Pattern Recognition Letters 000 (2017) 1–9

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Multi angle optimal pattern-based deep learning for automatic facial expression recognition

Deepak Kumar Jain[a,b], Zhang Zhang[a,b], Kaiqi Huang[a,b,*]

[a] CRIPAC & NLPR, CASIA, PR China
[b] University of Chinese Academy of Sciences, PR China

ARTICLE INFO

Article history:
Available online xxx

MSC:
41A10
65D05
65D17

Keywords:
STM
SURF
CNN
LSTM

ABSTRACT

Facial Expression Recognition (FER) plays the vital role in the Human Computer Interface (HCI) applications. The illumination and pose variations affect the FER adversely. The projection of complex 3D actions on the image plane and the inaccurate alignment are the major issues in the FER process. This paper presents the novel Multi-Angle Optimal Pattern-based Deep Learning (MAOP-DL) method to rectify the problem from sudden illumination changes, find the proper alignment of a feature set by using multi-angle-based optimal configurations. The proposed method includes the five major processes as Extended Boundary Background Subtraction (EBBS), Multi-Angle Texture Pattern+STM, Densely Extracted SURF+Local Occupancy Pattern (LOP), Priority Particle Cuckoo Search Optimization (PPCSO) and Long Short-Term Memory -Convolutional Neural Network (LSTM-CNN). Initially, the EBBS algorithm subtracts the background and isolates the foreground from the images which overcome the illumination and pose variation. Then, the MATP-STM extracts the texture patterns and DESURF-LOP extracts the relevant key features of the facial points. The PPCSO algorithm selects the relevant features from the MATP-STM feature set to speed up the classification. The employment of LSTM-CNN predicts the required label for the facial expressions.The major key findings of the proposed work are clear image analysis, effective handling of pose/illumination variations and the facial alignment. The proposed MAOP-DL validates its effectiveness on two standard databases such as CK+ and MMI regarding various metrics and confirm their assurance of wide applicability in recent applications.

1. Introduction

Rapid growth of Human-Computer Interfaces (HCI) induces the challenges in Facial Expression Recognition (FER). Even though the frontal view recognition is well on the basis of the appearance or geometry model, the Multiview Facial Expression Recognition (MFER) is the challenging issue [8]. Illumination conditions, age and size variations in the facial image and videos that lead to the development of Automatic FER in three stages such as face detection and tracking,feature extraction and emotion classification in [19] and [13].With numerous development of applications such as automatic video indexing, surveillance and assisted living, detection and recognition of actions in natural settings is an important stage. Recently, the two type of datasets with the same actions in the wild and video clips introduces the new complexities to the recognition community. The lack of constraints on such type of data makes the classification as challenging one. Recently, an Automated Face Recognition (AFR) [5] is an attractive research study due to the extensive applications of mobile phone authentication to surveillance. There are two types of environments available for AFR such as controlled (frontal, neutral expressions and uniform illumination) and uncontrolled environment (arbitrary poses, non-uniform illumination and occlusions.

AFR in an uncontrolled environment [16] is the challenging task due to its intense pose variations which are considered as major uncertain blocks in AFR. Temporal alignment and the semantic aware dynamic representation were the major issues in FER due to the repeatable property of low-level features. Manual designed cuboids have the capability to capture the low-level information instead of high-level information provided the less discriminative capability [12]. The major requirement like all the images is available prior to congealing leads to offline tasks with low efficiency [18]. The combination of expression dynamics with the 3D facial geometry [21] provided the wealth information that opens up the new researches [37] on the capture of a face in all motions and detection of cues.

* Corresponding author.
E-mail addresses: deepak.juet@cripac.ia.ac.cn, deepak.jain.juet@gmail.com (D.K. Jain), zzhang@nlpr.ia.ac.cn (Z. Zhang), kaiqi.huang@nlpr.ia.ac.cn (K. Huang).

http://dx.doi.org/10.1016/j.patrec.2017.06.025
0167-8655/© 2017 Published by Elsevier B.V.

The provision of temporal coherence and the achievement of required matching accuracy depend on the estimation of similarity between the expressions. The data driven approaches [11,32] predict the k-nearest neighbors which are considered as the candidates for each frame. The proposal of expression transfer method refined the expression retrieved from the database.The dense registration of point clouds is the computational task in AFR due to the large size 3D vertices in the facial mesh. Besides, the existence of landmarks is the pre-requisite for FER. The partial separation of facial cues induced the complications in automatic FER applications. Hence, the issues addressed in the automatic FER are foreground/background separation, facial key point's extraction, the dimensionality of features and multi-labeling-based recognition. The major contributions of proposed MAOP-DL model are listed as follows:

- The Extended Boundary Background Subtraction (EBBS) isolates the background and foreground of images that made the system as adaptive to sudden illumination changes.
- The combination of Multi-Angular Texture Pattern (MATP) and the Spatio-Temporal Matching (STM) algorithm extracts the different features that provide the clear image analysis and proper alignment.
- The Densely Extracted SURF model integrated with the Local Occupancy Pattern (LOP) algorithm govern the facial key points to accurately identify the muscle variations during the expression.
- Finally, the application of Priority Particle Cuckoo Search Optimization (PPCSO) algorithm on MATP-features reduces the dimensionality and the utilization of Long Short Term Memory-Convolution Neural Network (CNN) improves the classification performance effectively.

The rest of the paper is organized as follows. Section 2 presents a description about the previous research studies relevant to an automated FER and their issues. Section 3 describes the implementation process of proposed MATP-DESURF model for an effective FER. Section 4 investigates the performance of proposed system with the various performance metrics over the existing methods. Finally, Section 5 presents the conclusion and future enhancement of proposed research work.

2. Related work

Due to the small, partially visible and the occluded relevant objects in human interactions, the recognition of mutual contexts between each other is the difficult task. Yao et al. [31] proposed the mutual context model in which objects and human pose are jointly modeled in recognition of human-pose interaction activities. The interference learned from the Action Units (AU) was static and the temporal changes in facial expressions led to more discrimination in intensity levels. Rudovic et al. [20] proposed the novel Conditional Ordinal Random Field (CORF) model to model the context-sensitive approach for facial action recognition.). Oh et al. [17]proposed Radial Basis Function RBF-Neural network on the basis of the combination of Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA). The improper facial alignments and the occlusions are the major issues in traditional PCA-LDA-based dimensionality reduction techniques. Siddiqi et al. [27] introduced the robust FER system that employed the Step-Wise LDA (SWLDA) to select the localized features from the expression frames. Happy et al. [6] proposed the novel facial landmark detection and the salient patch-based FER under the diverse resolutions of the image. Fast network training with low human supervision in an optimization process decreases the network generalization ability.

On the basis of learning, several approaches are evolved in research studies to deal with the issues in an accurate recognition. Barakova et al. [1] proposed the interpretation method of emotions that were detected in facial expressions in the context of events. The genetic search and the optimization of parameters were regarded as the next step toward the automatic analysis of facial expressions. Iosifidis et al. [7] proposed the graph-based embedded ELM algorithm that exploited both the penalty and SL criteria in infinite dimensional spaces. The dense registration of point clouds is the computational task in AFR due to the large size 3D vertices in the facial mesh. Besides, the existence of landmarks is the pre-requisite for facial expression recognition. Zhao et al. [35] presented the probabilistic framework on the basis of Bayesian Belief Network (BBN) on the basis of Statistical Feature Models (SFM) and Gibbs–Boltzmann distribution for geometric and appearance features. Blur and illumination were the major problems and by using the blurring operation of given images convex set was formed. Vageeswaran et al. [28] proposed the blur-robust algorithm to solve the convex optimization problems. Recovery of High Resolution (HR) images from the Low-Resolution (LR) sequences was the major objective of superresolution and achieved by hallucination methods. Ma et al. [15] reviewed the several maps-based models under pose, illumination and expressions (hallucination problems). The modeling of mappings was performed by using the redundant and sparse representation.

The selection of representative features and the uncertain class labels made the feature selection as critical task. Wang et al. [30] proposed the unsupervised feature selection for the recognition of facial features in presence of class labels. The utilization of Zernike Moments (ZM) offered the promising recognition rates with its rotational invariances. Sariyanidi [22] modified the ZM to obtain the local representation by computing the moments of each pixel of face image on the basis of the local neighborhood. Sariyanidi et al. [23] reviewed and analyzed several states of art solutions by splitting up the pipeline of face recognition process into four stages as registration, representation, dimensionality reduction and recognition. Senechal et al. [24] utilized the multi-kernel SVM for each AU with the Local Gabor Binary Pattern histograms and the Active Appearance Model (AAM) coefficients. The capture of complex decision boundary is the difficult task in spontaneous emotions analysis. Shan [25]constructed the strong classifier by using the combination of weak classifiers with the AdaBoost approach. The features to recognize the smile expression are the intensity difference between the pixels exist in the grayscale images.

During the image annotation process, the selection of features prior to investigation of properties of feature combination do not provide any significant contribution. Zhang et al. [33] introduced the regularization-based feature selection algorithm to leverage both the sparsity and clustering properties of features for annotation phase. Complex variations, gross errors and preservation of local details of image introduced the difficulties in shape priors modeling. Zhang et al. [34] proposed the Sparse Shape Composition model (SSC) to address the aforementioned challenges. The partial separation of facial cues induced the complications in automatic FER applications and such problem is regarded as identity-independent ER problem. Kung et al. [10] proposed the Dual Subspace Non-negative Graph Embedding (DSNGE) for the representation of expressive images on two subspaces as identity and expression. Borude et al. [2] discussed the methods available for facial features tracking under various illuminations, pose variations and lighting conditions. Guo et al. [4] proposed the dynamic facial expression recognition which was formulated with the longitudinal group wise registration problem.The major contributions of the dynamic recognition are diffeomorphic growth model, built of salient longitudinal facial expression, sparse representation (spatial and temporal domain) guided the recognition effectively. The survey of traditional research studies indicated that the presence of

Please cite this article as: D.K. Jain et al., Multi angle optimal pattern-based deep learning for automatic facial expression recognition, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.06.025

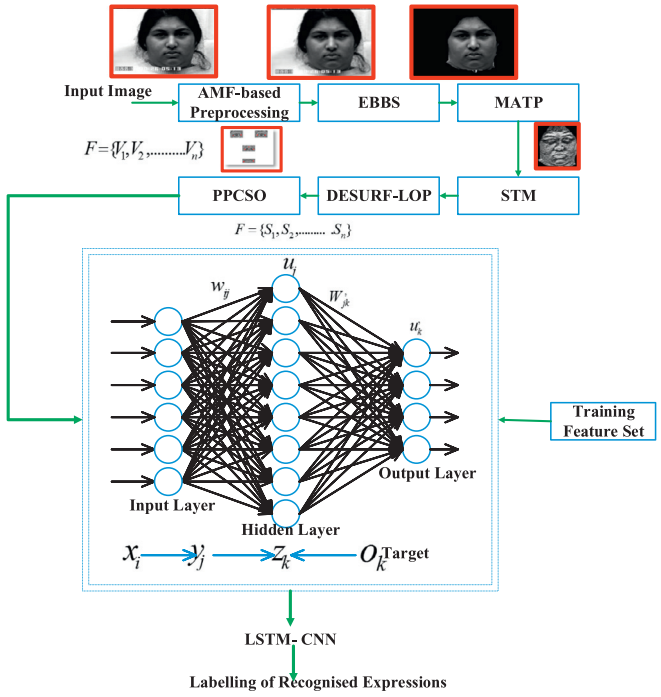


Fig. 1. Overall Flow of Proposed MAOP-DL for FER.

illumination and pose variations, dimensionality of features were the major issues in FER applications.

3. Overall flow of proposed MATP-DESURF-based FER

An accurate automated recognition of facial expressions with the diverse illumination/pose variations, high dimensionality of features and the exact facial key points is the major focus of the research work proposed in this paper. Fig. 1 shows the overall workflow of proposed MATP-DESURF-based FER.

Initially, the RGB image from the real-time database is considered as the input to the system. The presence of noise introduces the difficulties in pixel intensity estimation. Hence, the input image is preprocessed with the Adaptive Median Filtering (AMF) to remove the noise and smoothen the image that leads to intensity normalization. Then, the EBBS algorithm is applied to the noise-free image to isolate the background and foreground. The feature points are necessary to recognize the emotions in HCI applications.

This paper proposes the novel MATP combined with Densely Extracted SURF model, LOP and STM algorithm to extract the feature key points. The split-up of extracted patterns into several grids and the extraction of facial key point features increase the dimensionality. Hence, the PPCSO is used in a proposed system to select the optimal relevant features. The optimally selected result is matched with the training feature grid through LSTM-CNN classification to recognize the expressions. The isolation of background and foreground prior to the feature extraction and the dimensionality reduction through the optimization in proposed work provided the properly aligned feature set for FER.

4. Foreground extraction

The noise removal is the initial stage of proposed work. The presence of impulsive noise and the distortion in object boundaries induce the difficulties in image processing. In this paper, the AMF [26] is used to remove the noise and reduce the distortion (excessive thinning and thickening) in two levels. The application of size of the window to filter the image pixels is adaptive in nature. The noise present in the image is removed through AMF as follows:

Level 1: Check whether the median pixel value of the image (I_{med})is in between the ranges of minimum and maximum pixels ($I_{\min}, I_{\max}$)and if it exists, then the value is not considered as an impulse. Then, the algorithm goes to level 2 to check the next pixel. Otherwise, the size of the window is increased and the level 1 is repeated until the median value is not an impulse.

Level 2: Check the current pixel is in the ranges of minimum and maximum pixels and if it exists, then the filtered image pixel is unchanged. Otherwise, the pixel is regarded as corrupted and the filtered image pixel is assigned the median value for level 2.

An isolation of background from the foreground of noise-free image is the next stage in proposed work.This paper utilizes the EBBS algorithm to extract the foreground from the noise-free image. The algorithm comprises three essential processes such as cell formation, directionality estimation, and the intensity depth analysis to extract the foreground from the images. To implement this processes, the window formation is the prior step. Hence, the window necessary to project the input image is formed as follows:

$$W = I(i-1\ to\ i+1, j-1\ to\ j+1) \tag{1}$$

Once the window is formed, the magnitude that represents the multiple cells of window is computed. The normal distribution of (G) values in each cell and the weight for window are responsible for the magnitude estimation. In the proposed EBBS algorithm, the directionality-based ROI with depth analysis plays the major role in foreground extraction. The mathematical formulation of directionality ($\bar{D}_{x,y}$) is defined as follows:

$$\bar{D}_{x,y} = \sigma^2 \nabla^2 G(x, y, \sigma^2) \tag{2}$$

where, σ= Standard deviation and x,y = Size of Window By convoluting the directionality ($\bar{D}_{x,y}$) with the window formed (W), the weight corresponds to the magnitude estimation is updated such as

$$\bar{W}(x, y) = W(x, y) * \bar{D}_{x,y} \tag{3}$$

Where $W(x, y)$= Initial Window and $\bar{W}(x, y)$= Size of window. The updated weight values from the Eq. (3) are used to estimate the magnitude for depth analysis(Te) as follows:

$$Te(x, y) = \frac{\sum(W(x, y) \times \bar{W}(x, y))}{\sqrt{\sum W(x, y)^2 - \bar{W}^2}} \tag{4}$$

The magnitudes obtained from Eq. (4) are compared with the updated weight values. If the magnitude is greater than the updated weight value, then the encoded form of likelihood value (L) for the patterns necessary to isolate foreground and background is predicted. The Algorithm 1 to extract the foreground output as follows

The encoded values (E) are multiplied with the average value of magnitude of the intensity depth to extract the patterns. If the value of the patterns are greater than zero, then the foreground of the image. The values are replaced with zero otherwise.

The cells that represent the noise-free image are formed and the binary difference of pixel values is sequentially estimated as shown in Fig. 2.

Figs. 3 and 4 clearly depicts the foreground extraction based on the directionality values. From Fig. 3, the connected components from the magnitude estimation are predicted. The layers from the magnitude values are separated for foreground and background. By performing the deep intensity analysis through the encoding process, the background of the image is subtracted and foreground necessary for FER is extracted.

5. MATP-DESURF-based facial key points

The muscle variations of facial key points (left eye, right eye, nose and mouth) play the major role in FER analysis. Hence, the

Algorithm 1 EBBS algorithm.

Input: Image 'I'
Output: EBBS output 'Y'
S-1: Initialize the window matrix (W)
S-2: Estimate the directionality ($\bar{D}_{x,y}$) by using equation 2
S-3: Update the initial weight by convolution defined in equation 3
S-4: Compute the magnitude of window by equation 4
S-5: Extract the likelihood value 'L'
S-5: Extract the likelihood value 'L'

$$L_{x,y}=\begin{cases}1, & \text{if } Te(x,y) > \bar{W}\\ 0, & \text{Else.}\end{cases}$$

S-6: Encode the likelihood value
E= Bin2dec(L)
S-7: Extract the patterns by using following formula
P_{ij}=E$\times \frac{\sum Te}{N-1}$ //N=size of window
S-8: Separate foreground and background

$$Y_{ij}=\begin{cases}I_{ij}, & \text{if } P_{ij} > 0\\ 0, & \text{Else.}\end{cases}$$

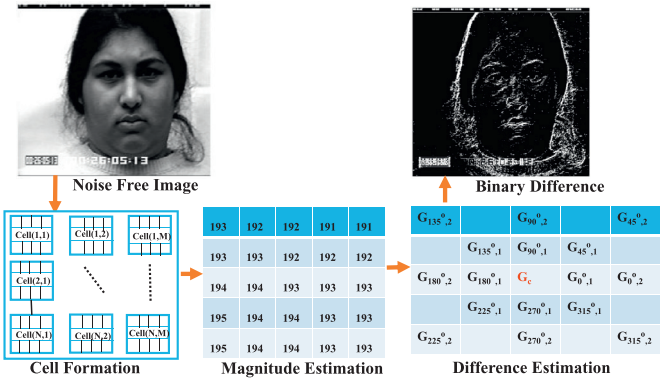


Fig. 2. Cell formation.

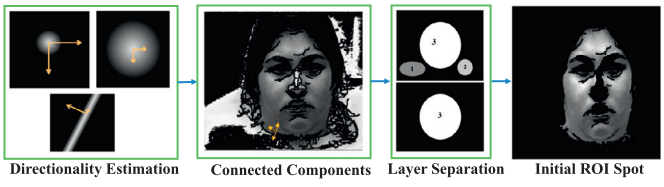


Fig. 3. Directionality-based ROI extraction.

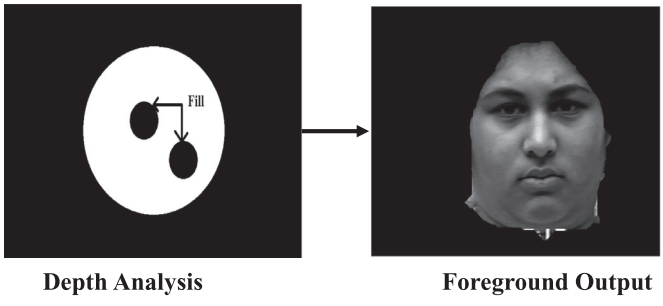


Fig. 4. Foreground extraction.

Fig. 5. MATP Pattern.

prediction of facial key points is the next stage of proposed work. The output from the EBBS algorithm is considered as the input to the MATP algorithm.

5.1. Multi-Angle texture pattern

In this stage, the window size is extended to 5 × 5 for projection of EBBS output. Then, the median value of projected image is computed. Initially, the window over the EBBS image is formed with the size of 5 × 5. Within this window, the cells with 3 × 3 is extracted separately. By applying the angle-based difference estimation, the rules required for the vector prediction is formed.

The algorithm to compute the patterns in multi-angular form is listed as follows (see Algorithm 2).

The magnitude value corresponding to the difference between the window formation (T1, T2) are mathematically expressed as follows:

$$mag = \sqrt{(double((T1(3,4) \sim T2(3,3))^2 + ((T1(3,4) \sim T2(3,3))^2)} \quad (5)$$

The comparison between the center pixel with the neighboring pixels is performed and then the decimal coding is performed to extract the patterns. The bitwise OR operation is performed with two types of patterns ($Pt2 \bigoplus Pt1$) extracted the relevant patterns and then mesh grid is constructed. The coordination among the values in the meshgrid is predicted with sparse form.

Finally, the angle-based texture patterns as shown in Fig. 5 that are necessary to predict the facial key points are predicted. The STM comprises the four major processes such as index estimation, cluster formation, template separation and tag representation. The STM is spanned by 3D blocks densely sampled from the dictionary that cover the local variations of spatial and temporal space. The application of bank of learned filters extract the low-level features from each block. The features are directly learned from the input dataset proved the generalizability. In proposed work, the query patterns derived from the MATP process is passed to the cluster formation. The index corresponding to the dense samples from the dictionary and the input MATP are estimated and then clustered.

The STM are arranged in four parts as Left eye, right eye, nose and mouth as shown in Figs. 6 and 7. Using this STM,the patterns relevant to the facial key points are extracted through the DESURF-LOP process.

5.2. DESURF-LOP

The histogram relevant to the facial key point features is the next stage of proposed work. The unique coordinates in sparse

Please cite this article as: D.K. Jain et al., Multi angle optimal pattern-based deep learning for automatic facial expression recognition, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.06.025

Algorithm 2 Multi-angular texture pattern.

Input: EBBS output 'Y'
Output: Texture pattern 'PT' and sparse data S

S-1: Initialize 5×5 window matrix
S-2: Project window over the EBBS output
For (i=3 to (*Row_size*)-2
For (j=3 to (*Column_size*)-2
$T1 = Y_{((i-1\ to\ i+1),(j-1\ to\ j+1))}$
S-3: Compute the median value for the window
mn=Median (T1)
S-4: Check the difference of center of pixel with the neighborhood over the angles of $0^\circ, 30^\circ, 45^\circ, 60^\circ, 90^\circ, 120^\circ, 135^\circ, 180^\circ$
if T1(1,2)>= mn && T1(2,3)>=mn
IR(1)=1;
elseif T1(1,2)<mn && T1(2,1)>=mn
IR(2)=2;
elseif T1(2,1)<mn && T1(3,2)<mn
IR(3)=3;
elseif T1(2,3)>= mn && T1(3,2)<mn
IR(4)=4;
Endif
S-5: Construct the next window
$T2 = Y_{((i-2\ to\ i+2),(j-2\ to\ j+2))}$
S-6: Compute the magnitude value from newly formed window by using equation˜5
S-7: Compute the vectors
$V = mag \times IR$
S-8: Compare the center and neighborhood pixels in the new window 'V'

$$f_1(g_{p1} \sim g_c) = \begin{cases} 1, & \text{if } g_{p1} \sim g_c \geq 0 \\ 0, & \text{Else.} \end{cases}$$

//g_c=center gradient pixel
$g_{p1} = Neighborhood\,pixel 0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ$,$180^\circ, 225^\circ, 270^\circ, 315^\circ$ for T2

S-9:Compute the Patterns
$Pt2_{ij} = bin2dec(f_1(g_{p1} \sim g_c))$
S-9:Compute the Patterns
$Pt2_{ij} = bin2dec(f_1(g_{p1} \sim g_c))$
End loop j
End loop i
S-10:Perform the bitwise OR operation between two patterns
$PT = Pt2 \bigoplus Pt1$
S-11:Construct the meshgrid for patterns
M=Meshgrid(PT)
S-12: Extract the sparse from meshgrid
S=Sparse(M)

form are extracted into separate clusters as follows:

$$C_{xy} = \begin{cases} i, & S_{xy} = S_{(x-1)(y-1)} \\ 0, & \text{else} \end{cases} \quad (6)$$

where i = 1 to N.

The pattern of image pixel at and and their distance (D_i) are used to find the coefficients of matrix called Hessian matrix (H(S,σ)) as follows:

$$H(S, \sigma) = \begin{pmatrix} L_{xx}(C, \sigma) & L_{xy}(C, \sigma) \\ L_{xy}(C, \sigma) & L_{yy}(C, \sigma) \end{pmatrix} \quad (7)$$

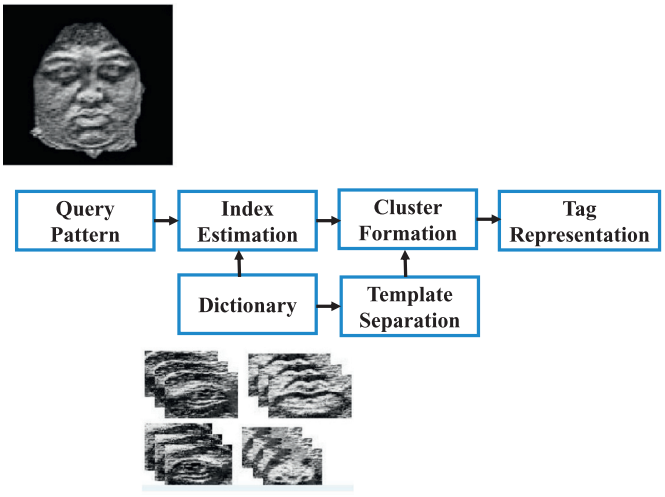


Fig. 6. STM Block Diagram.

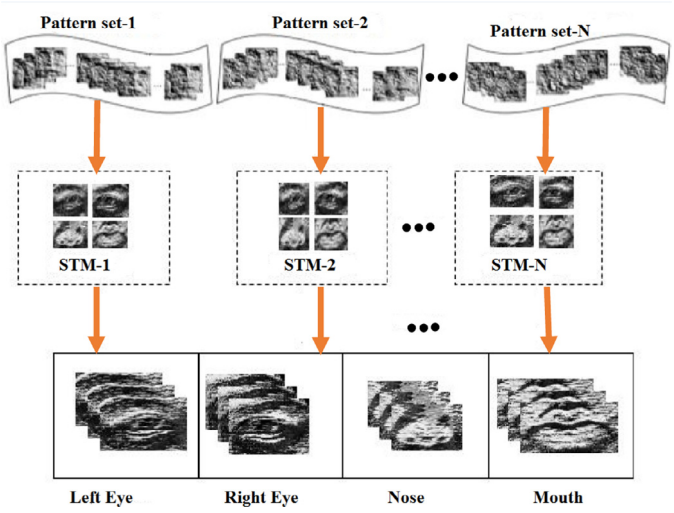


Fig. 7. STM arrangement.

where,

$$L_{xx}(C, \sigma) = \sqrt{\sum_{i=1}^{N} D_{i(PT(C)_{xy}) * \sigma}} \quad (8)$$

Then, the matching points of STM output with the testing input are predicted as follows:

$$V_s = \{\Sigma dx, \Sigma |dx|, \Sigma dy, \Sigma |dy|\} \quad (9)$$

The region corresponds to the matching points are extracted as follows:

$$R_{xy} = PT(V_s) \quad (10)$$

Finally, the features relevant to the histogram of matching points are extracted as follows:

$$F_i = B_i, where\ i = 1\ to\ N \quad (11)$$

Fig. 8 shows the pictorial representation of working process in DESURF-LOP-based facial key points prediction. With increase of features, the computations required for classification is more. Hence, the selection of number of relevant features is the prior stage to classification. Optimization procedures play the major role in feature selection process. Among various procedures, the Cuckoo

Please cite this article as: D.K. Jain et al., Multi angle optimal pattern-based deep learning for automatic facial expression recognition, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.06.025

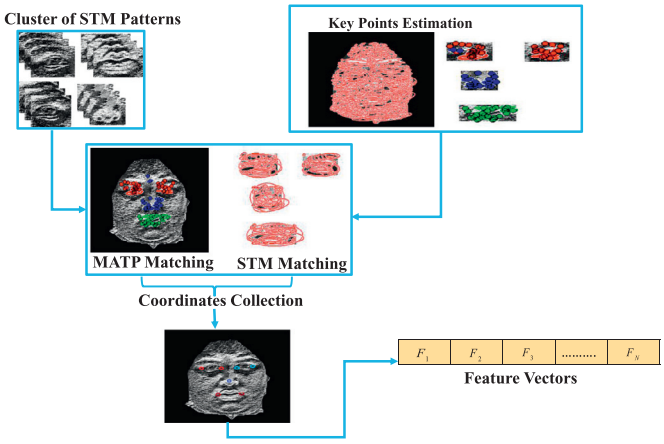


Fig. 8. DESURF-LOP Facial key point feature extraction.

Search (CS) is used to select the features necessary for classification. But, the enhanced form of CS is used in this paper to improve the classification performance further.

6. PPCSO-based feature selection

The characteristics of suitable fitness function assigning and the objective function formulation with either maximizing or minimizing objective is inherited to find the relevant features. The fitness of the solution is directly proportional to the objective function value of maximization problem. The fitness function formulation in traditional method are modified with the help of searching radius and the length of particles as

$$Ft = \{Mean(C), (k \times l)\} \tag{12}$$

Where, *C={F(1), F(2),... ... F(N)}* denotes the cuckoo particles

$$k = k + \frac{r}{(N-1) \times \Sigma C_i} \tag{13}$$

r=searching radius
N=Length of the particles

$$l = 1 - \sqrt{\frac{mean(C)}{k}} \tag{14}$$

The feature vectors are modeled as cuckoos in this CS algorithm and find the coordinates of each cuckoo is defined by using the fitness function. By using the prediction of best cluster point and the value (d) defined in (15), the radius of cluster is updated.

$$d = Min(Ft) \pm (\alpha \times (Max(Ft) - Min(Ft))) \tag{15}$$

where α=1 The index represent the Cluster Head (CH) is computed by comparing the value of current cluster with best cluster point. Once the Cluster head is formed, then the new value of coordinates to be found in order to reach the best cluster point. Mutation and crossover in Cuckoo search(CS) are responsible for updating the positions of cuckoos. The mathematical formulation of reproduction and mutation to select the cluster head and the updated radius are listed as follows:

$$X_{update}(i) = x(i-1) + ((O^{\frac{-1}{\alpha}}) * Cos(Ft)) \tag{16}$$

$$Y_{update}(n) = y(i-1) + ((O^{\frac{-1}{\alpha}}) * Cos(Ft)) \tag{17}$$

Based on the comparison between the fitness function with the probability of laying eggs defined in (18) serves the base for mutation and they described as follows:

$$C(m) = \left(1 - \frac{i-1}{(M-1)^{\frac{1}{\mu}}}\right) \tag{18}$$

$$X_{mut}(i) = x(i-1) + (C(m) \times (Max(x) - Min(x))) \tag{19}$$

$$Y_{mut}(n) = y(i-1) + (C(m) \times (Max(y) - Min(y))) \tag{20}$$

After the mutation is over, the mean value of fitness function corresponds to the new coordinates is predicted. Then, check whether the computed values are greater than zero or not. If they are greater than zero, then changes in radius is allowed and cuckoos corresponds to feature vectors are moved towards that point. Otherwise, the changes in radius are prevented. Then, the average magnitude of the fitness function is estimated which serves as the base value for selected feature set (SF) and tested feature set (TS). The Algorithm 3 for PPCSO is listed

Algorithm 3 Priority particle cuckoo search optimization.

Input: Texture Feature matrix 'F', and Testing Feature,'T'
Output: Optimal Selected Feature set for Training 'SF' and Testing 'TS'.

S-1: Initialize cuckoos and fitness value (Ft=0)
C={F(1),F(2),... ... F(N)}
S-2: Initialize the search radius as 1 (r=1) and constants ($\alpha = 0.1$ and $\beta = 0.1$)
S-3: Estimate the fitness by using (12)
S-4: Assign the coordinates of cuckoo by using the fitness value (x,y)= C(idx(Fitness))
S-5: Extract the best cluster point

$$O = \begin{cases} C_i, & \text{if } Dist(x,y) \le r \\ 0, & \text{Else} \end{cases}$$

S-6: Update the radius
r=d(1)+$\frac{O_{0tor} \times (d(2) \sim d(1))}{r}$
d=Min(Ft)±($\alpha \times (Max(Ft) - Min(Ft))$)
S-7: Select the cluster head by using the index value H=C(indx)

$$indx = \begin{cases} 1, & \text{if } (C \times e^{-\beta N}) < 0 \\ 0, & \text{Else} \end{cases}$$

S-8: Update the coordinates and compute the probability of laying eggs by using the (16) through (18)
S-9: check C(Ft)<C(m) and perform the mutation by using (19) and (20) if condition is satisfied.
S-10: Check whether the mean value of new fitness is greater than zero. Update the radius if the condition is satisfied.r=r+mean($Ft_{mutation}$),otherwise the changes in radius Unallowed.
S-11: Compute the average of fitness function
mn = $\frac{\Sigma Ft}{length(Ft)}$
S-12: Compute the index SL=idx(F(Ft>mn))
S-13: Extract the test features and relevant features SF=F(SL); TS=T(SL)

7. LSTM-CNN classification

The model proposed in this paper isolates the foreground and background prior to feature extraction. Besides, the relevant feature selection through the PPCSO reduces the dimensionality of features. Hence, the labelling of facial expression is the final stage of this work. The features (SF) and tested features (TS) from PPCSO are passed to the LSTM-CNN as shown in Fig. 9. Initially, the labels are initialized as "n" and L=1. The features corresponds to the F1

Please cite this article as: D.K. Jain et al., Multi angle optimal pattern-based deep learning for automatic facial expression recognition, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.06.025

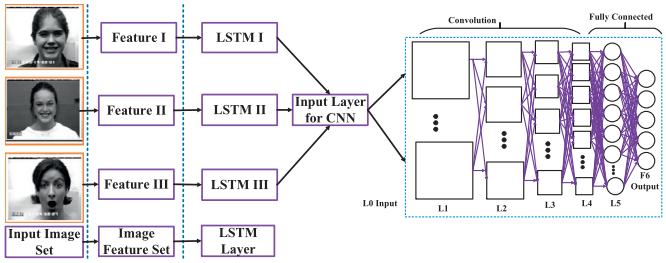


Fig. 9. LSTM-CNN classification.

are assigned as S. The maximum and the mean of selected features (SF) are computed and they can be regarded as M and N respectively. The limit of sub-divided intervals for the classification process lies in the ranges between $(1<\frac{1}{n}<N)$. Then, the rules necessary to perform the classification process are extracted as follows.

$$R = SF(M-N) * Lt \tag{21}$$

The neighbor link parameter (ρ_t) and the kernel function (K) are necessary for accurate classification

$$\rho_t = SF^{-1}TS(t) \tag{22}$$

$$K = R^{-1}\emptyset(t) \tag{23}$$

The training feature set with the neighbor link and the kernel parameter for the mapping process is constructed by

$$SF_i = K_i + \rho_i = R^{-1}\emptyset(i) + \rho_i \tag{24}$$

The probability distribution on feature set for neighboring features to update the kernel function is computed as follows:

$$\emptyset(t) = \frac{1}{(2\pi)^{\frac{n}{2}}}\frac{1}{n}\sum_{i=1}^{N_i} e^{\left[\frac{-(Tr_t-S_t)^{-1}(Tr_t-R_j)}{2\sigma^2}\right]} \tag{25}$$

Finally, the Point Kernel Classifier (PKC) checks whether the value of selected features is compared with the probability distribution for each column (t) ($S_t > \emptyset(t)$) defined in (25). The kernel function formulation for the class labels are listed as follows:

$$V_t(TS) = \Sigma_{n=1}^{N}\Sigma_{m=1}^{M}\left(\frac{\partial SF_{p,m}}{\partial t_i}TS_{p,m}\right) \tag{26}$$

If the probability distribution function is greater than count of selected features SF_t, then the corresponding labels ($C{=}L(\emptyset(t))$ are assigned to the images to indicate the expressions (anger, disgust, fear, happiness, sadness, and surprise).

8. Experiments

The proposed MAOP-DL is evaluated on two databases. The experimental details are shown and the comparative analysis is discussed in this section.

8.1. Databases

CK+ Database: The Cohn–Kanade database[9] contains the expressions of 100 university students with the age variations of 18 to 30. Among them, the 65% are female, 15% are African–American and 3% are either Asian or Latino. There are six emotions based on the prototype description are the subjects of this database. To validate the effectiveness of proposed work, image sequences from 96 subjects are selected.Fig. 10 shows some normalized samples with all expressions.

MMI Database: This database is the most challenging database [29] compared to CK+database. There are 30 students and research staff members with the age of 19 to 62 are included in this database. Among them, 44% are female with the background of European, Asian or South American ethnic background.Typical pictures of persons showing emotions can be seen in Fig. 11. There are 213 image sequences with the labels of six expressions are included in the database. To validate the proposed system, there are 205 images are selected. Table 1 Presents the abbreviations of Conventional learning methods used for Comparison.

Fig. 10. Example of six basic expressions from the Cohn–Kanade + database (anger, disgust, fear, happiness, sadness, and surprise).

Fig. 11. Example images from the MMI. The emotions from left to right are: Anger, Disgust, Fear, Happiness, Sadness, Surprise.

Table 1
Abbreviations of Methods.

Method	Description
ADL	Patches selected by ADaboost are used
AFL	All the patches of whole Face are used
CPL	Common Patches are used
CSPL	Common and Specific Patches are used

Table 2
Recognition Rate and F-1 measure analysis for CK+ database.

Expression	AFL	ADL	CPL	CSPL	MOAP-DL
Anger	0.6407	0.6281	0.7144	0.7440	0.9482
Disgust	0.8782	0.8776	0.8927	0.9134	0.9558
Fear	0.8235	0.8206	0.8209	0.8432	0.9487
Happiness	0.9416	0.9381	0.9305	0.9462	0.8592
Sadness	0.8204	0.8346	0.8515	0.8619	0.8816
Surprise	0.9806	0.9791	0.9827	0.9870	0.9440
Recognition Rate	0.8694	0.8226	0.8842	0.8989	0.9617

8.2. Results comparisons

We investigate the performance of Proposed MOAP-DL over the existing single scale methods regarding the recognition rate and F1-measure as in Table 2 for CK+ database in this section. In existing methods, the CSPL provides the better F1 measures of 0.7440, 0.9134, 0.8432, 0.9462, 0.8619 and 0.9870 for the expressions of anger, disgust, fear, happiness sadness and surprise respectively. The recognition rate of CSPL [37] is 0.8989 which is also higher than the existing methods. But, the proper alignment and optimized texture features in proposed MAOP-DL increased the F1-measure to 0.9482, 0.9558, 0.9487, 0.8592, 0.8816 and 0.9440 respectively. Besides, the recognition rate also improved to 0.9617. The comparative analysis shows that the proposed MAOP-DL improves the F1-measure by 20.42(anger), 4.24(Disgust), 10.55(Fear),1.97(sadness) compared to CSPL respectively,however the expression of Happy and surprise failed to recognise accurately. The recognition rate for MAOP-DL also 6.28% higher than the CSPL method.

Table 3 presents the comparative analysis of recognition rate analysis of proposed and the existing methods (sparse+A+T, sparse+A and conventional+A) [4] various expressions. In exist-

Table 3
Recognition Rate analysis for MMI database.

Expression	Conventional+A	Sparse+A	Sparse+A+T	MOAP-DL
Anger	92.2	93.8	95.8	98.67
Disgust	95.1	97.1	97.9	98.82
Fear	93.9	96	96.5	98.96
Happiness	97.5	97.6	98.2	98.67
Sadness	91.9	92.3	94.5	98.52
Surprise	94.5	95.3	97	98.67

Table 4
Recognition Accuracy Analysis for CK+ database.

Methods	CK+
GSNMF [36]	72.17
CSPL [38]	89.9
DSNGE [10]	94.82
Ours	96.17

Table 5
Recognition Accuracy Analysis for MMI database.

Methods	MMI
CSPL [38]	73.53
3D-CNN [3]	95
CNN [14]	97.81
Ours	98.72

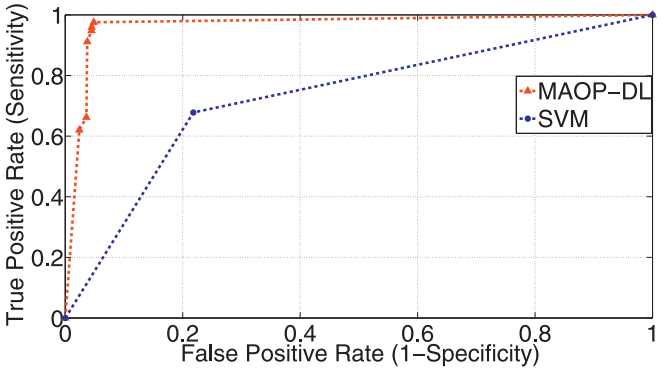
ing methods, the sparse+A+T provides the better recognition rate of 95.8, 97.9, 96.5, 98.2, 94.5 and 97% for the expressions of anger, disgust, fear, happiness sadness and surprise respectively. But, the proper alignment and optimized texture features in proposed MAOP-DL increased the recognition rate to 98.6745, 98.8218, 98.9691, 98.6745, 98.5272 and 98.6745% respectively. The comparative analysis shows that the proposed MAOP-DL improves the recognition rate by 2.91, 0.92, 2.51, 0.48, 4.02 and 1.7% compared to sparse+A+T respectively.

We compare the accuracy of recognition with the following methods to show the effectiveness of MAOP-DL: Graph-preserving Sparse NMF (GSNMF) [36], Common and Specific Patches (CSPL) [38] and Dual Sub-space NGE(DSNGE) [10] in Table 4.

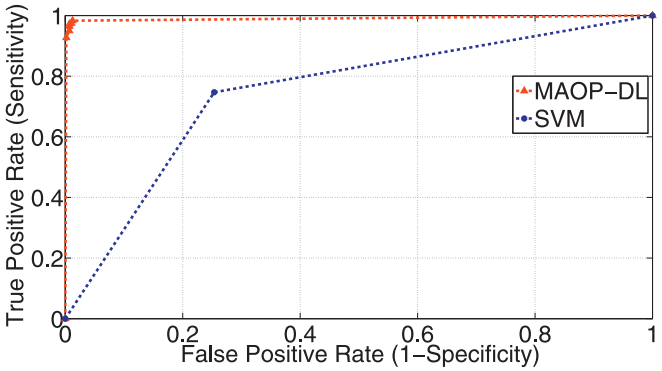
In existing methods, the DSNGE provides better performance (94.92%) than the existing methods. The foreground extraction prior to the pattern extraction and feature selection and the enhanced STM models with the multi-angular patterns in proposed MAOP-DL further improves the recognition accuracy to 96.17%. The comparison between the MAOP-DL with the DSNGE shows that the MAOP-DL provided the 1.3% improvement compared to DSNGE models respectively.

In existing methods, the CNN provides better performance (97.81%) than the existing methods. The foreground extraction prior to the pattern extraction and feature selection and the enhanced STM models with the multi-angular patterns in proposed MAOP-DL further improves the recognition accuracy to 98.72% as shown in Table 5. The comparison between the MAOP-DL with the CNN shows that the MAOP-DL provided the 0.91% improvement compared to CNN models respectively.

Fig. 12(a) and (b) shows the ROC performance analysis of proposed MAOP-DL with the existing SVM for CK+ database and MMI database respectively. From the Fig. 12(a) and (b), it is observed that the proposed MAOP-DL provides the high true positive rate for small values of false positive rate due to the angle-based texture pattern and optimal selected features compared to SVM models respectively for CK+ and MMI database.



(a) CK+ database



(b) MMI database

Fig. 12. ROC analysis of (a) CK+ database and (b) MMI database.

9. Conclusion

This paper proposed the MAOP-DL for the clear image analysis and effective learning of facial expressions. The selection of key point features based on intensity difference analysis, relevant feature selection and the neural network based classification carried out in proposed MAOP-DL enhanced the recognition performance compared to the existing methods. The effectiveness of proposed MAOP-DL is validated over the number of state-of-art methods on CK+ and MMI database with the parameters of recognition rate and F1- measure. The novel feature extraction based on the multi angle and the noise removal through the EBBS contributed towards the illumination variation handling and the matching accuracy effectively.The major limitations of proposed work are less performance in involuntary expression handling,low intensity and short duration expressions are not handled,Spot and recognition of muscle variations,due to these limitations observed in short duration FER analysis,Future work should focus on the analysis of microexpression(involuntary expression with low intensity and short duration) by extending the multi-angle-based feature descriptors to spot and recognize it.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (2016YFB1001005), the National Natural Science Foundation of China (Grant No. 61473290, Grant No. 61673375), the National High Technology Research and Development Program of China (863 Program) under Grant 2015AA042307, and the Projects of Chinese Academy of Science (Grant No. QYZDB-SSW-JSC006, Grant No. 173211KYSB20160008).

References

[1] E.I. Barakova, R. Gorbunov, M. Rauterberg, Automatic interpretation of affective facial expressions in the context of interpersonal interaction, IEEE Trans. Hum. Mach. Syst. 45 (4) (2015) 409–418.
[2] P.R. Borude, S. Gandhe, P. Dhulekar, G. Phade, Identification and tracking of facial features, Procedia Comput. Sci. 49 (2015) 2–10.
[3] Y.-H. Byeon, K.-C. Kwak, Facial expression recognition using 3D convolutional neural network, Int. J. Adv. Comput. Sci. Appl. 5 (12) (2014).
[4] Y. Guo, G. Zhao, M. Pietikäinen, Dynamic facial expression recognition with atlas construction and sparse representation, IEEE Trans. Image Process. 25 (5) (2016) 1977–1992.
[5] S. Hadfield, R. Bowden, Hollywood 3d: recognizing actions in 3d natural scenes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3398–3405.
[6] S. Happy, A. Routray, Automatic facial expression recognition using features of salient facial patches, IEEE Trans. Affective Comput. 6 (1) (2015) 1–12.
[7] A. Iosifidis, A. Tefas, I. Pitas, Graph embedded extreme learning machine, IEEE Trans. Cybern. 46 (1) (2016) 311–324.
[8] M. Jampour, V. Lepetit, T. Mauthner, H. Bischof, Pose-specific non-linear mappings in feature space towards multiview facial expression recognition, Image Vis. Comput. 58 (2017) 38–46.
[9] T. Kanade, J.F. Cohn, Y. Tian, Comprehensive database for facial expression analysis, in: Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on, IEEE, 2000, pp. 46–53.
[10] H.-W. Kung, Y.-H. Tu, C.-T. Hsu, Dual subspace nonnegative graph embedding for identity-independent expression recognition, IEEE Trans. Inf. Forensics Secur. 10 (3) (2015) 626–639.
[11] K. Li, F. Xu, J. Wang, Q. Dai, Y. Liu, A data-driven approach for facial expression synthesis in video, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 57–64.
[12] M. Liu, S. Shan, R. Wang, X. Chen, Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1749–1756.
[13] A.T. Lopes, E. de Aguiar, A.F. De Souza, T. Oliveira-Santos, Facial expression recognition with convolutional neural networks: coping with few data and the training sample order, Pattern Recognit. 61 (2017) 610–628.
[14] A.T. Lopes, E. de Aguiar, T. Oliveira-Santos, A facial expression recognition system using convolutional networks, in: Graphics, Patterns and Images (SIBGRAPI), 2015 28th SIBGRAPI Conference on, IEEE, 2015, pp. 273–280.
[15] X. Ma, H. Song, X. Qian, Robust framework of single-frame face superresolution across head pose, facial expression, and illumination variations, IEEE Trans. Hum. Mach. Syst. 45 (2) (2015) 238–250.
[16] K. Niinuma, H. Han, A.K. Jain, Automatic multi-view face recognition via 3d model based pose regularization, in: Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on, IEEE, 2013, pp. 1–8.
[17] S.-K. Oh, S.-H. Yoo, W. Pedrycz, Design of face recognition algorithm using PCA-LDAcombined for hybrid data pre-processing and polynomial-based RBF neural networks: design and its application, Expert Syst. Appl. 40 (5) (2013) 1451–1466.
[18] X. Peng, Q. Hu, J. Huang, D.N. Metaxas, Track facial points in unconstrained videos, arXiv:1609.02825 (2016).
[19] H. Qayyum, M. Majid, S.M. Anwar, B. Khan, Facial expression recognition using stationary wavelet transform features, Math. Prob. Eng. 2017 (2017).
[20] O. Rudovic, V. Pavlovic, M. Pantic, Context-sensitive dynamic ordinal regression for intensity estimation of facial action units, IEEE Trans. Pattern Anal. Mach. Intell. 37 (5) (2015) 944–958.
[21] G. Sandbach, S. Zafeiriou, M. Pantic, D. Rueckert, Recognition of 3D facial expression dynamics, Image Vis. Comput. 30 (10) (2012) 762–773.
[22] E. Sarıyanidi, V. Dağlı, S.C. Tek, B. Tunc, M. Gökmen, Local zernike moments: a new representation for face recognition, in: 2012 19th IEEE International Conference on Image Processing, IEEE, 2012, pp. 585–588.
[23] E. Sariyanidi, H. Gunes, A. Cavallaro, Automatic analysis of facial affect: a survey of registration, representation, and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (6) (2015) 1113–1133.
[24] T. Senechal, V. Rapp, H. Salam, R. Seguier, K. Bailly, L. Prevost, Facial action recognition combining heterogeneous features via multikernel learning, IEEE Trans. Syst. Man, Cybern. Part B (Cybern.) 42 (4) (2012) 993–1005.
[25] C. Shan, Smile detection by boosting pixel differences, IEEE Trans. Image Process. 21 (1) (2012) 431–436.
[26] C. Shan, Smile detection by boosting pixel differences, IEEE transactions on image processing 21 (1) (2012) 431–436.
[27] M.H. Siddiqi, R. Ali, A.M. Khan, Y.-T. Park, S. Lee, Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields, IEEE Trans. Image Process. 24 (4) (2015) 1386–1398.
[28] P. Vageeswaran, K. Mitra, R. Chellappa, Blur and illumination robust face recognition via set-theoretic characterization, IEEE Trans. Image Process. 22 (4) (2013) 1362–1372.
[29] M. Valstar, M. Pantic, Induced disgust, happiness and surprise: an addition to the MMI facial expression database, in: Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect, 2010, p. 65.
[30] L. Wang, K. Wang, R. Li, Unsupervised feature selection based on spectral regression from manifold learning for facial expression recognition, IET Comput. Vis. 9 (5) (2015) 655–662.
[31] B. Yao, L. Fei-Fei, Recognizing human-object interactions in still images by modeling the mutual context of objects and human poses, IEEE Trans. Pattern Anal. Mach. Intell. 34 (9) (2012) 1691–1703.
[32] X. Yu, J. Yang, L. Luo, W. Li, J. Brandt, D. Metaxas, Customized expression recognition for performance-driven cutout character animation, in: Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on, IEEE, 2016, pp. 1–9.
[33] S. Zhang, J. Huang, H. Li, D.N. Metaxas, Automatic image annotation and retrieval using group sparsity, IEEE Trans. Syst. Man, Cybern. Part B (Cybern.) 42 (3) (2012) 838–849.
[34] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D.N. Metaxas, X.S. Zhou, Towards robust and effective shape modeling: sparse shape composition, Med. Image Anal. 16 (1) (2012) 265–277.
[35] X. Zhao, E. DellandréA, J. Zou, L. Chen, A unified probabilistic framework for automatic 3Dfacial expression analysis based on a Bayesian belief inference and statistical feature models, Image Vis. Comput. 31 (3) (2013) 231–245.
[36] R. Zhi, M. Flierl, Q. Ruan, W.B. Kleijn, Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition, IEEE Trans. Syst. Man, Cybern.Part B (Cybern.) 41 (1) (2011) 38–52.
[37] L. Zhong, Q. Liu, P. Yang, J. Huang, D.N. Metaxas, Learning multiscale active facial patches for expression analysis, IEEE Trans. Cybern. 45 (8) (2015) 1499–1510.
[38] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, D.N. Metaxas, Learning active facial patches for expression analysis, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 2562–2569.

Please cite this article as: D.K. Jain et al., Multi angle optimal pattern-based deep learning for automatic facial expression recognition, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.06.025