Random walk-based feature learning for micro-expression recognition

Deepak Kumar Jain, Zhang Zhang, Kaiqi Huang

Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation CAS, University of Chinese Academy of Sciences, Beijing, China.

ABSTRACT

Facial expression recognition (FER) and its analysis becomes an attractive research study in the fields of computer vision applications and pattern recognition. These facial expressions are generally categorized into two kinds such as micro and macro-expressions. To detect the macro-expression effectively, an angle based pattern extraction models and Markov models are employed. But the micro-expression delivers more detailed information than the macro-expression. Other difficulties such as short durations and rapid spontaneous facial expression are induced due to the detection and analysis of the micro-expression. To solve these challenges, we propose the three novel techniques such as Active Shape Modeling (ACM), Random Walk (RW) and the Artificial Neural Network (ANN) which helps to improve the overall performance effectively. The key points from the facial expression over the video frames are predicted using ASM and are spatially associated with the original face through the procrustes analysis. Then the RW algorithm is used to learn the training features prior to ANN model. This RW is integrated with ANN model to improve the learning performance of micro-expression with minimum computation complexity. The experimental validation on two spontaneous micro-expression datasets such as Chinese Academy of Sciences Micro-Expression (CASME) and Spontaneous Micro-expression (SMIC) over the existing SVM classifiers shows its effectiveness in automatic micro-expression learning applications.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Facial Expression Recognition (FER) is an excellent source of non-verbal communication cues which defines the process of identifying the deformation in facial features and movements. The reproduction of distinct spatio-temporal pattern of expressions and the explicit emotion discrimination task are the major tasks that require the diagnostic feature extraction(Zhang et al., 2016). Due to the tie up of diagnostic feature extraction with the emotions, the time consumption for the recognition is more and hence the suitable framework is needed to reduce the time consumption. The early processing of happy expression requires the latest semantic processing of emotional content of face(Neath-Tavares and Itier, 2016). But, the signs from the mouth and the eyes have the great impact on the fearful content of the facial image. The examination of time requirement for the facial expression processing through the Event Related Potentials (ERPs) is the temporal dynamics of neuro-cognitive events (Khan et al., 2016). Face recognition, expression analysis and the facial units detection require the fundamental task called fitting facial landmarks on sequential images (Peng et al., 2016). The learning of part-based representation is required to model the facial shape and appearance effectively. The joint adaptation of filtering and the representation models is the major requirement to detect the misalignment through the deep neural network formulation. Fast network training is the major aspect in the neural network formulation. The relative new algorithm called Extreme Learning Machine (ELM) (Iosifidis et al., 2016) is used to reduce the time consumption and the low human supervision. The weight and bias values of each neuron are assigned randomly through the global approximates. The exploitation of sub-space learning (Wang et al., 2015) for the ELM process projects the linear and non-linear data effectively. The classification of facial expression comprises feature extraction and selection. The major approaches to find the facial features geometric-based and appearance-based methods(Lajevardi and Hussain, 2012). The utilization of data scaling in the preprocessing stage make the data reduction techniques such as Principal Component Analysis (PCA) as the invariant. The invariance towards the illumination changes and the simple computations through the Local Binary Patterns (LBPs) require...
an extension towards the volume-based LBP. The lightweight descriptor on the basis of the six-Interconnection points (Wang et al., 2014b) provides the consistent matching with the minimum computational complexity.

The afore-mentioned methods do not consider the color which is the fundamental aspect of the human perception in behavior analysis. The Tensor Discriminant Color Space (TDCS) model (Wang et al., 2014a) is used to denote the third-order tensor which provides more robustness to the noise variations. The illumination changes and cast shadows require the modification and the replacement of correlation values with the pixel intensities effectively (Tzimiropoulos et al., 2012). The explicit mapping of linear correlations with the gradient orientation is the major issues observed from the overview of traditional FER methods are the feature dimensionality and the trade-off between dimensionality and the recognition performance. This paper proposes the suitable framework by integrating the ASM with the RW-ANN classifiers. The novel contributions of proposed work are listed as follows:

- The incorporation of ASM model in procrustes analysis extracts the facial key points to describe the variations of facial deformations and the muscle variations effectively.
- The prior learning of features through the RW block reduces the feature dimensionality.
- The integration of RW with the ANN classifier highly contributes to the accurate feature learning with the random analysis.
- The merging of feature attributes with the mixture of feature set attributes corresponding to video frames through the RW-ANN improves the learning performance.

The organization of the paper comprises into nine sections. Section 2 describes the traditional FER techniques related to the dimensionality and deformations respectively. Section 3 illustrates the overall flow of the proposed ASM-RW-ANN for an accurate recognition. Section 4 illustrates the face detection-based on maximum block size in detail. Section 5 discusses the procrustes analysis using the ASM. Section 6 presents the RW-based feature learning and Section 7 describes the ANN-classification model. Section 8 discusses the evaluation of proposed work against the existing methods corresponding to the CASME dataset. Finally, the entire discussion of this paper concludes at section 9.

2. Related Work

In this section, the review of existing facial expression recognition techniques highlights the major issues. Ali et al (Ali et al., 2016) presented the boosted Neural Network Ensemble (NNE) collections-based technique for the multicultural facial expression recognition. The employment of Nave Bayesian (NB) classifier in the boosted technique predicts the number of NNE collections. The reduction of redundant information from the Local Binary Pattern (LBP) based plane formulation improved the recognition performance further. Huang et al (Huang et al., 2016) proposed the Spatio-temporal model with the Completed Local Quantized Pattern (CLQP) to analyze the micro-expressions. description of both appearance and motion information required the spatiotemporal monogenic binary patterns. Huang et al (Huang et al., 2012) utilized the monogenic signal analysis to extract the magnitude of the real and imaginary pictures which oriented with the frames. The integration of concatenation with the kernel learning contributed to the effective handling of features. Wong and Tan (Liong et al., 2016) introduced the automatic detection and recognition of spontaneous facial expressions using optical-strain information. The optical flow method was used to extract the features which were used to construct the histogram effectively. Nicolle et al (Nicolle et al., 2016) addressed three issues such as classification problem in Action Unit (AU), poor running ability for real-time applications and activation of AU for natural behavior.

The subtle head movements and the unconstrained lighting conditions affected the FER task adversely. Feng et al (Xia et al., 2016) presented the random walk model to calculate the probability of frame and the correlation between the frames. The discrimination of high-stakes from the true facial expression was the major issue in the Adaboost model. Su and Levine (Su and Levine, 2016) adopted 2D appearance model to characterize the 3D facial features. Instead of built from the 3D model, the extraction of invariant 2D facial features provided the accurate recognition effectively. Liu et al (Liu et al., 2014) presented the manifold modeling to solve the afore-mentioned issues. The three major components were involved in this work were Spatio-Temporal Manifold (STM) modeling of video frames, Universal manifold Model (UTM)-based feature learning and the fitting of local modes of STM in UTM. The discovery of common and specific patches through the exploration of commonalities was the major issue in STM-UTM models. Zhong et al (Zhong et al., 2015) proposed the two-stage sparse learning model with the prior knowledge of facial muscles in the AU. The employment of multi-scale face division strategy extracted the different facial patches under different coverage areas to represent the expression. But, this property was not adaptable in the frame-based methods. Siddiqui et al (Siddiqui et al., 2015) integrated the step-wise Linear Discriminant Analysis (LDA) with the Hidden Conditional Random Fields (HCRF) in the sequence-based FER system. The localized features extracted using this approach were minimum compared to previous method. Kamarol et al (Kamarol et al., 2016) proposed the appearance-based feature extraction model through the Spatio-Temporal Texture Map (STTM) to capture the subtle spatial and temporal variations. The dynamic feature extraction by using the block-based method construct the required histogram. High-quality database with the minimum training samples was the major issue in the traditional studies. Yan et al (Yan et al., 2014) reviewed the previous developed micro-expression databases with the high spatio and temporal resolutions. Shen et al (Shen et al., 2012) performed two experiments such as prediction of effect of the duration of expression on the recognition performance and the upper limit of micro-expressions. Sun et al (Sun et al., 2014) formed the objective function which included the non-local smoothness term and the information of boundaries to preserve the motion details effectively. The review of existing literatures stated that the maximum error rate, bulky feature set for classification are major problems in the
3. Overall flow of proposed ASM-RW-ANN based micro-expression recognition

In general, a micro-expression is described as an involuntary pattern of the human body that is significant enough to be observable, but may not fully convey the triggering emotion (Godavarthy, 2010). Micro-expression occurring on the face are rapid and are often missed during casual observation. Lasting from 1/25th to 1/3rd of a second (roughly 2-10 frames) (Ekman et al., 1992), micro-expressions can be classified, based on how an expression is modified, into three types (Ekman, 2009):

- Type 1. Simulated Expressions: When a micro-expression is not accompanied by a genuine expression.
- Type 2. Neutralized expressions: When a genuine expression is suppressed and the face remains neutral.
- Type 3. Masked Expressions: When a genuine expression is completely masked by a falsified expression.

Type 2 micro-expressions are not observable and type 3 micro-expressions may be completely eclipsed by a falsified expression. In this paper, we focus on type 1 micro-expressions, i.e., micro-expressions that correspond to rapid, but observable and non-suppressed motion on the face.

An accurate automated recognition of micro-expressions with the facial deformations, muscle variations and the extraction of the facial key points are the major focus of the research work proposed in this paper. Figure 1 shows the overall workflow of proposed ASM-RW-ANN-based micro-expressions. Initially, the RGB image is constructed from the set of frames in the real-time video database. The input RGB image is loaded to the proposed system. The detection of face is the preliminary stage in proposed work where the skin intensity is estimated and the Region of Interest (ROI) is tracked. Then, the ASM is applied to the extracted region for procrustes analysis. The matching between the initialized coordinates with the extracted coordinates searched the relevant key points of the facial image. The feature vector is constructed from the distance estimation through the geometrical measurement in the feature extraction block. The relevancy among the features corresponds to the key points which are the major focus of the research work proposed in this paper. An accurate automated recognition of micro-expressions with the facial deformations, muscle variations, and the extraction of the facial key points are the major focus of the research work proposed in this paper.

The major operational modules in proposed work are listed as follows:

- Skin-intensity-based face detection.
- ASM-based procrustes analysis.
- RW-feature learning.
- ANN-classification.

4. Skin Intensity-Based Face Detection

Initially, the face detection from the input frames is the major stage in expression analysis. The preliminary stage in face detection is to acquire an image and apply the preprocessing to remove the noise from the images. An object detection, tracking, and recognition requires the different color spaces such as RGB, HSV and YCrCb. In this paper RGB color space is used to measure the pixel intensity variations corresponding to the skin region. It is possible to detect the face just by cropping directly. But it does not detect the face accurately. The face detection using skin pixels provide variation in the skin region which vary the distance. Hence we can detect the face accurately. This face detection based on the intensity of the skin region increases the accuracy which leads to an increase in the performance. Hence we use skin intensity-based face detection in this proposed work. The extraction of probable regions in order to reduce the search space depends on the chrominance values of skin color pixel intensities. The basic objective of the skin color classification is to determine whether the color pixel has the color of human skin or not and hence the skin color classification is employed in the proposed work to alleviate the skin tone variations. Then, the location of face or ROI is tracked to locate the key points finally. The workflow of the skin intensity-based face detection is graphically depicted in Figure 2. The
pends on RGB color space as illustrated in equation (1) present in the input images. The skin pixel intensities related to the RGB-B color space are compared with the respective limits of 95, 40 and 20 respectively to assign the label.

\[ V = [R, G, B] \] (1)

The probable skin regions are identified on the basis of the skin color-based thresholding in morphological operation. Due to the application of morphological operation on the intensity images, the color image is necessarily converted into the grayscale image. The break-up of dark regions into the number of small regions through the intensity thresholding constitutes the morphological opening. The removal of small objects and the preservation of shape and size of the objects which are exactly less below the size of a face. In this paper, the morphological opening on the gray scale image is illustrated in Figure 3. Then, the comparison among the pixel intensities with the threshold value (on the basis of skin color) generates the mask image. The input image is considered as the image 1 and the grayscale image from the RGB-based skin intensity estimation is regarded as the image 2. Then, the comparison between the image 1 and 2 provides the segmented region of the facial image. Based on the segmented face region, the maximum blob size is used to track the region of interest and finally the face is detected as shown in Figure 4 (a) and (b).

\[ p_i = (x_i, y_i) \] (2)

\[ q = [p_1, p_2, ..., p_n]^T \] (3)

The ASM model construction depends on the manual drawn contours and finds the main variations in the training data in the Principal Component Analysis (PCA). The set of matrices that describe the texture of the lines perpendicular to the control point. Initially, the shape instance from the tracked ROI is modeled as judiciously set of points namely

\[ p_i = (x_i, y_i) \] (2)

The set of n feature points are stacked into the long vector as follows:

\[ q = [p_1, p_2, ..., p_n]^T \] (3)

The utilization of procrustes algorithm through Active Shape Model conveys that the sum of distances to the mean of each shape is minimized. The Algorithm 1 is to estimate the ASM-based procrustes analysis. In this paper, the ASM requires an approximate of any instance of the shape which includes the training instances with the projection of the first eigenvectors

\[ q = \tilde{q} + \sum_{i=1}^{l} b_i u_i \] (4)

The promoted ASM model extracts the landmarks of face

Fig. 3. Directionality-Based ROI Extraction.

Fig. 4. (a)Segmented face region (b)Face detected output.

Fig. 5. ASM-based procrustes Analysis.

5. ASM-Based Procrustes Analysis

The tracked ROI is passed to the ASM block to identify the matching between the coordinates initialization and the extract-

Algorithm 1 ASM-based procrustes

Input: Tracked ROI
Output: Active Shape Model(ASM)
S-1: Translate the shape instance for the input ROI as the stacked vectors (q)
S-2: Select the one example as the mean shape \( \tilde{q} \)
S-3: Record the estimates \( q_0 \) for default reference frame.
S-4: Align all the instances of the shape with the current estimate.
S-5: Compute the mean for aligned instances.
S-6: Check the condition \( \sum_{n}(q_i - \tilde{q})^2 = 0 \)
S-7: Rescale and reiterate the process until the condition is satisfied.
shapes. The robust and rapid matching of shape model through the multi-resolution approach searched the landmarks on the pyramid shape of the input image. The total landmarks located on the facial image are 77 in ASM model. But, the small head movements induce the changes in the shape points in ASM. Hence, the procrustes analysis is employed to align the shape points in frames. Figure 6 (a), (b) and (c) show the ASM matching lines, binary representation and the procrustes points.

In procrustes analysis, the normalization of orientation, scale and the translation parameters help to align the shape points to the original shape as in Figure 7. Once the facial key points are extracted, then the feature extraction process is necessary. The representation of image or video in the form of the sequential information refers feature extraction process. In this paper, the geometrical information of the key points is regarded as the features corresponding to the expression in each frame.

The distance between the facial key points as shown in Figure 7 provides the geometrical information. In Figure 8, the points 1 and 2 represent the facial key points from the procrustes analysis. The distance between the points 1 and 2 from the origin is denoted by the h1 and h2. Then, the angle between these points is denoted by ϕ. The distance and angle are regarded as the geometrical features for classification. In this paper, the origin is considered as the centroid of key points and the geometrical features are extracted for current frame in the video from the Figure 9. Finally, the features are aggregated into the feature vector (F).

6. RW-based feature learning

Image annotation and retrieval processes are based on the suitable learning and this paper utilizes the Random Walk(RW) model(Xia et al., 2016) to compute the probability of the frames containing micro expressions with the correlation among the frames in the temporal window through the blocks as shown in figure 9. This RW algorithm is majorly used in the fields of image segmentation, computer networks, vision science, wireless networks, population genetics, polymer physics, and mathematical ecology. For i-th frame in a video sequence, the L frames are extracted and make them as the temporal window centered on this frame. The RW process is formulated as:

$$p_{t+1}(i) = \alpha \sum_{j \in \Omega_j} p_t(j) \phi(i, j) + (1 - \alpha) p_0(i)$$

where, $\Omega_j$=Temporal neighbor set (the neighbor set adjacent to the i-th frame).
The transition probability is computed by using the deformation similarity as follows:

$$\phi(i, j) = \frac{p_0(j) s_{ij}}{\sum_{k \in \Omega} p_0(k) s_{jk}}$$  \hspace{1cm} (9)$$

where $s_{ij}$=Geometrical deformations among the frames.

The Algorithm to apply the labels based on RW is expressed as follows:

**Algorithm 2 Random Walk Algorithm**

**Input:** Feature set(F)

**Output:** Label(L)

S-1: Initialize the seed particles as feature set(F).Initialize the values.

S-2: Compute the weight of initialized particles from feature set

$$W_k = \exp(-\sigma((1-\tau)\|I(k)-I(j)\|^2 + \sigma(p(k)-p(j))^2))$$

where, $\sigma, \tau$=Adjusting parameter,$I$=original image,$k,j$=indices of pixel intensities,$p$=probabilistic image.

S-3: Estimate the likelihood estimation is based on the minimization function as follows:

$$\arg \min_E = \arg \min x^T L$$

$$L = \begin{cases} 
\sum_j W_{kj}, k = j \\
-\sum_j W_{kj}, k \neq j \\
0, \text{ otherwise}
\end{cases}$$

S-4: Compute the best relevancy based on minimization function.

S-5: Cluster label L=1 for best relevancy and 0 for others.

expression is computed and then the binary operation is performed on probabilities of frames on the basis of the threshold value predicts the presence of micro-expression and obtains the consecutive frame clips. The probability value of frames containing micro-expression is modelled as the weight value of the particles. The particle for RW is initialized from the feature vector. The relevancy between the each particle is computed to construct the cluster. During each iteration, the best relevancy among the particles is predicted that form the cluster data and they are labelled accordingly as shown in Figure 10. Figure 11 shows the relationship between the input particles and the clustered particles prior and after RW. Among the feature set, the best selected feature set necessary to form the cluster is pictorially represented in figure 11. The total number of random walks performed in RW model are 136 and the geometrical deformation of the particles is noted accordingly. Once the final iteration or random walk is performed, all the particles are aggregated into the origin. Initially, the particles of RW are scattered (in green color) and they are aggregated sequentially towards the center until the final iteration count is reached. The particles illustrated in blue color indicates the feature set which forms the cluster. However, RW is applied to learn the features for proper labelling.

7. ANN Classification

The RW data is passed to the Artificial Neural Network (ANN) which includes the kernel model, network formation, feature matching, and the classification. The extracted features from the each frame of the input video are classified using this model. The learned features from the RW model are passed to the kernel model. The cluster is formed with the label. The matching between the testing feature set and the learned set are predicted. The relevance of testing vector and the label are assigned based on the relevancy. The labeling of facial expression is done according to the relevancy metric. Figure 12 shows the ANN formulation for the proposed work.

The network of neurons is constructed with the feature set. The clustered attributes that described the feature set with the learned feature set forms the hidden layer and the weight value
corresponds to the neuron update is estimated. ANN with the prior RW learning is used to compute the classified output with the testing vectors. The output \( y \) is modelled as:

\[
y = MLP(x, w)
\]

(10)

where, \( x = (F_1, F_2, ..., F_n) \) The MLP uses the gradient back propagation algorithm for training to update the weights and transfer the output into the fitness function as follows:

\[
f(t) = \frac{1}{1 + e^{-t}}
\]

(11)

The maximum and the mean of selected features (x) are computed and they can be regarded as M and N respectively. The rules necessary to perform the classification process of the RW learning features \( (L_{RW}) \) are extracted as follows:

\[
R = L_{RW}(M - N) * LT
\]

(12)

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

\[
k_i = \frac{1}{N_i} \sum_{i=1}^{N_i} e^{\frac{-1}{\sigma^2}(g_{iu} - g_{ij}^2)^2}
\]

(13)

The estimate of classified output for each level of hidden layers as follows:

\[
y_j = \sum \beta_j L_{i-1} w_{ij}
\]

(15)

From the output samples, the ANN model provides the recognized expressions from the database. The rearranging of feature set and the learning set form the subset which is necessary for the classification.

8. Experimental Evaluation

The proposed ASM-RW-ANN is evaluated on two spontaneous databases. The experimental details are shown and the comparative analysis is discussed in this section.

8.1. Experimental setup

Two spontaneous micro-expression datasets are used to evaluate the performance of our proposed approach in our experiments: CASME dataset Yan et al. (2013) and SMIC dataset Li et al. (2013). Both of them are designed to detect and recognize micro-expressions, which are constructed by inducing subjects micro-expressions.

- The Chinese Academy of Sciences Micro-Expression (CASME) Yan et al. (2013) contains the micro-expressions of 195 under 60 fps. The facial movements elicited are ranges from 1500 and the corresponding samples are coded with the onset, apex and offset frames with the AU the micro-expressions with the duration of below the 500 ms are selected from the database. For our proposed work, there are 19 number of subjects used which contains the 8 different expressions of both frames and videos.

- High Speed (HS) data are segmented and labeled by using the two annotators on the basis of the 164 self-reported emotions of the 16 participants to construct the Spontaneous Micro-expressions (SMIC) database Li et al. (2013). The high-emotional clips are constructed under an interrogation room setting with the punishment threats. The 60% of Images are used for training and the remaining 40% used for testing to validate the performance of proposed work on different methods.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMIC</td>
<td>Strain Pattern(SP)</td>
<td>0.5371</td>
</tr>
<tr>
<td></td>
<td>Feature Difference Analysis(FDA)</td>
<td>0.8229</td>
</tr>
<tr>
<td></td>
<td>Adaboost</td>
<td>0.7131</td>
</tr>
<tr>
<td></td>
<td>RW-SVM</td>
<td>0.7951</td>
</tr>
<tr>
<td></td>
<td>RW-Adaboost</td>
<td>0.8693</td>
</tr>
<tr>
<td></td>
<td>Ours</td>
<td>0.8812</td>
</tr>
<tr>
<td>CASME</td>
<td>Strain Pattern(SP)</td>
<td>0.7076</td>
</tr>
<tr>
<td></td>
<td>Feature Difference Analysis(FDA)</td>
<td>0.8785</td>
</tr>
<tr>
<td></td>
<td>Adaboost</td>
<td>0.7384</td>
</tr>
<tr>
<td></td>
<td>RW-SVM</td>
<td>0.8689</td>
</tr>
<tr>
<td></td>
<td>RW-Adaboost</td>
<td>0.9208</td>
</tr>
<tr>
<td></td>
<td>Ours</td>
<td>0.9456</td>
</tr>
</tbody>
</table>

Table 1 summarizes the AUCs of various approaches on SMIC, CASME datasets. The local binary patterns and the optical flow algorithm utilization in FDA and SP provides minimum AUC. The consideration of deformable correlations among the frames in RW models boost the AUC performance compared to the traditional models. Among the existing models, the RW-Adaboost provides the AUC of 0.8693 and 0.9208 for SMIC and CASME datasets respectively. But, the proposed ASM-RW-ANN improves the AUC performance into 0.8812 and 0.9456 respectively. The comparative analysis between the proposed ASM-RW-ANN with the existing RW-Adaboost states that the RW-ANN offers 1.35 and 2.62 % improvement compared to RW-Adaboost models for SMIC and CASME datasets respectively.
8.2. Results

Figure 13 and Figure 14 show the ROC performance analysis of proposed ASM-RW-ANN with the existing adaboost, RW-adaboost, SP, FDA for CASME and SMIC databases respectively. From the Figure 13 and Figure 14, it is observed that the proposed ASM-RW-ANN provides the high true positive rate for small values of false positive rate due to the deformable correlation prior learning of RW models and ANN. With the increase in false positives, the linear increase of True positive rate is observed up to certain point. Beyond that, the performance of true positive is maintained in the constant state at the value of 18 in false positive rate. Figure 15 and Figure 16 show the True positive rate analysis for the different noise ratio values for CASME and SMIC databases respectively. With the increase in noise ratio values, the TPR values are linearly decreased from maximum values. For the minimum noise ratio value for CASME as shown in Figure 15, the TPR values
of existing RW-adaboost and proposed work are 0.72 and 0.78 respectively. Similarly, they are 0.26 and 0.42 respectively for maximum noise ratio values. For the minimum noise ratio value for SMIC as shown in Figure 16, the TPR values of existing RW-adaboost and proposed work are 0.72 and 0.89 respectively. Similarly, for the maximum noise ratio values, they are 0.26 and 0.88 respectively.

The Prior learning by the RW model improves the TPR values for minimum and maximum noise ratio values respectively for the CASME and SMIC databases. With the increase in noise ratio values, the TPR values are linearly increased from minimum values. For CASME, the minimum weak classifiers as shown in Figure 17, the TPR values of existing RW-adaboost and proposed work are 0.658 and 0.705 respectively. Similarly, they are 0.705 and 0.89 respectively for maximum weak classifiers. For SMIC, the minimum weak classifiers as shown in Figure 18, the TPR values of the existing RW-adaboost and proposed work are 0.658 and 0.752 respectively. Similarly the maximum weak classifiers for SMIC, they are 0.705 and 0.899 respectively. The Prior learning by the RW model improves the TPR values for minimum and maximum weak classifiers respectively for CASME and SMIC databases.

9. Conclusion

The issues in the micro-expression recognition framework are addressed in this paper. The novel model ASM-RW-ANN proposed for the effective learning of micro-expressions in this paper. In this model, several processes such as procrustes analysis, key point extraction, RW-based prior learning and the ANN classification are carried out to enhance the micro-expression recognition performance compared to the existing methods. The effectiveness of proposed ASM-RW-ANN is validated over the number of state-of-art methods on CASME and SMIC databases with the parameters of accuracy, true positive rate, ROC against the SVM classifiers. Besides, the variations of TPR due to the variations in the weak classifier count and the noise ratio are investigated. The comparison of proposed ASM-RW-ANN with the existing classifiers assured the effectiveness of proposed method in the automatic FER applications. The major findings of the proposed work are the accurate detection of facial muscle variations, prior learning of features and the better recognition of the facial expressions. The proposed work will be extended by using the relevant attribute merging technique with the optimal selection of feature set. The combination of optimal selected features with the geometrical feature set will be used for analyzing the facial muscle variations in the future work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants 61473290, the National High Technology Research and Development Program of China (863 Program) under Grant 2015AA042307, and the international partnership program of Chinese Academy of Sciences, Grant No. 173211KYSB20160008.

References
