entropy-based texture ([Tsallis et al., 2009](#_bookmark54)) for studying variations in background as well as facial regions in family and non-family images. An overview of the Tsallis fractional entropy is presented in the following.

The Tsallis fractional entropy ([Tsallis et al., 2009](#_bookmark54)) measures the amount of uncertainty acting in the valuation of a random variable

Σ *f* cos *θ i*

=

*x*

*j*

*i*=1 *j , y*

*f*

Σ *f* sin *θ i*

or the consequence of a random process. The general discrete form

of this entropy is given in [Eq. (10)](#_bookmark18).

*v* = ,

=

*j*

*i*=1 *j*

*f*

(6)

*j*

*j*

*j*

*x* 2 + *y* 2 (7)

*T (ρ)(x)* 1

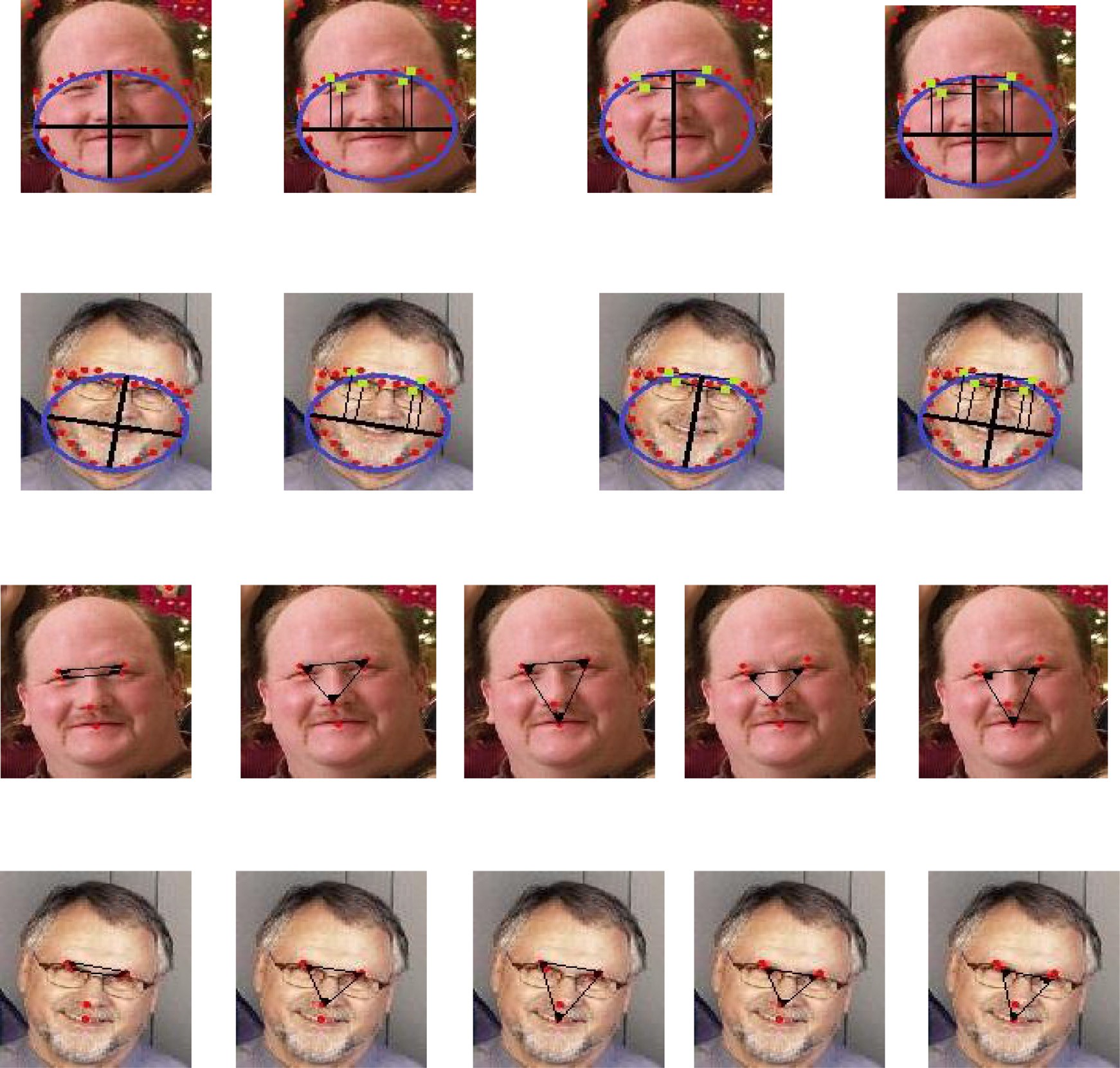
*β* 1

=

*β*

Ã1 — Σ *ρiβ (x)*!*,* (10)

*i*







**Fig. 6.** Geometric features using facial points for family and non-family image discrimination.

where *ρ* is the *q*-Gaussian probability of pixel *x*, q /= 1 and In our discussion, let *β*=*q*, then we conclude

!

*β* /= 1 are the fractional powers of the entropy, and the quantity 1/(*β* — 1) is the capacity of the image. The q-Gaussian is a proba- bility distribution ascending from the growth of the Tsallis entropy under suitable restrictions. It has the formal function as defined in

,

*Tβ,β (x)* =

*β*

C*β (β* — 1*)*

Ã1 —

Σ*i*=1

*Z* £

1 + *(*1 — *β)*¡—*βxi*2

*β*

¢¤

1—*β*

# (15)

[Eqs. (11)–(13)](#_bookmark21), where C*q* is a normalization factor. where *Z* is the total number of pixels in the image. The proposed

,

*β* ¢

*ρ (x)* = *e* —*βx*2 *,* (11)

C

*q*

¡

*q*

method calculates the above Tsallis fractional entropy based on fre- quency details of the input image, which gives a texture property to study the structure of it. The advantage of Tsallis fractional en-

¡ *β* ¢ £

2

2

¡ *β* ¢¤ 1

tropy is that it is sensitive to non-textured regions (low frequency).

*eq* — *x* =

√*π*

1 + *(*1 — *q)* — *x*

*Г*  3—*q*

¡ ¢

2*(q*—1*)*

¡ ¢

1—*q , q* /= 1*,* (12)

In addition, it sharpens any changes in texture details in the re- gions, where pixel values are changing sharply (high frequency). The sample illustration for Tsallis fractional entropy for family and

C*q* = ,*q* 1

—

*Г*  1 *,* (13)

*q*—1

non-family images is shown in [Fig. 7](#_bookmark22), where we can see all the dominant information represented by edges in the background and

Since the variable is the pixel which has a positive value in the

maximum entropy procedure, the q-exponential distribution is de- rived. Applying [Eqs. (11)–(13)](#_bookmark21) in [(10)](#_bookmark18), we have the following gen- eralized formula of the fractional entropy:

,

*Z*

¢¤  !

the facial regions are highlighted. [Fig. 8](#_bookmark23) shows the clear discrimi- nating power of Tsallis fractional entropy texture features for fam- ily and non-family images. Therefore, for the feature matrix given

by the Tsallis fractional entropy texture, we first split the input im-

*Tβ,q*

*(x)* =

*β*

C*q (β* — 1*)*

Ã1 —

Σ*i*=1

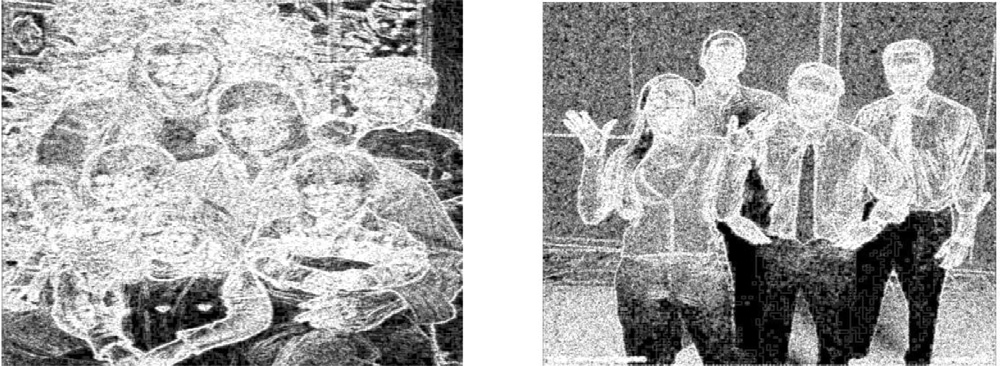
£1 + *(*1 — *q)*¡—*βxi*2

*β*

1—*q*

# (14)

age into blocks with a size of *a*×*a* pixels, then the Tsallis fractional entropy for each block is computed. For all the blocks of the input image, the “mean” and the “variance” are computed and saved as



**Fig. 7.** Fractional entropy features for family and non-family images.

a large number of samples for training and labeling samples, we prefer to use the combination of the proposed features and the CNNs rather than raw pixels with the recent deep learning mod- els. The main objective of the proposed work is to propose fea- tures that can classify the family and non-family photos. Thus, the proposed features are fed to a pre-defined CNN classifier which is available online ([Arora & Suman, 2012](#_bookmark38)) for classification in this work. For learning parameters of the classifier, we follow a 10-fold cross-validation procedure, which splits the dataset into training and testing components. The training samples are used for learn- ing and adjusting the parameters of the classifier and the testing samples are used for evaluation. The complete algorithmic steps of the proposed method for classifying family and non-family images are presented below.

**Algorithm: Feature Extraction for the Proposed Method**

10: Initialization: *I*=Input image, *m* ={1,…,68}, set of points given by [Ren et al. (2014)](#_bookmark53)

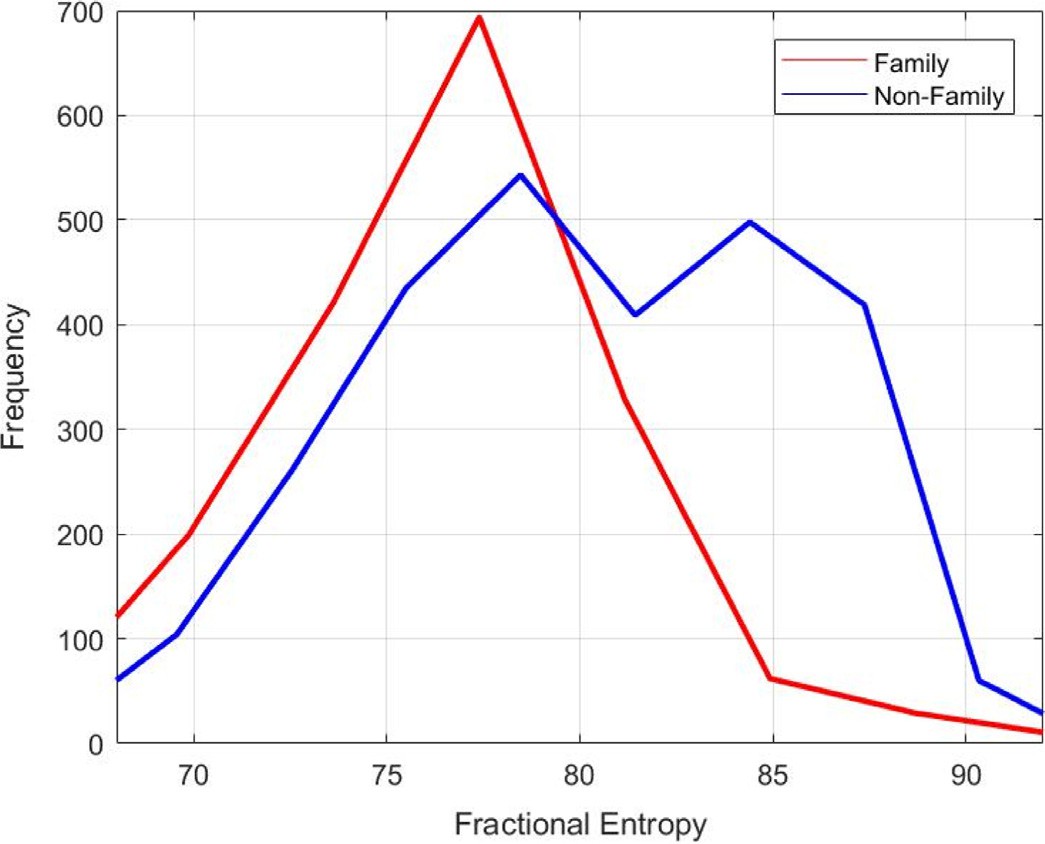
11: *f* ← Number of faces 12: **For** *i*=1 to *f* **do**

13: (X*Pm ,* Y*Pm* )← facial points

14: *B*1*, B*2*, E*1*, E*2*, N, M, C* ← Key facial points, which includes eyebrows, eyes, nose, mouth and center of the all the key points, respectively.

{ }

15: (*Ama jor* , *Aminor* ) ← (*xc, yc*) as defined in [Eq. (1)](#_bookmark6)

16:

*d→i* ← {*B*1*, B*2*, E*1*, E*2} &{*Ama jor* , *Aminor* }:*d→i*

as defined in [Eq. (2)](#_bookmark7)

17:

*k*

*θ→i* ← {*B*1*, B*2*, E*1*, E*2*, N, M*}:*θ→i*

*(*1×8*)*

as defined in [Eqs. (4)](#_bookmark14) and [(5)](#_bookmark15)

*j*

18: End For

*(*1×16*)*

19:

20:

*D*¯ *k* ←mean (*d→ )* : ***D→*** *(*1×8*)* as defined in [Eq. (3)](#_bookmark9)

*γ*¯ *j* ← circular mean*(θ→j )* : ***→γ****(*1×16*)* as defined is [Eqs. (6)–(9)](#_bookmark16)

*k*

**Fig. 8.** Histogram of Fractional entropy features for family and non-family images. .

the output texture features *MT* and *VT*, respectively. The pseudo- code for the proposed Tsallis fractional entropy algorithm is de- scribed as follows:

21: *MT* , *V T* ←Fractional Entropy feature extraction as defined in the above algorithm

22: —*f*—*e*—*a*—*tu*—→*re(*1×26*)* ← *D*¯ || *γ*¯ || *MT* || *V T* // Final feature vector having

dimension, 1×26.

23: CNN classification //Classification of Family and Non-family photos.

### Experimental results

For experimentation, we created our own dataset by collect-

ing images from social media, such as Facebook, Flickr, Instagram

**Algorithm: Fractional Entropy Feature Extraction**

1: Initialization: *I*=Input image, *a*=3; *β* = 0.5 2: **For** each Input image I **do**

3: **{***B1****,*** *B2, …, Bn***}**←split I into n blocks size of *a*×*a* pixels 4: **For** *i*=1 to *n* **do**

and from our own camera. This dataset includes indoor/outdoor scenes and images with 3–25 people. In addition, the dataset in- volves family and non-family photos of different cultures, such as Hindu and Chinese, and modern styles of family/non-family pho- tos. This makes the dataset challenging for experimentation. For

5: *Tβ,β (Bi*

3 3

×

*)* ← *I* // Fractional entropy is calculated as defined in

labeling the data as either family or non-family, we followed the

[Eq. (15)](#_bookmark20), where *i* denotes the *i*th block of 3×3 dimension.

6: **End For**

7: *MT* ← mean (*TBi* ), *i*={1,2,…*n*}// Mean of Fractional entropy of all (*n*) blocks

*β,β*

8: *V T* ←variance (*TBi* ), *i*={1,2,…*n*}) //Variance of Fractional entropy of

*β,β*

all (*n*) blocks 9: **End For**

Feature distributions of the spatial, angle and texture features for family and non-family images are shown in [Fig. 9](#_bookmark24)(a)–(c), re- spectively, where one can see that the feature distributions of geo- metric and Tsallis fraction entropy provides a clear distinction be- tween family and non-family images in terms of histogram behav- ior.

The concatenated features are then passed to a fully connected Convolutional Neural Network (CNN) for classifying family and non-family images ([McAllister et al., 2016](#_bookmark45)). Inspired by the method ([Nanni, Chidoni, & Brahnam, 2018](#_bookmark47)), where it is mentioned that the combination of handcrafted features and the ensemble of CNNs give better results than deep learning tools such as GoogleNet, ResNet50 that use raw pixels of the input images for bioimage clas- sification, we explore the same idea of combining the proposed features with the CNN for family and non-family image classifi- cation in this work. Since the proposed work does not provide

[instructions suggested in](#_bookmark55) [Gallagher et al. (2009)](#_bookmark43)[, Wang et al. (2017, 2015). Furthermore, the dataset includes one photo for one family.](#_bookmark55) In other words, the dataset does not have multiple photos of the same family. In total, our dataset consists of 388 family images and 382 non-family images, which gives a total of 770 images.

To demonstrate that the proposed method is effective, we also considered the benchmark dataset collected from publicly available data in [Gallagher et al. (2009)](#_bookmark43), [Wang et al. (2017, 2015)](#_bookmark55). This pub- lic data provides a large number of images, which include many groups of photos and images containing both family and non- family categories. As a result, we chose the relevant family and non-family images and labeled these manually according to the instructions in [Gallagher et al. (2009)](#_bookmark43), [Wang et al. (2017, 2015)](#_bookmark55). We consider this dataset as the benchmark dataset, which con- sists of 1790 family and 2753 non-family images. In total, there are 4543 images, which is larger than the dataset considered in [Haghighat et al. (2015)](#_bookmark44), [Mehta et al. (2018)](#_bookmark48). Overall, we considered 5263 (770 from our dataset and 4543 from benchmark dataset) images for experimentation in this work. Sample images of fam- ily and non-family photos for ours and the benchmark dataset are shown in [Fig. 10](#_bookmark25)(a) and (b), respectively, where we can see intra- and inter-class variations. It is also observed from [Fig. 10](#_bookmark25) that



**Fig. 9.** The proposed feature distribution for family and non-family images classification.



**Fig. 10.** Sample images of our dataset and the benchmark dataset ([Gallagher et al, 2009](#_bookmark43)).

family and non-family images have both indoor and outdoor scenes as backgrounds. It is also true that height distribution of persons in a hierarchical order for family and a non-hierarchical or- der for non-family is not necessarily true as shown in [Fig. 10](#_bookmark25). The detailed statistics of ours and the benchmark dataset are listed in [Table 1](#_bookmark27), where we calculate the ratios (E1 and E2) as respectively defined in [Eqs. (16)](#_bookmark26) and [(17)](#_bookmark28) using the count images with indoor

height orders. The ratio in [Table 1](#_bookmark27) indicates that our dataset is much more complex than the benchmark dataset because the ratio with respect to indoor backgrounds and the hierarchical order of our dataset are greater than those of the benchmark dataset. Note that in [Eqs. (16)](#_bookmark26) and [(17)](#_bookmark28), *total* denotes the size of the dataset as given in [Table 1](#_bookmark27) in the bracket.

and outdoor scenes, and hierarchical or non-hierarchical persons’

*E Outdoor( f amily)* + *Indoor(non* — *f amily)*

1 *total*

=

# (16)

**Table 1**

Statistics for ours and the benchmark dataset for family and non-family image classification.

Dataset Family Non-Family

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | (Total) | Indoor | Outdoor | Hierarchical | Non-hierarchical |  | Indoor | Outdoor | Hierarchical | Non-hierarchical | *E*1 | *E*2 |  |
| Our (770) | 255 | 133 | 273 | 115 |  | 201 | 182 | 110 | 272 | 43.37 | 29.22 |  |
|  | Benchmark (4543) | 1172 | 618 | 1378 | 412 |  | 924 | 1829 | 513 | 2240 | 33.94 | 20.36 |  |

## *E Non* — *Hierarchical( f amily)* + *Hierarchical(non* — *f amily)*

=

2 *total*

# (17)

To show that the proposed method is superior in comparison to existing methods, we implemented two state-of-the-art meth- ods, namely, [Wang et al. (2015)](#_bookmark57), which explores facial geometric features and facial appearance model-based features. The features are passed to an SVM classifier for family and non-family image classification. Please note, the same idea is extended and the re- sults are improved in [Wang et al. (2017)](#_bookmark55) for the purpose of family and non-family image classification. However, both the ideas fo- cused only on facial regions for achieving results; these also ig- nored background clues.

To measure the performance of the proposed and existing methods, we generate confusion matrices for family and non- family classification and the classification rate. The Classification Rate (CR) is defined as the number of images classified correctly by the proposed method (*R*) divided by the total number of im- ages in the class (*MG*) as defined in [Eq. (18)](#_bookmark29). The Average Classifi- cation Rate (ACR) is calculated for diagonal elements of the confu- sion matrices to evaluate the overall performance of the proposed and existing methods.

In this work, we undertake 10-fold cross-validation for choosing the number of training and testing samples. The criteria divides the whole dataset into 10 equal-sized sub-folds. For each iteration, im- ages from each sub-fold are considered as testing samples, while images from the other sub-folds are considered as training sam- ples for classification, which results in a confusion matrix for one sub-testing fold out of 10 sub-folds. This process indicates that the chosen training samples are used for training the classifier and the testing samples are used for evaluation. In this way, the process considers every sub-fold as testing samples at each iteration, which results in 10 confusion matrices i.e. 10-fold. The average of all the 10 confusion matrices are considered as the final confusion matrix for evaluation in this work.

*R*

*CR* = *MG* (18)

* 1. *Evaluating the proposed classification*

The proposed method consists of three key steps, namely, extracting spatial/angle-based geometric features and fractional entropy-based texture features for classifying family and non- family images. In order to assess the contribution of each key step, we conducted experiments on both our dataset and the benchmark dataset individually to calculate average classification rates. The re- sults reported in [Table 2](#_bookmark30) show that the combined Spatial + Angle achieves the best results compared to the individual features for both our dataset and the benchmark dataset. It is also noted from [Table 2](#_bookmark30) that the ACR of angle-based features is better than Spa- tial, but lower than Spatial + Angle for both the two datasets. This shows that angle-based features are better than spatial-based fea- tures, and the combination is better than both individual features. This is understandable because the spatial structure alone is not suﬃcient for handling the problem of complex backgrounds as it

only extracts 8 features. However, the improvement is marginally different. Therefore, we can conclude that spatial and angle-based features contribute equally for achieving the best results.

In this work, we extracted fractional entropy-based texture features for the whole image, which includes facial regions and background information. We conducted experiments for calculat- ing classification rates only for the Facial region (FEF), Background (FEB), and the whole image (FEW) to identify the effectiveness of the facial region and background information, individually. Note: the facial regions detected by facial point detection are consid- ered as foreground, and the rest of the region is considered as the background for experimentation. The results of FEF, FEB and FEW are reported for both our dataset and the benchmark dataset in [Table 3](#_bookmark31). It is observed from [Table 3](#_bookmark31) that the FEF is the best at ACR compared to FEB for both the datasets. This shows that fa- cial regions contribute more compared to the background. This is justifiable because sometimes, family and non-family photos may share the properties of the background. It is evident from the re- sults of FEB for the benchmark dataset in [Table 3](#_bookmark31), where most fam- ily images are misclassified as non-family. This shows that the fea- tures of the background of family images overlap with the features of the background of non-family images. However, facial regions alone are not suﬃcient to achieve the best ACR compared to FEW. Therefore, we can conclude that the features of the foreground and background are important to achieve the best results for classifica- tion.

It is noted from [Tables 2](#_bookmark30) and [3](#_bookmark31) that Spatial + Angle and FEW are

better compared to individual features on both our dataset and the benchmark datasets for family and non-family image classification. In order to decide the best combination, we conducted experi- ments for the following combinations: Spatial + FEF, Spatial + FEB, Spatial + FEW, as reported in [Table 4](#_bookmark32) and Angle + FEF, Angle + FEB and Angle + FEW, as reported in [Table 5](#_bookmark33). When we look at the ACR of all the combinations in [Tables 4](#_bookmark32) and [5](#_bookmark33), Spatial + FEW and Angle + FEW are the best compared to the other combina- tions for both our dataset and the benchmark dataset. It is justi- fiable because Spatial + FEW and Angle + FEW include features of facial regions and background information. Therefore, to achieve the best results, we propose the combination of Spatial + FEW and Angle + FEW, which is the proposed method and the results are reported in [Table 6](#_bookmark34) for our dataset and the benchmark dataset.

When we compare ACR of Spatial + FEW and Angle + FEW with the results of the proposed method (Spatial + Angle + FEW), ACRs for the respective three experiments on our dataset are al- most the same. This is because the proposed method has been de- veloped based on our dataset. However, when we compare the ACR of Spatial + FEW, Angle + FEW, and the proposed method on the benchmark dataset, there is a significant improvement for the pro- posed method compared to Spatial + FEW and Angle + FEW. Hence, we can conclude that the proposed method is capable of handling complex datasets. It is observed from the ACR of the proposed method on our dataset and the benchmark dataset reported in [Table 6](#_bookmark34) that the proposed method scores highly on the benchmark dataset compared to our dataset. The reason is that our dataset in- cludes diverse images such as those of different culture, modern families and non-family photos. At the same time, the benchmark dataset provides a large number of images for training, i.e., 1790

**Table 2**

Confusion matrices of spatial, angle and spatial + angle on ours and the benchmark dataset in (%).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Spatial |  | | | |  | Angle |  | | | |  | Spatial + Angle | |  | | | |
| Classes | Our | Benchmark | | | |  | Our | Benchmark | | | |  | Our | | Benchmark | | |  |
|  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  |
| Family | 66.5 | 33.50 |  | 32.62 | 67.37 |  | 74.22 | 25.77 |  | 39.40 | 60.59 |  | 76.80 | 23.2 |  | 43.92 | 56.07 |  |
| Non-family | 35.07 | 64.92 |  | 39.88 | 60.11 |  | 21.46 | 78.53 |  | 33.92 | 66.12 |  | 20.94 | 79.05 |  | 34.96 | 65.08 |  |
| ACR | 65.70 |  |  | 46.36 |  |  | 76.37 |  |  | 52.76 |  |  | 77.92 |  |  | 54.49 |  |  |
| **Table 3**  Confusion matrices of FEF, FEB and FEF+FEB on ours and the benchmark dataset in (%). | | | | | | | | | | | | | | | | | | |
|  | FEF |  | | | |  | FEB |  | | | |  | FEW (FEF + FEB) | |  | | |  |
| Classes | Our | Benchmark | | | |  | Our | Benchmark | | | |  | Our | | Benchmark | | |  |
|  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  |
| Family | 75.51 | 24.48 |  | 50.09 | 49.01 |  | 64.43 | 35.56 |  | 4.99 | 95.33 |  | 80.92 | 19.07 |  | 54.65 | 45.35 |  |
| Non-family | 28.53 | 71.46 |  | 14.67 | 85.52 |  | 44.24 | 55.75 |  | 5.43 | 94.61 |  | 12.82 | 87.17 |  | 12.71 | 87.29 |  |
| ACR | 73.48 |  |  | 67.80 |  |  | 60.09 |  |  | 49.63 |  |  | 84.04 |  |  | 70.97 |  |  |

**Table 4**

Confusion matrices of Spatial + FEF, Spatial + FEB and Spatial + FEW on ours and the benchmark dataset in (%).

|  |  |  |
| --- | --- | --- |
| Spatial + FEF | Spatial + FEB | Spatial + FEW |
| Classes Our Benchmark | Our Benchmark | Our Benchmark |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  |
| Family | 73.71 | 26.28 |  | 33.68 | 66.31 |  | 62.62 | 37.38 |  | 37.35 | 62.57 |  | 85.56 | 14.44 |  | 71.04 | 28.88 |  |
| Non-family | 27.74 | 72.25 |  | 16.48 | 83.51 |  | 44.50 | 55.49 |  | 29.12 | 70.92 |  | 14.39 | 85.60 |  | 15.94 | 84.10 |  |
| ACR | 72.98 |  |  | 58.59 |  |  | 59.42 |  |  | 54.13 |  |  | 85.58 |  |  | 77.57 |  |  |

**Table 5**

Confusion matrices of Angle + FEF, Angle + FEB and Angle + FEW on ours and the benchmark dataset in (%).

|  |  |  |
| --- | --- | --- |
| Angle + FEF | Angle + FEB | Angle + FEW |
| Classes Our Benchmark | Our Benchmark | Our Benchmark |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  |
| Family | 84.02 | 15.97 |  | 43.64 | 56.35 |  | 74.48 | 25.52 |  | 45.83 | 54.16 |  | 83.24 | 16.75 |  | 67.86 | 32.13 |  |
| Non-Family | 19.10 | 80.89 |  | 14.26 | 85.73 |  | 26.17 | 73.82 |  | 14.22 | 85.52 |  | 11.78 | 88.21 |  | 19.80 | 80.20 |  |
| ACR | 82.46 |  |  | 64.68 |  |  | 74.15 |  |  | 65.67 |  |  | 85.72 |  |  | 74.03 |  |  |

**Table 6**

Confusion matrix and classification rate of the proposed method (spatial + angle + FEW) without an averaging operation, with CNNs, SVMs and CNN on raw pixels in the images on ours and the benchmark dataset (in %).

Proposed method (with averaging) Proposed without averaging

Proposed method CNN on raw

pixels Proposed method with SVM

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classes | Our |  |  | Benchmark | |  | Our |  |  | Benchmark | |  | Our |  |  | Benchmark | |  | Our |  |  | Benchmark | |  |
|  | Family | Non- family |  | Family | Non- family |  | Family | Non- family |  | Family | Non- family |  | Family | Non- family |  | Family | Non- family |  | Family | Non- family |  | Family | Non- family |
| Family | 88.40 | 11.59 |  | 96.36 | 3.63 |  | 77.31 | 22.68 |  | 48.92 | 51.08 |  | 62.62 | 37.37 |  | 91.58 | 8.42 |  | 80.87 | 19.13 |  | 91.67 | 8.32 |  |
| Non-family | 15.70 | 84.29 |  | 1.16 | 98.83 |  | 80.36 | 19.63 |  | 19.22 | 80.78 |  | 56.70 | 43.29 |  | 88.84 | 11.16 |  | 12.56 | 87.43 |  | 5.70 | 94.29 |  |
| ACR | 86.34 |  |  | 97.59 |  |  | 49.86 |  |  | 64.85 |  |  | 52.59 |  |  | 53.53 |  |  | 84.13 |  |  | 93.26 |  |  |

for family and 2753 for non-family compared to 388 for family and

382 for non-family images of our dataset. This is the advantage of the benchmark dataset for achieving the best results compared to ours. This is because when we feed a large number of training samples to the classifier, it covers more possible variations in im- ages. Therefore, a large number of training samples and more vari- ations led to achieving the best results for the benchmark dataset by the proposed method compared to our dataset.

In the case of spatial and angle feature extraction discussed in the Proposed Methodology Section, the proposed method com- putes the mean for distances and angles of all the faces in the respective images separately. To assess the influence of averag- ing (the mean), we conduct experiments for calculating the clas- sification rate using the proposed method without averaging. In

other words, the proposed method considers all the distance and angle features of faces in images as distance and angle feature vectors respectively for classification. The results are reported in [Table 6](#_bookmark34), where one can see the proposed method without averag- ing the scores providing very poor results compared to the pro- posed method with averaging for both the datasets. This shows that the operation, namely, averaging, plays a vital role in achiev- ing better results for family and non-family classification.

Since we use the CNN for classification, to show its effective- ness compared to the SVM and the use of raw pixels with the CNN, the proposed method is used for experimentation of the proposed features with an SVM as well as feeding raw pixels to a CNN for ours and the benchmark dataset. For experiments using raw pix- els of the images, the proposed method considers each pixel value





**Fig. 11.** Qualitative results of successful and unsuccessful classification employing the proposed method on our dataset and the benchmark dataset.

as a feature and it converts a two-dimensional image matrix to single-dimensional feature vector in a row-wise fashion. The con- verted single-dimensional feature vector is passed to a CNN for classification. This experiment does not involve the proposed dis- tance, angle and fractional entropy-based features for calculating the measures. The results are reported in [Table 6](#_bookmark34). It is noted from [Table 6](#_bookmark34) that the results of feeding raw pixels directly to a CNN per- forms poorly in terms of classification rate compared to the pro- posed features with a CNN. The main reason for the poor results is that since the number of samples for the training set is small, it may not cover all the possible variations of images when we feed raw pixels to the CNN directly. For experimentation with an SVM and for a fair comparative study with the CNN, the proposed method uses a polynomial kernel as it is non-linear like the CNN classifier. When we compare the results of the proposed features with an SVM and the proposed features with a CNN, the proposed features with an SVM achieve poorer results compared to the pro- posed features with a CNN. It is justifiable because the SVM does not have a generalization ability as is the case with a CNN. In addi- tion, the performance of the SVM depends on the kernel type and size. On the other hand, the CNN can learn complex non-linear in-

put and output relationships. Therefore, for the proposed problem, which is complex in terms of foreground and background varia- tions according to the statistics reported in [Table 1](#_bookmark27), the proposed features with a CNN perform better than the proposed features with an SVM.

We also report the results of two existing methods on our [dataset and the benchmark datasets in](#_bookmark55) [Table 7](#_bookmark36)[. Since (Wang et al., 2017) is the improved version of](#_bookmark55) [Wa](#_bookmark57)[ng et al. (2015), whereby](#_bookmark55) [Wang et al. (2017)](#_bookmark55) gives better results in terms of ACR. When we compare the ACR of the proposed method with two existing meth- [ods, the proposed method gives better results than (Wang et al., 2015) and (Want et al., 2017). This is understandable as both the](#_bookmark57) existing methods use only facial regions for classification, while the proposed method uses both facial and background regions. In ad- dition, the proposed method extracts geometric features based on spatial and angle information, and the new fractional entropy fea- ture are an enhancement on existing methods and hence it makes a positive difference.

Sample qualitative results of the proposed method on our dataset and the benchmark dataset are shown in [Fig. 11](#_bookmark35). [Fig. 11](#_bookmark35) also includes the results of misclassifications by the

**Table 7**

Confusion matrix of the proposed (spatial + angle + FEW) and existing methods on our dataset and the benchmark dataset in (%).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Proposed |  |  | [Wang et al. (2015)](#_bookmark57) |  |  | [Wang et al. (2017)](#_bookmark55) |  |
| Our | Benchmark |  | Our | Benchmark |  | Our | Benchmark |

Classes

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  | Family | Non-family |  |
| Family | 88.40 | 11.59 |  | 96.36 | 3.63 |  | 76.28 | 23.71 |  | 88.54 | 11.45 |  | 82.73 | 17.27 |  | 94.52 | 5.47 |  |
| Non-family | 15.70 | 84.29 |  | 1.16 | 98.83 |  | 37.43 | 62.56 |  | 24.95 | 75.04 |  | 22.51 | 77.49 |  | 14.38 | 85.61 |  |
| ACR | 86.34 |  |  | 97.59 |  |  | 69.42 |  |  | 81.79 |  |  | 80.11 |  |  | 90.06 |  |  |

**Table 8**

Confusion matrices and classification rate for family and non-family images with different foreground and background patterns on ours and the bechmark dataset (in %).

Family-Hierarchical *vs.* Non-family Non-Hierarchical

Family-Non-Hierarchical *vs.* Family

Hierarchical Family-Indoor *vs.* Non-family-Outdoor Family-Outdoor *vs.* Non-family-Indoor

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Classes | Our |  | Benchmark | | Our |  | Benchmark | | Our |  | Benchmark | | Our |  | Benchmark | |  |
|  |  | Non- |  | Non- |  | Non- |  | Non- |  | Non- |  | Non- |  | Non- |  | Non- |
|  | Family | family | Family | family | Family | family | Family | family | Family | family | Family | family | Family | family | Family | family |
| Family | 72.0 | 27.0 | 77.0 | 23.0 | 66.0 | 34.0 | 73.0 | 27.0 | 68.0 | 31.0 | 76.0 | 24.0 | 57.0 | 42.0 | 66.0 | 34.0 |  |
| Non- Family ACR | 31.0  70.5 | 69.0 | 33.0  72.0 | 67.0 | 35.0  65.0 | 64.0 | 43.0  65.0 | 57.0 | 29.0  69.5 | 71.0 | 23.0  76.5 | 77.0 | 26.0  65.5 | 74.0 | 27.0  69.5 | 73.0 |  |

proposed method on our dataset and the benchmark dataset. The reason for misclassification is that when the images of family and non-family images share geometric structures of the faces and the properties of backgrounds, the proposed method fails to perform correct classification. Therefore, there is scope for improvement in the future.

The existing methods ([Wang et al., 2015, 2017](#_bookmark57)) that work based on the fact that the height distributions of persons in images should satisfy a hierarchical order for family, while it does not for non-family. In the same way, according to the statistics in [Table 1](#_bookmark27), the benchmark dataset contains more images with in- door scenes for the family class, and more images with outdoor scenes for the non-family class. However, the proposed method does not consider these two constraints for the classification of family and non-family images. It is evident from the statistics re- ported in [Table 1](#_bookmark27) for our dataset, where it can be seen that the ratio of hierarchical to the total number of family images and non- hierarchical to the total number of non-family images is greater compared to that from the benchmark dataset. The same thing is true for images with indoor scenes for family and outdoor scenes for non-family images. To validate the statement, we conducted ex- periments for Family-Hierarchical *vs.* Non-family-Non-hierarchical and Family-Non-hierarchical *vs.* Non-family-Hierarchical on both ours and the benchmark dataset, and the results are reported in [Table 8](#_bookmark37). It is observed from [Table 8](#_bookmark37) that the classification rate for the expected order is higher than that of the other order. Therefore, we can conclude that there is not much influence on the overall performance of the proposed method. Similarly, im- ages with indoor and outdoor scenes do not have much of an ef- fect on the overall performance of the proposed method. It is ev- ident from the results reported in [Table 8](#_bookmark37) for Family-Indoor *vs.* Non-family-Outdoor and vice versa. In summary, for all the exper- iments listed in [Table 8](#_bookmark37) for both ours and the benchmark dataset, the proposed method achieves almost consistent average classifica- tion rates. This demonstrates that the proposed method works well irrespective of the background complexities and hierarchical distri- bution of heights. However, when we compare the results of the proposed method on the whole dataset ([Table 7](#_bookmark36)) without separa- tion and the results in [Table 8](#_bookmark37), the results of the proposed method in [Table 7](#_bookmark36) are higher than those in [Table 8](#_bookmark37) due to fewer training samples which represent the variations in the case of individual experiments listed in [Table 8](#_bookmark37).

### Conclusions and future work

In this paper, we have proposed a new idea for classifying fam- ily and non-family photos by combining facial structure and back- ground texture. The proposed method explores distances between facial key points for extracting spatial features. In addition, angles between facial key points are also explored for studying the struc- tures of faces, which are called geometric features. To make use of the background information and textural properties of facial re- gions, we have proposed novel Tsallis fractional entropy-based fea- tures. Furthermore, the proposed method combines spatial, angle and fractional entropy features to obtain the feature vector. The feature vector is applied to a conventional convolutional neural network for classification. Experimental results on our own dataset and the benchmark datasets show that the proposed method is better than two state-of-the-art methods in terms of average clas- sification rate.

The main contributions are the following. It is inherent that fa- cial regions are the key factor for family and non-family photo im- age classification. Based on this observation, we explore distance features for facial key points as spatial features to study the struc- ture of facial regions. We have used angle information for facial key points to make spatial features robust to extract the detailed structure of facial regions. The way we combine spatial and angle- based features as geometric ones is novel and an interesting ap- proach to tackle the issues of family and non-family photo classifi- cation. To extract regular patterns in facial and background regions (other than facial region), we propose a novel idea of introducing Tsallis fractional entropy for extracting texture properties of facial regions and other background regions. Furthermore, the proposed method combines geometric and fractional entropy features in a different way for achieving the best results.

Despite having proposed a new idea for family and non-family images classification, there are some limitations to the proposed approach. Sometimes, when family and non-family image share the same properties of facial regions with the background, the pro- posed method fails to yield good results. This is understandable because one can expect similar patterns of foreground and back- ground for both family and non-family images. In this case, we need a method, which can work irrespective of background and facial regions. One way is to introduce context features using fore- ground and background information to find a solution regarding

context, which can be independent of facial regions and the back- ground.

When photos contain both family and non-family members, the proposed method may not work well. It is beyond the scope of the proposed work as it is hard to separate family or non-family mem- bers in the same image. To find a solution, one possible way is to bring multimodal concepts, such as face, skin, dress, and structure of the body. This is due to the potential of sharing personal traits and habits with members belonging to the same family. If individ- uals do not belong to a particular family, we can expect different habits, structures (apart from the face), skin, etc.

In summary, this paper presents a new idea for finding a solu- tion to family and non-family photo classification. The proposed work demonstrates a promising direction for solving a number of issues, including human traﬃcking. There are several potential concepts, which can be considered as new research directions for future study. In order to support reproducible research, the dataset and code will made available to readers upon request.

### Conflict of interest

None.

### CRediT authorship contribution statement

**Maryam Asadzadeh Kaljahi:** Implementing Methodology, data curation. **Palaiahnakote Shivakumara:** Conceptualization, Formal analysis, Supervision, Draft writing, Investigation. **Tiangping Hu:** data curation, Validation. **Hamid A. Jalab:** Formal analysis, Model- ing. **Rabha W. Ibrahim:** Formal analysis, Modeling. **Michael Blu-** **mentsein:** Review editing. **Tong Lu:** Validation and reviewing. **Mohamad Nizam:** Validation and reviewing.

### Acknowledgments

[This work was supported by the Natural Science Foundation of China under Grant](http://dx.doi.org/10.13039/501100001809) [6](#_bookmark59)[1672273 and Grant](http://dx.doi.org/10.13039/501100001809) [6](#_bookmark59)[1832008, and the Science](http://dx.doi.org/10.13039/501100001809) Foundation for Distinguished Young Scholars of Jiangsuunder Grant BK20160021.

### Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.eswax.2019.100008](https://doi.org/10.1016/j.eswax.2019.100008).

### References

[Arora, R., & Suman (2012). Comparative analysis of classification algorithms on dif- ferent datasets using WEKA. *International Journal of Computer Applications, 54*, 21–25.](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0001) <https://github.com/amten/NeuralNetwork>[.](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0001)

[Cai,](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0002) [G., Hio, C., Bermingham, L., Lee, K., & Lee, I. (2014). Sequential pattern mining of geo-tagged photos with an arbitrary regions-of-interest detection method. *Ex- pert Systems with Applications, 41*, 3514–3526.](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0002)

[Dai,](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0003) [Q., Carr, P., Sigal, L., & Hoiem, D. (2015). Family member identification from photo collections. In *Proc. WCACV* (pp. 982–989).](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0003)

[Dandekar, A. R., & Nimbarte, M. S. (2014a). A survey: Verification of family relation- ship from parents and child facial images. In *Proc. SCEECS*.](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0004)

[Dandekar, A. R., & Nimbarte, M. S. (2014b). Verification of family relation from par- ents and child facial images. In *Proc. INPAC* (pp. 157–162).](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0005)

[Gallagher, A. C., & Chen, T. (2009). Understanding images of groups of people. In](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0006)

[*Proc. CVPR* (pp. 256–263).](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0006)

[Haghighat,](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0007) [M., Zonouz, S., & Mottaleb, M. A. (2015). CloudID: Trustworthy cloud-based and cross-enterprise biometric identification. *Expert Systems with Applications, 42*, 7905–7916.](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0007)

[Ibrahim, R. W., Moghaddasi, Z., Hamid, A. J., & Noor, R. M. (2015). Fractional dif- ferential texture descriptors based on the machado entropy for image splicing detection. *Entropy, 17*(7), 4775–4785.](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0008)

[McAllister, P., Zheng, H., Bond, R., & Moorhead, A. (2016). Towards personalized training of machine learning algorithms for food image classification using a smartphone camera. In *Proc. ICUCAI* (pp. 178–190).](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0009)

[Mehta, J., Ramnani, E., & Singh, S. (2018). Face detection and tagging using deep learning. In *Proc. ICCCCSP*.](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0010)

[Nanni, L., Chidoni, S., & Brahnam, S. (2018). Ensemble of convolutional neural net- works for bioimage classification. *Applied Computing and Informatics, 15*.](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0011)

[Ng,](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0012) [W. W. Y., Zheng, T. M., Chan, P. P. K., & Yeung, D. S. (2011). Social relation- ship discovery and face annotations in personal photo collection. In *Proc. ICMLC* (pp. 631–637).](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0012)

[Qin,](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0013) [X., Tan, X., & Chen, S. (2015). Tri-subject kinship verification: Understanding the core of a family. *IEEE Trans, Multimedia, 17*, 1855–1867.](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0013)

[Ren, S., Cao, X., Wei, Y., & Sun, J. (2014). Face alignment at 3000 fps via regressing local binary features. In *Proc. CVPR* (pp. 1685–1692).](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0014)

[Robinson, J. P., Shao, M., Wu, Y., Gillis, T., & Fu, Y. (2018). *Visual kinship recognition of families in the wild: 40* (pp. 2624–2837).](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0015)

[Shen,](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0016) [C. T., Liu, J. C., Shih, S. W., & Hong, J. S. (2009). Towards intelligent photo com- position-automatic detection of unintentional dissection lines in environmental portrait photos. *Expert Systems with Applications, 36*, 9024–9030.](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0016)

[Tsallis,](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0017) [C. (2009). *Introduction to nonextensive statistical mechanics*. Springer, Sci- ence](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0017) + [Business Media.](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0017)

[Wang, X., Guo, G., Merler, M., Codella, N. C., Rohith, M., Smith, J. R., et al. (2017). Leveraging multiple cues for recognizing family photos. *Image and Vision Com- puting, 58*, 61–75.](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0018)

[Wang, X., Guo, G., Rohith, M., & Kambhamettu, C. (2015). Leveraging geometry and appearance cues for recognizing family photos. In *Proc. ICWAFG* (pp. 1–8).](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0019)

[Xia,](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0020) [S., Pan, H., & Qin, A. K. (2014). Face clustering in photo album. In *Proc. ICPR*](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0020)

[(pp. 2844–2848).](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0020)

[Zhen,](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0021) [L., Caiming, Z., & Caixian, C. (2018). MMDF-LDA: An improved multi-modal la- tent dirichlet allocation model for social image annotations. *Expert Systems with Applications, 104*, 168–184.](http://refhub.elsevier.com/S2590-1885(19)30008-3/sbref0021)