IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 1, 2019, accepted December 17, 2019, date of publication December 30, 2019,
date of current version January 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2963113

Emotion Recognition From Body Movement

FERDOUS AHMED, A. S. M. HOSSAIN BARI, AND MARINA L. GAVRILOVA
Department of Computer Science, University of Calgary, Calgary, AB T2N 1N4, Canada

Corresponding author: Ferdous Ahmed (ferdous.ahmed1@ucalgary.ca)

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant (DG) (Machine on Intelligence for Biometric Security), in part by the Natural Sciences and Engineering Research Council (NSERC) ENGAGE on Gait Recognition, and in part by the Natural Sciences and Engineering Research Council (NSERC) Strategic Projects Grant (SPG) on Smart Cities.

ABSTRACT Automatic emotion recognition from the analysis of body movement has tremendous potential to revolutionize virtual reality, robotics, behavior modeling, and biometric identity recognition domains. A computer system capable of recognizing human emotion from the body can also significantly change the way we interact with the computers. One of the significant challenges is to identify emotion-specific features from a vast number of descriptors of human body movements. In this paper, we introduce a novel two-layer feature selection framework for emotion classification from a comprehensive list of body movement features. We used the feature selection framework to accurately recognize five basic emotions: happiness, sadness, fear, anger, and neutral. In the first layer, a unique combination of Analysis of Variance (ANOVA) and Multivariate Analysis of Variance (MANOVA) was utilized to eliminate irrelevant features. In the second layer, a binary chromosome-based genetic algorithm was proposed to select a feature subset from the relevant list of features that maximizes the emotion recognition rate. Score and rank-level fusion were applied to further improve the accuracy of the system. The proposed system was validated on proprietary and public datasets, containing 30 subjects. Different action scenarios, such as walking and sitting actions, as well as an action-independent case, were considered. Based on the experimental results, the proposed emotion recognition system achieved a very high emotion recognition rate outperforming all of the state-of-the-art methods. The proposed system achieved recognition accuracy of 90.0% during walking, 96.0% during sitting, and 86.66% in an action-independent scenario, demonstrating high accuracy and robustness of the developed method.

INDEX TERMS Emotion recognition, feature selection, gait analysis, genetic algorithm, information fusion, human motion, kinect sensor, biometrics.

I. INTRODUCTION

Emotion recognition based on human body movement is an emerging area of research. The interest in emotion recognition focusing only on body movement, posture, and gesture has risen dramatically over the last few years. This growing interest is due to several reasons. Many psychological studies have found evidence that the human perception can discern various affective states expressed only through body movements [1]–[3]. Body movement information can be a better alternative for recognizing emotions from a distance [4]. Most of the recent research on emotion recognition is focusing on developing a system that can recognize emotions based on nonverbal cues expressed through body movements [5]. Development of a computer system capable of predicting emotion through observation of a human body movement would significantly change the way humans interact with the computers [5], [6].

Based on the above discussion, an increasing number of applications, that use body movement information for emotion recognition, has emerged. One of the recent works used a robot as a social mediator to increase the quality of human-robot interaction [7]. Emotion recognition from body movement encompass a large number of applications including biometric security, healthcare, gaming, and behavior modeling [5]. Examples of applications of emotion recognition in biometric security domain include body movement and facial expression analysis for video surveillance [8], [9]. Use of emotion recognition in the medical domain includes identification of the signature behavior of patients having specific psychological conditions [10]. Despite an abundant demand for an accurate emotion recognition from body movement, this topic became trending only very recently.

The associate editor coordinating the review of this manuscript and approving it for publication was Jihwan P. Choi.

Researchers have mostly attempted to recognize emotions from various modalities, such as the face, head, and hand [11], [12]. Very few studies have focused on whole-body expressions for emotion analysis. However, as stated in [13], a computer model is not only suitable but may even exceed a human observer ability to recognize emotion, as it can detect subtle movement changes not readily apparent to the naked eye. Moreover, body movement information can be obtained noninvasively from a distance which may be beneficial for many practical applications.

Previous research focused only on a limited number of movement features from a vast number of computable features [12], [14]. A successful attempt to understand human emotion from actor's expressive body movements was carried out in [12]. Authors introduced a model based on Laban Movement Analysis (LMA), that integrated Body, Effort, Shape and Space features. However, the list of features was very broad, unstructured, and some of the features were never before used for human emotion. In addition, the relevance factor of the features was never considered. The challenge is thus to create a comprehensive list of motion features that encompass all nuanced movement-related information relevant to the emotional state of an individual [14]. Then, the best combination of movement features obtained using an effective feature selection algorithm can be used to train machine learning algorithms to recognize human emotions accurately. This paper solves all the above-mentioned challenges successfully.

In this paper, a comprehensive list of body movement features is computed and categorized into ten unique groups based on the type of movements. We leveraged the knowledge acquired from other disciplines such as computer animation and graphics for computing the body movement features [14]–[16]. A filter-based feature selection algorithm, Analysis of Variance (ANOVA) [17] was proposed to select relevant features from each of the movement feature groups. Several top features from each feature group were used as inputs to the second layer of the framework. The number of features considered from each group was derived using normalized Multivariate Analysis of Variance (MANOVA) [18] score computed for each group separately. Several popular feature ranking algorithms were investigated, including Mutual Information [19], Chi-squared Score [20], ReliefF [20], and Ensemble of Decision Tree [21]. Based on the two criteria: monotonicity and reliability, ANOVA was chosen to be the most suitable feature selection algorithm. The number of relevant features selected from each group was based on the normalized MANOVA score computed for each motion feature group. A binary chromosome based genetic algorithm was utilized to extract a feature subset maximizing the emotion recognition rate. Finally, a supervised machine learning algorithm, previously proven to be effective in the biometric domain, was used to recognize human emotions.

Based on the proposed framework, we achieved the highest emotion recognition accuracy of 90.0% during walking action sequences, 96.0% during sitting action sequences, and 86.66% in action-independent cases. The method outperfomed all of the state-of-the-art approaches tested on our proprietary dataset. Information fusion techniques such as score and rank-level fusion further improved the emotion recognition accuracy of the proposed system. The proposed system also achieved 81.25% accuracy on a public dataset [22], outperforming existing state-of-the-art methods reported on this dataset. The overall contributions of the presented research are summarized as follows:

- Proposal of a unique structuring of motion features into ten groups, each describing a different aspect of a human body movement.
- Development of a two-layer feature selection architecture that combines the power of a traditional filter-based approach with a genetic algorithm.
- Identification of the most relevant motion features for emotion recognition from a comprehensive list of motion features. The relevance factor was computed for a univariate case where the features were considered independently, and a multivariate case, where features were considered as part of a group.
- Computation of feature relevance during two action scenarios, which provides an additional insight on importance of features during emotion recognition.
- Proposing a unique combination of score and rank-level fusion with two-layer feature selection algorithm to maximize the emotion recognition accuracy.
- Introduction of several new temporal features that exhibited improvements over temporal features, used previously in the literature.

Preliminary work on this subject was carried out and published in [23].

II. PREVIOUS WORK

Emotion can be expressed through eye gaze direction, iris extension, postural features, and movement of the human body [5]. Pollick *et al.* [2] showed that arm movements are significantly correlated with the pleasantness dimension of the emotion model. Bianchi-Berthouze et al. introduced an incremental learning model through gestural cues and a contextual feedback system to self-organize postural features into discrete emotion categories [24]. However, those works were limited to only parts of the body. Several researchers attempted to recognize emotion from dance movement. Camurri et al. in [25] extracted the quantity of motion and contraction index from 2D video images depicting dance movements of the subjects to recognize discrete emotion categories. Very recently, Durupinar et al. in [26] conducted a perceptual study to establish a relationship between the LMA (Laban Movement Analysis) features and the five personality traits of a human. Senecal et al. in [12] analyzed body motion expression in theater performance based on LMA features. Researchers have also focused on recognizing emotion in arbitrary recording scenarios using deep learning architectures [27]. However, those attempts were limited to specific dance movements.

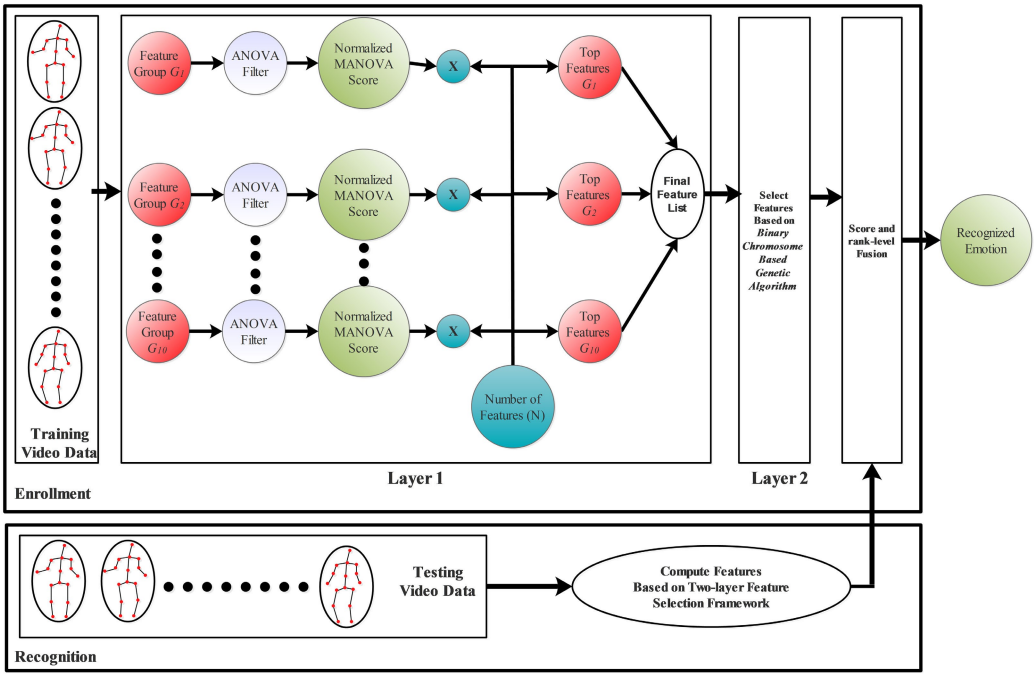

FIGURE 1. An overview of the proposed framework for emotion recognition from body motion.

One of the biggest challenges of emotion recognition is the high dimensionality representation of the motion features. Also, the literature provides very little guidance as to what type of motion features are suitable for emotion classification. Most of the existing research have considered a very limited number of features. Feature relevance was also not considered for emotion recognition. Therefore, most of the existing research is biased towards a particular set of motion features. For instance, Glowinski et al. in [15] extracted energy, spatial extent, symmetry, and smoothness related features and then used Principal Component Analysis (PCA) to create a minimal representation of affective gestures. Saha et al. in [28] picked nine features related to velocity, acceleration, and angular features to identify six emotions. This work successfully addresses the above deficiencies through the proposed comprehensive framework for emotion recognition, described in details in the next section.

III. PROPOSED METHODOLOGY

The first challenge is to create a complete description of a human body movement with emotion-specific identifying information. This problem was overcome by identifying and computing a comprehensive list of movement features. These movement features were then grouped into ten unique categories in such a way that each category represented a special aspect of a body movement (i.e. symmetry, space, speed of motion etc.). The final list of features was computed based on the relevance factor of these features using a two-layer feature selection algorithm. The feature selection framework introduced in section III-A overcomes the difficulty of identifying emotion-specific body movement information.

A. OVERVIEW

The first step of the proposed system involved the extraction of various geometric and kinematic features. Some of these features were previously introduced for 3D motion synthesis, classification, and indexing [16]. Researchers have yet to establish a consensus on the right combination of various motion features. Therefore, in the proposed emotion recognition system, a comprehensive list of motion features was extracted maximizing the available body movement information. The motion features were computed either on a single frame or over a sequence of frames spanned over a short period. As a result, computed motion features characterize various aspects of human motion, such as trajectories or geometric properties of the postures. These features were grouped into ten unique proprietary groups which will be discussed in section III-B.

Moreover, a *temporal profile* was computed for each of the features. The temporal profile consists of twelve, time series functions, as described in section III-C. A temporal profile computed in this way performs better than a histogram with fixed number of bins. The number of bins of a histogram determines the level of discretization of the calculated features. A limitation of using histogram is that the number of bins must be set empirically for the dataset. The values of a histogram are also sparse and most of the bins remain empty after the histogram computation.

The main component of the proposed framework involves a two-layer feature selection process, as shown in Figure 1. In the first layer, irrelevant features are eliminated using a combination of ANOVA and MANOVA. ANOVA is used to sort the features according to their relevance at recognizing

emotions [17]. ANOVA provides two measures: f-score and p-score to compute relevance of a feature. The f-score is a measure of total variation that exists among the arithmetic means of the target emotions. The p-score is a measure that determines the probability associated with rejecting the f-score. The features that failed to pass a *significance test* are discarded immediately. After removal of these features, remaining features are considered as statistically relevant for further analysis.

MANOVA was used to compute group significance and to distribute features among various feature groups. The number of features considered from each group was derived using normalized MANOVA score computed for each group separately [18]. The first layer may not be enough to attain optimum model performance based on the computed relevant features. The reason for this may be attributed to the performance improvement of specific feature combination for certain expert models. Thereby, several top features from each feature group were used as an input to the second layer of the framework. The objective of the second layer is to find the best subset of features that maximizes the emotion recognition rate of the expert models.

Statistically relevant features were ordered based on the computed f-score. To reduce the number of possible combinations for computing feature subset that maximizes the emotion recognition rate, a predefined number of features were selected from the top ANOVA features. Typically, this number is set empirically. According to [29], the number of features can be selected as a function of the sample size, N, and the maximum feature size is N. In the proposed system, the total number of computed features was set based on the sample size of N. Since each group of features describes a different aspect of human body movement, the total number of features were distributed among the feature groups. Top ANOVA features were selected from each feature group based on the total number of features and the normalized MANOVA score computed for each group. MANOVA was used to quantify group significance, and the number of features computed from each group was based on the computed MANOVA score of the group. The group significance scores were normalized so that each score ranges from 0 to 1 and their sum equals to 1. Then, the computed MANOVA score was used to distribute the total number of features from each motion feature group. This way only some of the top features from each motion feature group remained for the subsequent steps. If the number of features for a motion group exceeded the number of features that passed the ANOVA significance test for that group, then all of the features that passed the test in that group were selected.

The reason for using a filter-based approach in the first layer of the framework is to eliminate irrelevant features as much as possible. Most of the filter-based techniques compute rank of the features based on their ability to distinguish among the target categories. The features can be ranked based on their relevant factors, and irrelevant features can be removed based on an empirical threshold value. In our experiments, the features were ranked based on the computed f-score. From the top features based on the computed f-score, the features that produced a p-score, which was higher than a predefined threshold, was chosen for the genetic algorithm. During the experiment, the p-score was chosen as 0.005. This ensures that there exists a minimal chance that the computed f-score was produced from a different distribution. In this way, the first layer used the relevance of the features to prepare for the second layer of the proposed feature selection framework. The second layer uses the genetic algorithm that evaluates the distinctive ability of the features to maximize emotion recognition accuracy.

In the second layer of the two-layer framework, a binary chromosome-based genetic algorithm was used to identify the optimal feature subset that maximizes the emotion recognition rate. The genetic algorithm used in the proposed system achieved a plateau within 800 generations. The mutation rate was set to 0.03, as described in section IV. A detailed explanation of the genetic algorithm is presented in section IV. Finally, the expert models were fused using score and rank-level fusion as described in section IV-A. Figure 5 shows how features for a feature group were selected using the first layer of a two-layer framework.

B. MOTION FEATURE GROUPS

Based on a thorough analysis of the existing literature, a comprehensive list of 3D motion features was extracted. These features were grouped into ten unique categories minimizing the number of overlapping features describing various body movement types as much as possible.

- **Group of Features 1** This group of features consists of low-level feature descriptors that measure the speed of the motion, such as velocity, acceleration, and jerk. If X defines a motion that is described as n consecutive poses, where $X = x(t_1), x(t_2), x(t_3), ..., x(t_n)$. Then, the velocity is defined in equation 1 and the magnitude of the velocity is determined using the equation 2 according to [14]. In equations 1 and 2, $v^k(t_i)$ is the velocity of the k^{th} joint at time t_i, $v_x^k(t_i)$ is the x-component of the velocity of the k^th joint at time t_i, and δt refers to a small fraction of time required for transitioning between consecutive frames. Usually, δt is set to a very small value. During the experiment, the value was set to $\frac{1}{30}$ seconds as Kinect v2 has a frame rate of 30 fps.

$$v^k(t_i) = \frac{X^k(t_{i+1}) - X^k(t_i)}{2\delta t} \tag{1}$$

$$\|v^k(t_i)\| = \sqrt{v_x^k(t_i)^2 + v_y^k(t_i)^2 + v_z^k(t_i)^2} \tag{2}$$

  The acceleration and the jerk were computed based on the second and third order derivatives of the position vector using similar equations.
- **Group of Features 2** This feature group is related to the trajectory of the movement. It is expected to have a higher curvature of the hands that follows a contour of a circle compared to the hands that follow a straight

line [15]. The curvature was calculated using equation 3.

$$\kappa^k(t_i) = \frac{||v^k(t_i) \times a^k(t_i)||}{(\sqrt{v_x^k(t_i)^2 + v_y^k(t_i)^2 + v_z^k(t_i)^2})^3} \tag{3}$$

In equation 3, $\kappa^k(t_i)$ corresponds to curvature of k^{th} joint at time t_i, $v^k(t_i)$ corresponds to the velocity of k^{th} joint at time t_i, and $a^k(t_i)$ corresponds to the acceleration of k^{th} joint at time t_i.

- **Group of Features 3** This feature group represents an aggregated speed over a set of joints and defined as the Quantity of Motion (QoM) in the literature [14]. The QoM is calculated as a weighted sum of velocities of groups of joints. QoM for K number of joints is defined using equation 4.

$$QoM(t_i) = \frac{\sum_{k \in K} w_k v_k(t_i)}{\sum_{k \in K} w_k} \tag{4}$$

In the most recent work on the subject, all body joints are typically assigned uniform value (equal to 1), since all the joints contribute equally to the identification of specific patterns observed during human motion [14], [30]. Other studies [31]–[36] on a human body motion in sports performance, physiotherapy rehabilitation, and emergency response followed the suit and did not isolate any specific body joints from others. Therefore, in our work, the weights (w_k) were assigned uniform value of 1. The joints of the body were segmented into five groups: arm region, head region, upper body, lower body, and finally the whole body encompassing all major joints. Features were extracted separately from each of the body segments.

- **Group of Features 4** This group of features represents how the surrounding space is utilized by a person during movement, as shown in Figure 2. Space utilization of the human body can be estimated by the *bounding volume* of various segments of the body defined over the temporal domain [37]. A rectangular box, enclosing the region of the body joints, was computed for various segments of the body including arm region, head region, upper body, lower body, and whole body. The joints included in various body segments are described in Figure 3. Equations 5 and 6 were used for computing the bounding volume. In equation 5, $|*|$ denotes the absolute value.

$$\begin{aligned} d_x &= |\max_{k \in K} j_x - \min_{k \in K} j_x|, \\ d_y &= |\max_{k \in K} j_y - \min_{k \in K} j_y|, \\ d_z &= |\max_{k \in K} j_z - \min_{k \in K} j_z| \end{aligned} \tag{5}$$

$$BoundingVolume(BV) = d_x * d_y * d_z \tag{6}$$

- **Group of Features 5** This feature group represents the *displacement of the major joints* of the human body. Equation 7 is used to calculate the displacement according to [14].

$$D(t_i) = \|X^l(t_i) - X^r(t_i)\| \tag{7}$$

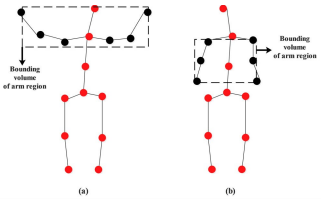

FIGURE 2. Bounding volume of the arm region of two human 3D skeletons (a) and (b). Joints included in the arm region are highlighted using black circles.

In equation 7, r is the reference joint, and l is any other joint of the body. The base of the spine was chosen as the reference joint. The joints considered for displacement computation are Head, Neck, Shoulder, Elbow, Wrist, Hand, Knee, Ankle, Foot, and Center of Mass (COM). COM was computed by calculating a weighted sum of all the joints in 3D Cartesian coordinates.

- **Group of Features 6** In this category, we computed the *motion features*: verticality (maximum distance of the y components for all the joints), extension (maximum distance from the center of mass to all other joints), elbow flexion (elbow was used as the reference joint while a joint relative angle is formed by shoulder, elbow, and hand) [38], arm shape (magnitude of the vector from hand to base of the spine), hand relationship (distance between left and right hands) and feet relationship (distance between left and right feet). These features are similar to motion features described in [14].
- **Group of Features 7** This feature group quantifies the *effort component of the Laban Movement Analysis*. In the proposed system, the analysis was applied to four subcategories of effort. These subcategories include weight, time, space and flow of the effort. The weight subcategory explains the strength of the movement. The two extremes of this movement are light and strong movement. The strength of the movement was quantified by measuring the maximum kinetic energy by various segments of the body observed within a small period. Equations 8 and 9 were used to compute this feature. In equation 8, $E(t_i)$ is kinetic energy of a particular segment of the body at time t_i and α_k represents mass coefficient of various body joints. A uniform value of 1 was used to keep the calculation simple. In equation 9, T indicates a time window within which the maximum kinetic energy, $Weight(t_i)$, at time t_i was computed [14]. This feature was computed for $N - T$ consecutive frames in a walking sequence.

$$E(t_i) = \sum_{k \in K} E_k(t_i) = \sum_{k \in K} \alpha_k v^k(t_i)^2 \tag{8}$$

$$Weight(t_i) = \max_{i \in [1,T]} E(t_i), \quad i = 1, 2, 3, ..., N \tag{9}$$

The time subcategory of effort explains whether the movement was sudden (quick) or sustained (steady). The quantification of this feature was

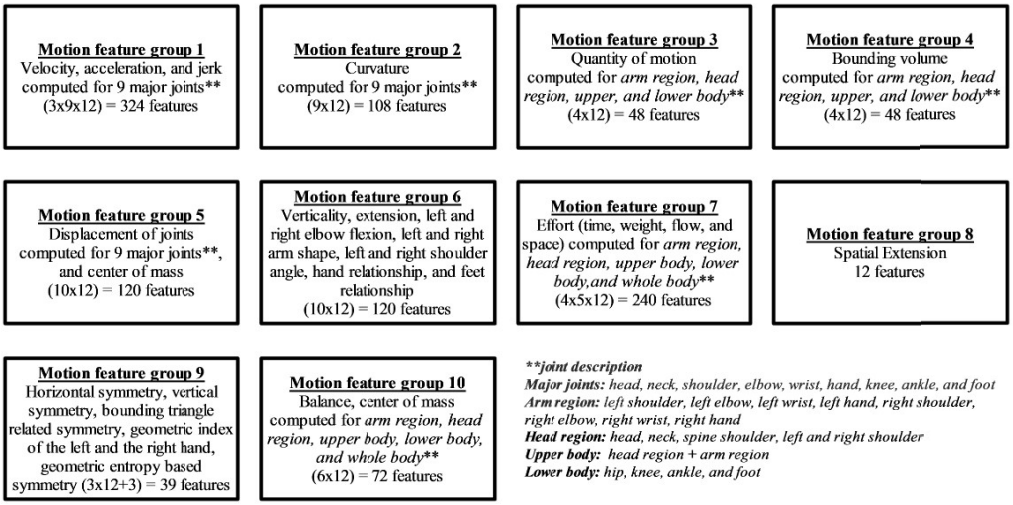

FIGURE 3. A taxonomy of the motion features computed for emotion recognition from body motion features.

accomplished by measuring the acceleration of various body segments over a predefined period. Small values of this feature characterize steady movements. In contrast, higher values indicate a sudden change in body movements. To compute the features of time subcategory, equation 10 was used [14]. In equation 10, $a^k(t_i)$ is the acceleration of the k^{th} joint at time t_i.

$$Time^k(t_i) = \frac{1}{T}\sum_{i=1}^{T} a^k(t_i) \tag{10}$$

The space effort explains whether motion effort was focused towards a particular spot (direct) or several spots (multi-focused and flexible). Equation 11 shows how space effort was computed [14].

$$Space^k(t_i) = \frac{\sum_{i=1}^{T-1} ||x^k(t_{i+1}) - x^k(t_i)||}{||x^k(t_T) - x^k(t_i)||} \tag{11}$$

The final subcategory of effort quantifies the fluidity of the movement. The fluidity can either jerky or smooth and is computed as follows [14]:

$$Flow^k(t_i) = \frac{1}{T}\sum_{i=1}^{T} j^k(t_i) \tag{12}$$

where $j^k(t_i)$ is the jerk of the k^{th} joint at time t_i. In all of the equations mentioned above, T is the total number of the frame under consideration. All the subcategories of effort computed in this section were applied to various body segments separately. The body joints included in each of the body segments are shown in Figure 3.

- **Group of Features 8** This feature group provides an estimate of the *spatial extent* of the bounding triangle formed by the hands and the head, similar to the feature descriptor introduced in [15]. This group of features explains the coverage by hands and head over time. The spatial extent was computed using equations 13 and 14. In equation 14, $X^c(t_i)$ indicates the barycenter of the bounding triangle formed by hands and head. $X^r(t_i)$ denotes 3D Cartesian coordinates of the reference joint r at time t_i. J_{head} indicates the 3D Cartesian coordinate of the head. In the experiment, the base of the spine was used as the reference based on the recommendation of the research conducted in [39].

$$C = \frac{1}{3}(J_{head} + J_{hand\_left} + J_{hand\_right}) \tag{13}$$

$$SpatialExtent(t_i) = ||(X^c(t_i) - X^r(t_i))|| \tag{14}$$

- **Group of Features 9** This feature group represents the *symmetry of the movement*. Some previous works in the literature seem to indicate that the correlation of asymmetry with a relaxed attitude and high social status of a person exists [40]. Therefore, spatial asymmetry was calculated to measure the expressivity of the body movement. There are two methods of calculating this feature [15]. Discarding one method of computation may result in a loss of important information regarding the emotion. Therefore, two types of spatial asymmetries were computed using two of those methods. In the first method, horizontal and vertical asymmetries were calculated from the barycenter of the bounding triangle formed by two hands and head [15]. Then, horizontal, vertical, and bounding triangle-based asymmetries were computed using equations 15, 16, and 17 [15]. In equations 15 and 16, $j1$ and $j2$ refer to the coordinates of the left and the right hand.

$$SI_{horizontal}(t_i) = \frac{((j1_x(t_i) - j_x(t_i)) - (j2_x(t_i) - j_x(t_i)))}{(|j1_x(t_i) - j_x(t_i)| + |j2_x(t_i) - j_x(t_i)|)} \tag{15}$$

$$SI_{vertical}(t_i) = \frac{((j1_y(t_i) - j_y(t_i)) - (j2_y(t_i) - j_y(t_i)))}{(|j1_y(t_i) - j_y(t_i)| + |j2_y(t_i) - j_y(t_i)|)} \tag{16}$$

$$SI(t_i) = \frac{SI_{horizontal}(t_i)}{SI_{vertical}(t_i)} \tag{17}$$

In the second method, the geometric entropy of each hand was computed as defined in [41]. This measure represents the spread of the movement in the available space. Equation 18 was used to calculate the geometric entropy of hand [41]:

$$H = \frac{2 * LP}{c} \tag{18}$$

where LP is the path length of the center of the mass of the left or right arm region and c is the perimeter of the convex hull of the selected region. The measure was taken for both left and right arm region separately. Then, the overall spread was calculated using the equation 19 based on [41].

$$SI = \frac{H_{lefthand}}{H_{righthand}} \tag{19}$$

Temporal profile was not computed for the geometric entropy-based features described in equations 18 and 19, as these features were computed based on the overall time sequence.

- **Group of Features 10** This feature group describes how balance of various segments of human body changes during movement. In this feature group, temporal *ceter of mass displacement (COMD)* and *balance* of the body were computed during movement. The center of mass was computed by measuring the arithmetic mean of the Cartesian 3D coordinates of all the joints. COMD was computed by calculating the magnitude difference of center of mass for two consecutive frames. Equations 20, 21, and 22 were used to compute COMD. The balance of the body was computed by measuring the difference between the center of mass of the upper and the lower body. Equation 23 was used to compute the balance [14].

$$COM(t_i) = \frac{1}{\sum_{k \in K} 1} \sum_{k \in K} J_k(t_i) \tag{20}$$

$$C(t_i) = \sqrt{COM_x(t_i)^2 + COM_y(t_i)^2 + COM_z(t_i)^2} \tag{21}$$

$$COMD(t_i) = C(t_{i+1}) - C(t_i) \tag{22}$$

$$Balance(t_i) = ||COM_{upperbody}(t_i)|| - ||COM_{lowerbody}(t_i)|| \tag{23}$$

C. TEMPORAL PROFILE

Some of the features introduced in the previous section were defined over the time domain. In order to reduce the noise without eliminating the high-frequency components, Savitzky-Golay filter [15] was applied. We introduced twelve statistical measures to characterize feature over time domain as shown in Figure 4. These measures capture how motion features evolve. The time features computation is explained as follows:

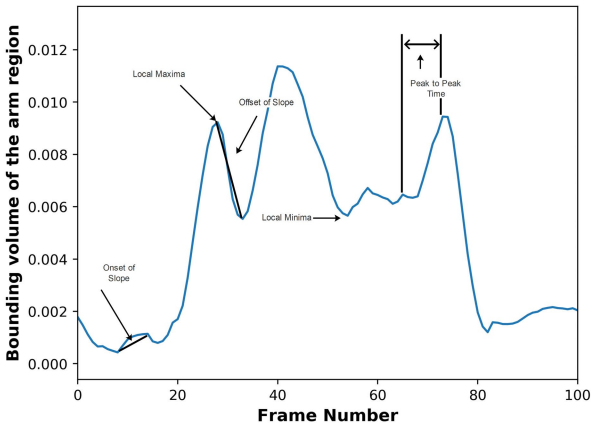

FIGURE 4. Visualization of temporal features computed from a time series.

Features 1–5: The following features were calculated to capture the overall behavior over time: min, max, mean, standard deviation, and min versus max ratio. In equations 24–28, X depicts the time series data and i depicts a specific point in time.

$$Min(X) = \min_{i=1}^{N} X(i) \tag{24}$$

$$Max(X) = \max_{i=1}^{N} X(i) \tag{25}$$

$$Mean(X) = \frac{1}{N} \sum_{i=1}^{N} X(i) \tag{26}$$

$$StandardDeviation(X) = \sqrt{\sum_{i=1}^{N} \frac{(X(i) - Mean(X))^2}{N-1}} \tag{27}$$

$$MinVersusMaxRation(X) = \frac{Min(X)}{Max(X)} \tag{28}$$

Feature 6: This feature measures the amount of white noise present in the time series. The spectral flatness was computed by taking the ratio of the geometric and the arithmetic mean of the power spectrum of the time series. Spectral flatness was computed using equation 29.

$$SpectralFlatness(X) = \frac{\sqrt[N]{\prod_k f_x(k)}}{\frac{1}{N} \sum_k fx(k)} \tag{29}$$

In equation 29, f_x is the power spectrum of x, k is an index into the spectrum, and N is the number of non-zero elements of the signal.

Features 7–8: The mean of the extreme values (local minimum and maximum values) were calculated and then added to the temporal profile. Features 7 and 8 were computed using equations 30 and 31, respectively.

$$MeanOfLocalMinimum = \frac{1}{M} \sum_{i=0}^{M-1} X(Lmin(i)) \tag{30}$$

$$MeanOfLocalMaximum = \frac{1}{M}\sum_{i=0}^{M-1} X(Lmax(i)) \quad (31)$$

Features 9–10: To characterize the transition from one extreme value to the next, the slope of their movement over time was computed. "Onset (τ)" and "offset (Ψ)" slope were computed based on the direction of the slope between the two consecutive local minimum and local maximum values. Then, the mean value was added to the temporal profile. These features were computed using equations 32 and 33.

$$\tau = \frac{1}{M-1}\sum_{i=0}^{M-2} \frac{(X(Lmax(i+1)) - X(Lmin(i)))}{|Lmax(i+1) - Lmin(i)|} \quad (32)$$

$$\Psi = \frac{1}{M-1}\sum_{i=0}^{M-2} \frac{(X(Lmin(i+1)) - X(Lmax(i)))}{|Lmin(i+1) - Lmax(i)|} \quad (33)$$

In equations 32 and 33, *Lmax* and *Lmin* depict local maxima and local minima of the time series.

Feature 11: Average time between two consecutive extreme values was computed using equation 34.

$$\Omega = \frac{1}{2M}\sum_{i=0}^{M-1} (|Lmin(i+1) - Lmin(i)| + |Lmax(i+1) - Lmax(i)|) \quad (34)$$

Feature 12: This feature characterizes whether local minimum and local maximum values were reached using similar speed. This feature is defined as the ratio of the onset and the offset slopes previously calculated. This feature is computed using 35.

$$RatioOfSlopes = \frac{\tau}{\Psi} \quad (35)$$

D. FIRST LAYER OF THE TWO-LAYER FEATURE SELECTION FRAMEWORK

In the first layer of the feature selection framework, ANOVA was used to compute feature significance of each of the motion features. Suppose, there is N number of observations from k different emotion groups. An observation is denoted by x_{ij}, where i indicates the emotion category, while j indicates the index of an observation. If the overall mean of all observations is denoted by $\bar{x}$ and group mean is denoted by $\bar{x_i}$ then the equation of an observation can be as [17]:

$$x_{ij} = \bar{x} + (\bar{x_i} - \bar{x}) + (x_{ij} - \bar{x_i}) = \bar{x} + (\bar{x_i} - \bar{x}) + \epsilon_{ij} \quad (36)$$

The error term, ϵ_{ij}, used in equation 36 is equal to $(x_{ij} - \bar{x_i})$ and it is assumed to have a Gaussian distribution with zero mean and unit variance [17]. ANOVA attempts to compute quantitative measure in support of this hypothesis. For this reason, two quantitative measures: the sum of squares between groups (SSB) and the sum of squares within groups (SSW) were computed using equations 37 and 38, respectively. In equation 37, n_j refers to the number of elements of group j.

$$SSB = \sum_{j=1}^{k} n_j(\bar{x_j} - \bar{x}) \quad (37)$$

$$SSW = \sum_{j=1}^{k}\sum_{i=1}^{N}(x_{ij} - \bar{x_j}) \quad (38)$$

It is difficult to compare two measures when they are not normalized properly. In statistics, degrees of freedom is defined as the number of values in a calculation that can vary [17]. *SSB* and *SSW* have different degrees of freedom. Therefore, comparison can only be made once these two measures are normalized. The degrees of freedom of *SSB* is $k - 1$, and of *SSW* is $N - k$. Therefore, mean SSB (MSSB) and mean SSW (MSSW) can be obtained as [17]:

$$MSSB = \frac{1}{k-1} * SSB \quad (39)$$

$$MSSW = \frac{1}{N-k} * SSW \quad (40)$$

Further, MSSB and MSSW can be combined to calculate one uniform scalar value. This value is called f-score in the literature [17]. F-score can be computed using equation 41. If the f-score value is close to 1, it means that significant difference among the means of different emotion classes was observed. A high f-score value indicates a high relevance of the feature.

$$f\_score = \frac{MSSB}{MSSW} \quad (41)$$

In ANOVA, another measure called the p-score is computed alongside the f-score. The p-score denotes the risk of rejecting the f-score, which indicates how the means are different from each other. The p-score value is computed by observing the standard normal distribution. Having a small p-score value indicates a high probability of accepting the calculated f-score value. In other words, p-score value determines whether any statistical significance exists in the observed samples. A threshold value of 0.005 is usually set for the p-score. The p-score value of 0.005 can be interpreted as there is a 0.5% probability that no statistical significance exists. Based on the p-score, statistically insignificant features were removed from each feature group. MANOVA is the multivariate extension of ANOVA. The discrimination ability to distinguish various emotions by the feature groups was computed using MANOVA [18]. Therefore, unlike ANOVA, in which each observation is represented using a single scalar value x_{ij}, in MANOVA each observation is represented by a vector X_{ijp} in which j denotes the emotion category, i denotes the index of an observation, and p denotes the index of a particular feature within a motion feature group. The values of i and j are constrained by $0 \leq i \leq k$ and $0 \leq j \leq n$. In MANOVA, the total

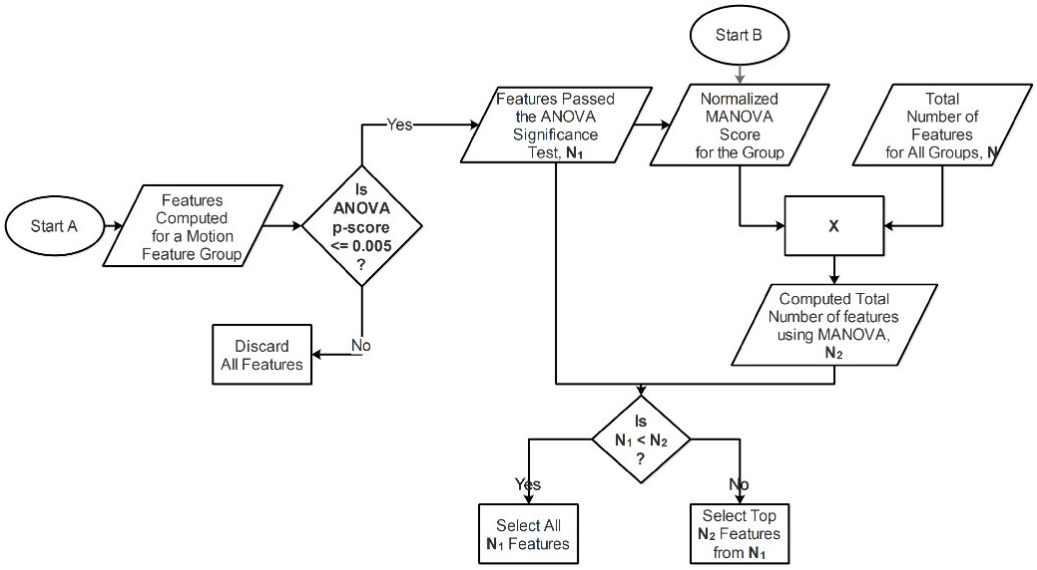

FIGURE 5. Selection of features in a motion feature group using the first layer of a two-layer feature selection framework.

sum of squares and cross product (TSSCP) is represented by equation 42.

$$\begin{aligned} &TSSCP \\ &= \sum_{j=1}^{k}\sum_{i=1}^{n}(X_{ij}-\bar{X_T})(X_{ij}-\bar{X_T})^T \\ &= \sum_{j=1}^{k}\sum_{i=1}^{n}[(X_{ij}-\bar{X_j})+(\bar{X_j}-\bar{X_T})][(X_{ij}-\bar{X_j})+(\bar{X_j}-\bar{X_T})]^T \\ &= \sum_{j=1}^{k} n_j(\bar{X_j}-\bar{X_T})(\bar{X_j}-\bar{X_T})^T+\sum_{j=1}^{k}\sum_{i=1}^{n}(X_{ij}-\bar{X_j})(X_{ij}-\bar{X_j})^T \\ &= H+E \end{aligned} \tag{42}$$

In equation 42, H is the hypothesis matrix, and E is the error matrix. The degrees of freedom for the *TSSCP*, H, and E are $n-1$, $k-1$, and $n-k$, respectively. F-score for the MANOVA is computed using Pillai's trace value from H and E matrices as shown in equation 43. In equation 43, λ_i is the i^{th} eigenvalue and q is the number of eigenvalues extracted.

$$f\_score = trace(\frac{H}{H+E}) = \sum_{i=1}^{q}\frac{\lambda_i}{1+\lambda_i} \tag{43}$$

F-score was computed for each motion feature group separately. Then these values were normalized and mapped to a range between 0 and 1. The sum of the f-scores resulted in a score of 1. Individual normalized f-scores were further multiplied with the total number of features to compute the overall feature for each motion feature group. If this value exceeded the total number of features computed for a particular group, all the features corresponding to the group were computed. Figure 5 illustrates this process.

IV. SECOND LAYER OF A TWO-LAYER FEATURE SELECTION FRAMEWORK

Features selected in the first layer of the two-layer framework contained statistically significant information regarding various emotion categories. However, it is not guaranteed that all features considered would be equally effective at recognizing emotions. Moreover, the f-score value computed using ANOVA only ensures variations between a single pair of emotion categories. A binary chromosome-based genetic algorithm was proposed to address this problem. This algorithm automatically selects a subset of features that, as a group, is more powerful compared to all features selected together at once. The main purpose of using the genetic algorithm is to explore whether a subset of features exists that maximizes the emotion recognition rate of all expert models.

The main components of a genetic algorithm with binary coding scheme include a population of chromosomes, a fitness function for optimization, a selection process for the reproduction of chromosomes, a crossover operator that produces the next generation of chromosomes, and a mutation operator to introduce variability [42]. The population of chromosomes can be considered as a sequence of ones and zeros that indicates whether a feature is selected. The indices of the ones' correspond to the indices of the selected features. The population size of a genetic algorithm affects the ability of exploration of the feature space. If the value is set too low, the genetic algorithm may not produce enough variability among the chromosomes. For this reason, the population size of the chromosomes was set empirically to 30.

The fitness function determines the ability of a chromosome to survive a generation of reproduction. The main goal of using the genetic algorithm is to find a subset of features that maximizes the emotion recognition rate. Therefore,

the emotion recognition rate was an obvious choice for the fitness function. Each chromosome in the population was evaluated using the fitness function. The list of chromosomes was sorted based on the result of the fitness function. Half of the population chromosomes were automatically scheduled to survive when the next generation of chromosomes were reproduced. The remaining half of the population was chosen based on a crossover operator performing recombination among the top half of the chromosomes.

The crossover operation used in the proposed system is shown using equation IV. Equations IV shows how crossover operator reproduces chromosomes at a crossover point x from two chromosomes C_1 and C_2 in a m-dimensional feature space. The crossover operator chooses a crossover point with a uniform probability distribution for each pair of consecutive chromosomes survived for the next generation [42]. Thus, the crossover point was chosen randomly for each pair of chromosomes for reproduction.

The mutation operator introduces randomness so that the crossover operation can avoid repeated reproduction of the same chromosomes. Mutation rate is a hyperparameter to balance exploitation and exploration ability of a genetic algorithm. If this value is set too high, genetic algorithm may not converge to a plateau during which the maximum recognition rate is not changed. If this value is set too low, genetic algorithm may get stuck in a local maxima. Typically, the mutation rate is set to a small value of 2–5% [43]. In our experiments, mutation rate of 0.03 was chosen. The mutation rate indicates that there is a 3% probability of reversing a single bit value randomly in a chromosome. The number of features N was more than 150. The brute-force algorithm would require exploring of 2^{150} combinations of chromosomes. If the number of features increases, the number of combinations of chromosomes will also increase exponentially. The binary chromosome-based genetic algorithm used in the proposed system produced a plateau within 800 generations. In this context, plateau represents the number of generations during which the maximum emotion recognition rate remains unchanged. The choices of hyperparameters have an impact on the performance of genetic algorithm. However, during our experiment, we found that the impact was quite low. Therefore, only the final accuracy after the fine-tuning the hyperparameters was reported in Section V of this paper.

$$\begin{aligned} C_1 &= [C_{1,1}, C_{1,2}, C_{1,3}, ..., C_{1,m}] \\ C_2 &= [C_{2,1}, C_{2,2}, C_{2,3}, ..., C_{2,m}] \\ C_1(x) &= [C_{1,1}, C_{1,2}, C_{1,3}, \ldots, C_{1,x}, C_{2,x+1}, C_{2,x+2}, ..., C_{2,m}] \\ C_2(x) &= [C_{2,1}, C_{2,2}, C_{2,3}, \ldots, C_{2,x}, C_{1,x+1}, C_{1,x+2}, ..., C_{1,m}] \end{aligned} \tag{44}$$

A. INFORMATION FUSION

One of the methods to boost the overall recognition rate of the expert models is to fuse decisions obtained from the expert models. Information fusion is a powerful technique that combines the decisions from multiple expert systems to improve the overall recognition rate. Fusion of decisions can be achieved in several ways depending on how the decisions are combined. Score and rank-level fusion are popular fusion techniques, that have been successfully adopted in similar applications [44]. In the proposed system, five expert models including Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Decision Tree (DT), Gaussian Naive Bayes (GNB), and K-Nearest Neighbor (KNN) were fused and performance of score-level and rank-level fusion were compared. A real-coded genetic algorithm was used to estimate the weight parameters. Genetic algorithm allows convergence to the solution much quickly than a brute-force approach. Replacing a fixed number of discrete values with real values increases the domain of possible weights. Another advantage of using real values is that they can exploit the graduality of the fitness function [45]. In a real-coded genetic algorithm, a transformation is required to convert a chromosome consisting of zeros and ones to real values. The transformation function is defined using equation 45 [45].

$$T(c) = \frac{1}{2^{m-1}-1} \sum_{j=1}^{m} c_j 2^{j-1} \times where, \quad \forall c = (c_1, c_2, ..., c_m) \in [0, 1]^m \tag{45}$$

In equation 45, m denotes the level of discretization used for the weights. T denotes the function to transform decision genes to real values. In the proposed system, the value for m was set to 8 or 2^8 discrete levels of real-valued weights. Since five expert models were used, the size of the chromosome was set to 40, where each expert model is associated with a chromosome of size 8. The remaining steps for the genetic algorithm were similar to the method described in section IV. The population size and mutation rate was set to 30 and 0.03, respectively. The terminating condition for the algorithm was established using the same approach as described in section IV. In the experiment, the plateau was reached within 300 generations of reproductions.

V. EXPERIMENTAL ANALYSIS

A. PROPRIETARY DATASET

The first experiment was conducted on 30 subjects of a proprietary dataset. Each subject performed five different emotionally expressive walking sequences including a separate neutral walking sequence. Laban Movement Analysis (LMA) framework was used as a guideline to synthesize human motion styles similar to paper [12]. We focused on subjects' structural and physical properties of body shape, dynamic quality of movement, and surrounding space utilization during movement. None of the subjects had any prior acting experience. Each emotional walking sequence was recorded for 20 seconds using Microsoft Kinect v2. A total of 3000 seconds of recorded video data containing approximately 90,000 frames were recorded. Subjects walked in front of Kinect in a circular fashion showing both sides of the body. Figure 6 shows snapshots of the human skeleton joint

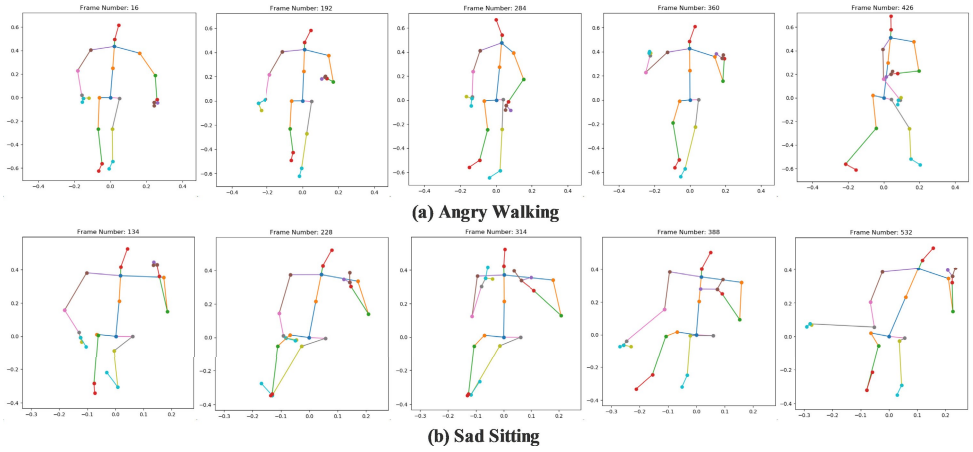

FIGURE 6. Two emotionally expressive action sequences. (a) Five representative frames from an angry walking sequence. (b) Five representative frames from a sad sequence while the subjects were in a sitting position.

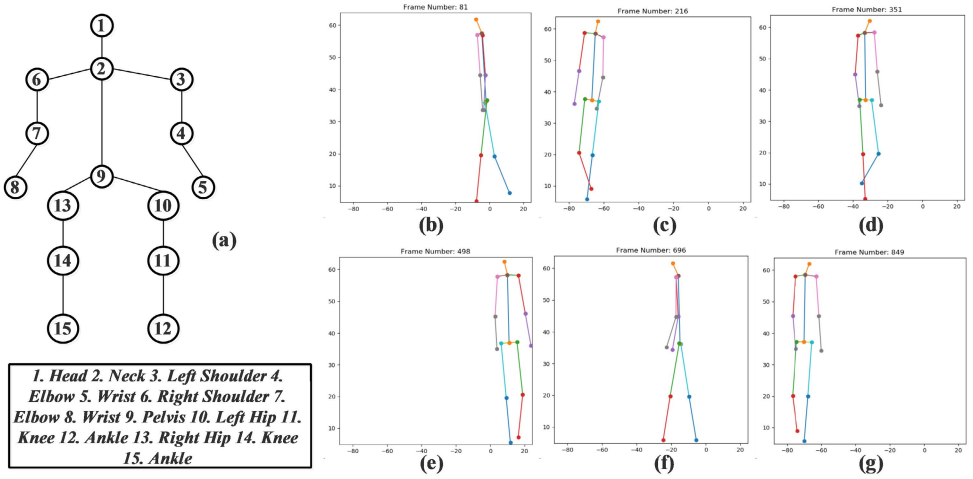

FIGURE 7. (a) The name of the body joints considered for emotion recognition through movement analysis. Snapshot of frames containing 3D body coordinates computed from a sad walking sequence are shown in (b), (c), (d), (e), (f), and (g).

coordinates computed during a sad and an angry emotional walking sequences.

B. PUBLIC DATASET

The dataset collected movements from 30 nonprofessional actors [22]. According to the researchers, the reasons for choosing nonprofessionals were to avoid systematic exaggeration of movements and to increase variability of movement expression. Each actor performed movements expressing four different emotions including neutral, happiness, sadness, and anger. Only verbal instructions were provided to the subjects. Emotions were observed during walking scenarios. Two recordings were collected from each subject for each emotional walk. Therefore, a total of 240 samples of various emotional walking sequences were considered for the experiment. Each subject wore a suit with retroreflective markers. These markers were able to reflect light spots to a two-dimensional space. Overall, 35 markers were placed at various places of the human body, but only 15 joint coordinates were converted to 3D space. This dataset can be considered very challenging, since a reduced number of body joints were considered for the experiment. Figure 7 shows snapshots of the human skeleton joint coordinates computed during a sad emotional walk.

C. SELECTION OF THE FILTER-BASED TECHNIQUE

In order to eliminate irrelevant features before applying any expert model, two general categories of feature selection algorithms were used: filter-based and wrapper-based approaches [46]. There are many filter-based techniques developed over the years. We considered five different techniques for this research. Typical filter-based techniques include ReliefF based approaches [20], information theoretic approaches [19], statistical approaches [20] and ensemble of decision tree-based approaches. Feature subset from the relevant list of features was selected using recently proposed genetic algorithm based framework [47]. Out of

the five techniques analyzed, one technique was chosen based on two criteria.

1) CONSISTENCY

This criterion examines whether the feature ranking algorithms provide consistent rank over various subsets of the dataset. The dataset was split randomly into two folds 100 times. Each time, the filter-based algorithm computed the rank for each fold. The expected outcome is to have minimum disagreements between the computed ranks. The scores computed were normalized using min-max normalization. Suppose, an algorithm generated R_1 and R_2 feature ranks for two subsets of the same dataset as shown in equations 46 and 47. Consistency score was computed using equation 48 where N denotes the number of features and *Rscore* denotes the consistency score for a single iteration.

$$R_1 = (R_{11}, R_{12}, R_{13}, ..., R_{1N}) \tag{46}$$

$$R_2 = (R_{21}, R_{22}, R_{23}, ..., R_{2N}) \tag{47}$$

$$Rscore = \sqrt{\sum_{i=1}^{N}(R_{1,i} - R_{2,i})^2} \tag{48}$$

The results shown in Table 1 prove that ANOVA provided the most consistent ranking among all the algorithms considered. The closest second algorithm was ReliefF with neighbor size 30. From the observed consistency scores, it can be inferred that the algorithms considered provided more consistent ranks based on emotion expressed during the sitting action compared to the walking action.

TABLE 1. The table shows measured consistency scores for various filter-based feature selection algorithm.

Filter-Based Method	Walking Action	Sitting Action
ReliefF (number of neighbors = 30)	0.403 ± 0.067	0.241 ± 0.072
Mutual Information (MI)	0.477 ± 0.051	0.311 ± 0.058
Analysis of Variance (ANOVA)	$\boldsymbol{0.262 \pm 0.062}$	$\boldsymbol{0.210 \pm 0.085}$
CHI-squared Score (CHI2)	0.476 ± 0.051	0.279 ± 0.061
Ensemble decision tree-based Method (EDT)	0.907 ± 0.039	0.772 ± 0.042

2) MONOTONICITY

We examined the level of monotonicity by various feature ranking algorithms. The monotonicity indicates the gradual performance decline along the ranked order of features generated by the algorithm. We utilized LDA and SVM to measure the monotonicity within an interval of 30 consecutive frames. Then, Pearson correlation (to measure the linear relationship) and Spearman rank-order correlation (to measure the non-parametric measure) were computed for each of the filter-based algorithms. ANOVA exhibited the highest level of average monotonicity among the feature ranking algorithms as shown in Table 2.

Based on the experimental result for showing the consistency and monotonicity properties of filter-based methods (shown in Table 1 and Table 2), it is evident that ANOVA is the most appropriate method for the filter-based layer of the proposed framework.

TABLE 2. The table shows measured monotonicity scores for various filter-based feature selection algorithm for expert models: LDA and SVM.

Filter-Based Method	Walking Action			Sitting Action		
	LDA	SVM	Average Correlation	LDA	SVM	Average Correlation
ReliefF (number of neighbors = 30)	-0.878	-0.931	-0.9045	-0.936	-0.965	-0.9505
Mutual Information (MI)	-0.873	-0.926	-0.8995	-0.951	-0.951	-0.9510
Analysis of Variance (ANOVA)	***-0.908***	***-0.952***	***-0.9300***	***-0.963***	***-0.969***	***-0.9660***
CHI-squared Score (CHI2)	-0.841	-0.941	-0.8910	-0.925	-0.916	-0.9205
Ensemble decision tree-based Method (EDT)	-0.834	-0.861	-0.8475	-0.917	-0.889	-0.9030

D. FEATURE ANALYSIS

From the analysis presented in the previous section, ANOVA was used to select relevant features. A high score of ANOVA value signifies a high relevance of a feature. ANOVA also provides p-score that describes the statistical significance of the result. The p-value was set to 0.005. Based on the measured p-value, any feature failing to pass the significance test was automatically discarded from consideration. The remaining features were sorted based on their feature relevance. Then, the top features from each of the motion feature groups were chosen based on normalized MANOVA score and passed onto the second layer of the framework. The total number of computed motion features was 1131. More features passed the significance test for the sitting action sequences compared to the walking sequences. In case of emotion expressed during walking sequences, 436 features passed the significance test compared to 633 during the sitting sequences as shown in Table 3.

One of the key motivations behind feature analysis is to gain a deeper understanding of human emotion and related body movement information. Therefore, the top two features from each motion feature group were separately computed for further analysis. Figure 8 shows the top two features computed from each group during walking and sitting actions separately. Some conclusions can be drawn by observing the f-scores of the motion features. During walking action, the movements observed in the arm and the upper body region were key factors to perceive emotion. The highest importance was assigned to the jerkiness of the wrist. ANOVA assigned high importance to QoM observed in the upper body and the arm region. The time subcategory of the effort component was given high importance to perceive emotion during walking sequences. It can be inferred from the results that whether the perceived movements are sudden or sustained can be an important factor in recognizing human emotion during walking.

During sitting action, movements related to hands and utilization of surrounding space were assigned more importance compared to the suddenness or quickness of the movement. The important feature for recognizing emotion during sitting action was the minimum elbow flexion observed during a sequence. Other prominent features include maximum displacement of the hand and the wrist, bounding volume of the arm region and the upper body, and the maximum spatial extent of the whole body. The most prominent subcategory in the effort component is space, which indicates whether the movement is focused on a particular spot or flexible.

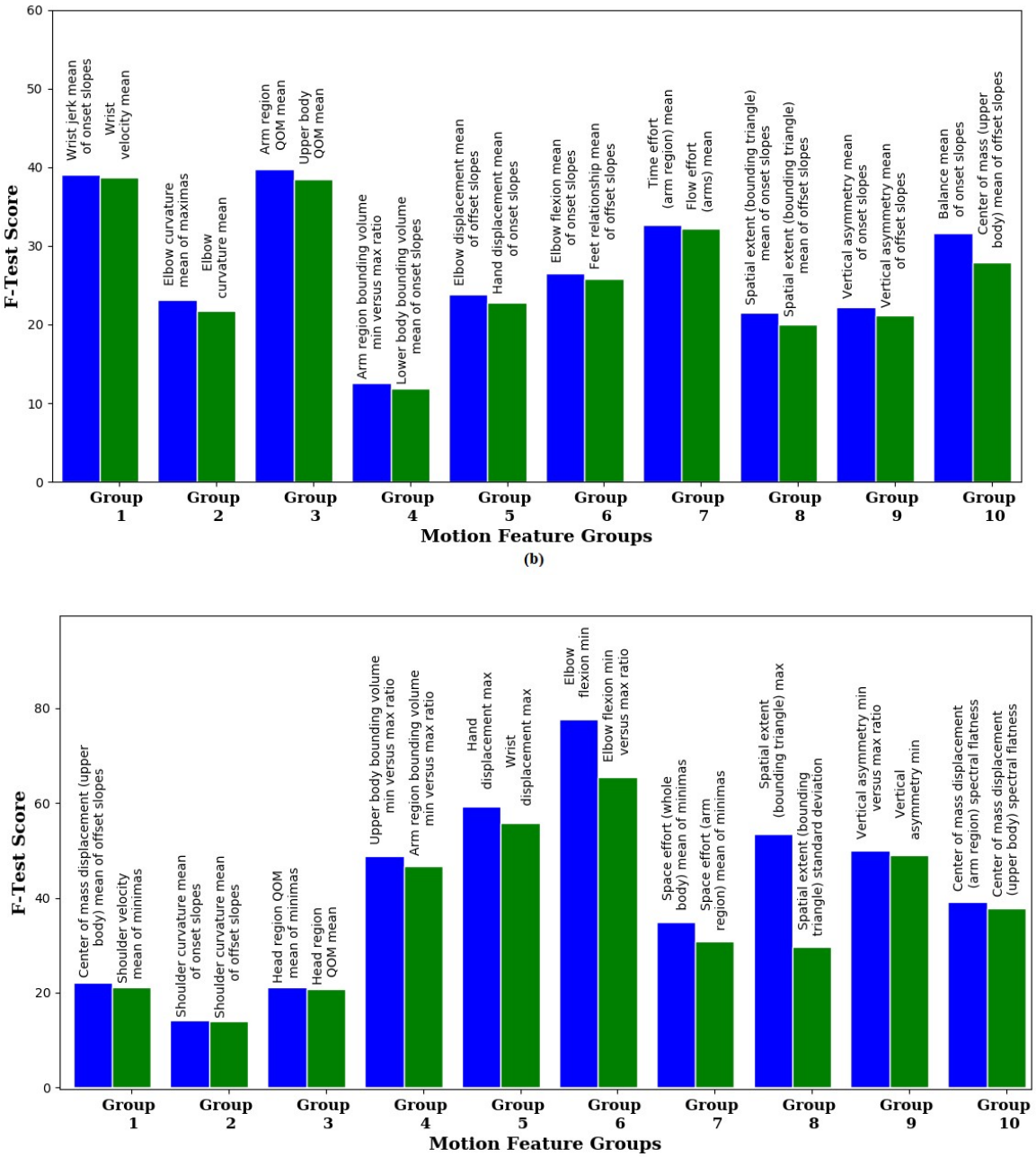

FIGURE 8. Top two motion features from each group computed using ANOVA for emotion expressed during (a) sitting and (b) walking action.

TABLE 3. The table shows the number of features that passed the ANOVA significance test.

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	Total
Total Extracted features	324	108	48	48	120	120	240	12	39	72	1131
Features selected (walking)	136	18	20	11	43	70	87	8	14	29	436
Features selected (sitting)	178	30	31	29	71	80	129	9	26	50	633

Another observation is that the overall f-scores were higher for the sitting action compared to the walking action. Higher f-scores during sitting action sequences indicate that motion information contained more cues to recognize emotion during sitting action in contrast with the walking action.

As was discussed before, a combination of motion features having high ANOVA score may not perform the same when considered as a group. Therefore, the MANOVA score was computed for each motion group to analyze group significance and to reduce the number of motion features

TABLE 4. MANOVA *f*-score and *p*-score computed for various motion feature groups.

Motion Feature Group	Sitting				Walking			
	MANOVA f-score	MANOVA p-score	MANOVA f-score Normalized	Top 150 features	MANOVA f-score	MANOVA p-score	MANOVA f-score Normalized	Top 150 features
G1	14.05	1.02e-9	0.10	15	31.46	1.11e-16	0.18	28
G2	19.68	5.86e-13	0.14	21	18.00	5.05e-12	0.10	16
G3	18.30	3.42e-12	0.13	19	41.15	1.11e-16	0.24	36
G4	9.39	8.89e-7	0.06	10	7.67	1.23e-5	0.04	7
G5	8.38	4.07e-6	0.06	9	13.74	1.58e-9	0.08	12
G6	14.81	3.56e-10	0.11	16	6.85	4.4e-5	0.03	6
G7	5.39	0.0004	0.03	6	28.11	1.11e-16	0.16	25
G8	23.54	5.10e-15	0.16	25	15.67	1.11e-10	0.09	14
G9	25.25	6.66e-16	0.18	27	3.62	0.0075	0.02	3
G10	1.51	0.2	0.01	2	5.102	0.0007	0.03	4

considered for the second layer of the framework. MANOVA f-scores were computed using Pillai's trace as discussed in section III-D. Since 150 samples were available for both the sitting and the walking sequences, top 150 features were distributed among the feature groups based on the normalized MANOVA f-scores. Individual MANOVA f-scores and the distribution of features are shown in Table 4. From Table 4, it is observed that G8 and G9 exhibited the highest f-scores obtained during sitting action and therefore, more features were distributed in these two categories. The feature groups: G1, G3, and G7 contained the maximum number of motion features for emotion expressed during walking action.

Further insight regarding the feature significance of emotions can be gained using a post hoc analysis of the significant features. Figure 9 shows how means of different emotions vary based on the computed feature. The distribution of the features for happy and angry emotions during walking action was very similar. The distribution of features for sadness, fear, and neutral emotions was also very similar during walking action. Based on the distribution of features for different emotions, it is difficult to discriminate emotions using only a single significant feature. However, combining these features allows distinguishing among various emotional expressions better. Overall, recognizing neutral emotion during sitting action was easier compared to other emotions. Although elbow flexion was the most significant feature, it can only distinguish neutral and angry emotions from other emotional expressions. However, maximum hand displacement was able to distinguish more emotion groups since the distribution of this feature is quite different for various emotion groups. From the analysis, we observed that features computed from the upper section of the body is quite important in overall emotion recognition.

E. FEATURE SUBSET SELECTION

From the previous section, it was concluded that the filter-based feature selection algorithm alone could not reliably distinguish among various emotion categories. Therefore, a genetic algorithm discussed in section IV was utilized. In order to test the performance of a feature subset, machine learning models such as Support Vector Machine (SVM),

TABLE 5. Performance improvement of the individual expert models after using the proposed feature selection framework (PFSF) during walking, sitting, and action-independent scenario.

Expert Models		Emotion Recognition Rate (%)		
		Walking	Sitting	Action-independent
SVM	Before using PFSF	52.0	70.67	62.0
	After using PFSF	80.0	92.67	80.33
LDA	Before using PFSF	55.33	76.67	65.0
	After using PFSF	84.66	94.67	80.33
GNV	Before using PFSF	41.33	52.0	40.67
	After using PFSF	66.67	86.67	64.67
DT	Before using PFSF	50.0	68.67	52.0
	After using PFSF	76.0	78.67	65.0
KNN	Before using PFSF	43.33	70.0	60.0
	After using PFSF	76.67	91.33	79.0

Linear Discriminant Analysis (LDA), Naive Bayes classifier, K-Nearest Neighbor, and Decision Tree were used. The parameters for the SVM classifier, such as C (margin maximization of the decision function), γ (influence of the training samples), decision type (one-versus-all or one-versus-rest), and kernel type (radial basis or linear function) were chosen based on an exhaustive grid search. We used the singular value decomposition as the solver for the LDA. The number of neighbors for the KNN classifier was chosen as eleven based on the recommendation from [47]. We used five-fold cross-validation during the experiment. We avoided biased learning by not taking samples from the same subject during both the testing and the training sets.

A binary encoded genetic algorithm was used for each expert model. The size of the chromosome populations was kept at 30, and the mutation rate was set to 0.03 as discussed in Section IV. Reproduction of populations using crossover operators was performed 1000 times to achieve the maximum recognition rate for the classifiers as shown in Figure 10. It can be observed from Table 5 that the proposed feature selection method with LDA classifier achieved the highest emotion recognition rate of 94.66% during sitting action and 84.66% during walking action. Based on Table 6, all the expert models achieved maximum accuracy using only 52–84 features. This is a substantial reduction of the number of features from the original 1131 feature set.

While conducting experiments, we discovered that the top 4.9% (55 features) of all the features computed were enough

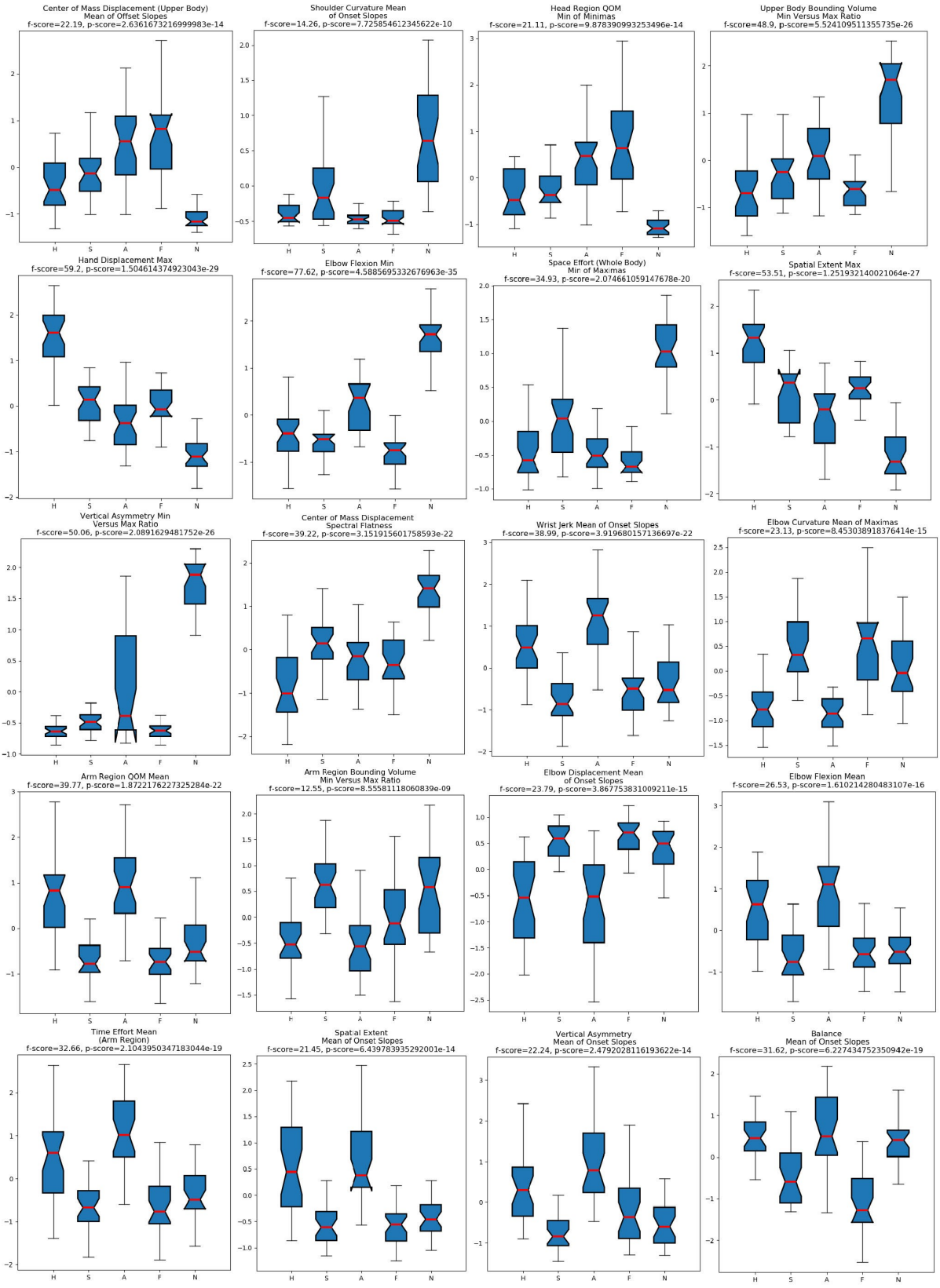

FIGURE 9. Post hoc analysis of the most significant features for each motion group.

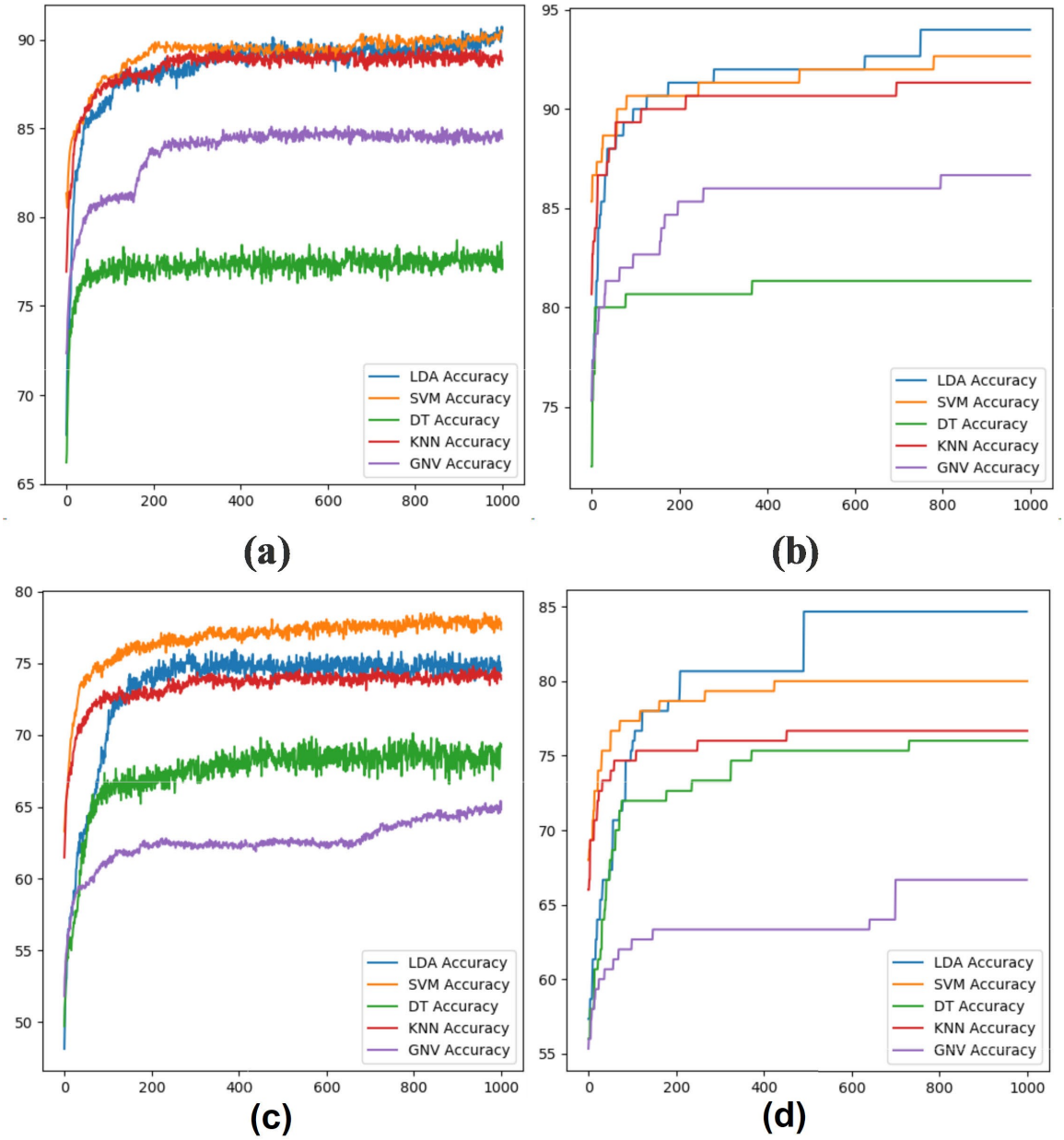

FIGURE 10. Improvement of emotion recognition rate of different classifiers over various generations (number of chromosomes = 30): (a) average emotion recognition accuracy during sitting action, (b) maximum emotion recognition accuracy during sitting action, (c) average emotion recognition accuracy during walking action, (d) maximum emotion recognition accuracy during walking action.

to produce the best result for walking scenario. Moreover, top 5.8% (66 features) and 5% (57 features) of all the features were sufficient to achieve the best results for sitting and action-independent scenarios, respectively. TABLE 6 shows the exact number of top features that produced the best emotion recognition accuracy.

Furthermore, the proposed method with SVM classifier is considered as the second best classifier achieving 92.66% and 80.00% emotion recognition rates during sitting and walking action sequences, respectively. A separate analysis was conducted to evaluate the proposed system's performance in an action-independent scenario. The SVM outperformed other expert models during an action-independent scenario achieving 83.33% emotion recognition rate. The proposed LDA and KNN based methods also achieved a good recognition rate of 80.33% and 79.00% respectively. However, the recognition accuracy of the naive Bayes and the decision tree slightly degraded because of action-independence consideration.

F. PERFORMANCE ENHANCEMENT USING SCORE-LEVEL AND RANK-LEVEL FUSION

Recognition accuracy was enhanced by applying information fusion of the prediction scores obtained using multiple

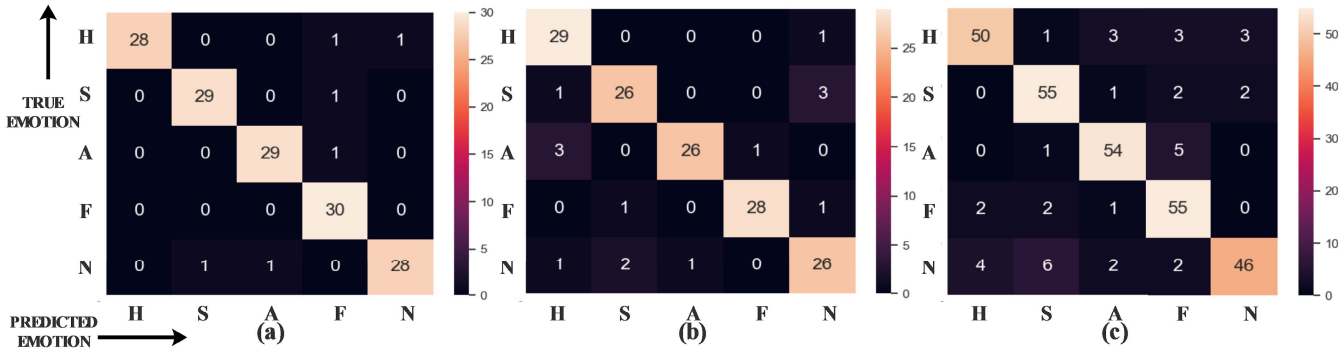

FIGURE 11. The confusion matrix computed for five emotions (H: happiness, S: sadness, A: anger, F: fear, and N: neutral) during (a) sitting, (b) walking, and (c) action-independent scenarios.

TABLE 6. Number of selected features of the individual expert models after using the proposed feature selection framework (PFSF) during walking, sitting, and action-independent scenario.

Expert Models	Number of Selected Features			Total Extracted Features
	Walking	Sitting	Action-independent	
SVM	56	66	84	1131
LDA	55	66	57	
GNV	56	57	72	
DT	52	82	69	
KNN	77	70	78	

TABLE 7. The table shows the improvement of the emotion recognition rate after using score-level and rank-level fusion with weights obtained using a real-coded genetic algorithm.

Expert Models	Sitting	Walking	Action Independent
	Emotion Recognition Rate (%)	Emotion Recognition Rate (%)	Emotion Recognition Rate (%)
Best accuracy by an expert model	94.66	84.66	83.33
Score-level fusion	96.0	90.0	86.66
Rank-level fusion	96.0	89.33	86.33
Improvement by score-level fusion	***1.34***	***5.34***	***3.33***
Improvement by rank-level fusion	***1.34***	***4.67***	***3.0***

expert models. In the proposed system, the recognition ability of five expert models was combined using score and rank-level fusion. These expert models are Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Linear Discriminant Analysis (LDA), Gaussian Naive Bayes (GNB), and Decision Tree (DT). The score in this context was represented by a binary value to express the presence or absence of a particular emotion category. The number of generations of reproductions, mutation rate, and the population were set to 300, 0.03, and 30, respectively (as discussed in Section IV). The emotion recognition rate was improved in all action scenarios after applying fusion methods. The proposed system achieved 5.34% improvement for the walking action scenarios, 1.34% for the sitting action scenarios, and 3.33% in action-independent scenarios using score-level fusion (see Table 7). The proposed system achieved 4.67% improvement for the walking action scenarios, 1.34% for the sitting action scenarios, and 3.00% in action-independent scenarios using rank-level fusion, as shown in Table 7. The final emotion recognition rate for walking, sitting, and action-independent scenarios are 90.0%, 96%, and 86.66%, respectively.

The confusion matrix of the proposed system is defined as an $N \times N$ square matrix where each row corresponds to the target emotion category, and each column represents the predicted emotion. The confusion matrix was computed for each action scenario separately to detect the emotion categories that were difficult to recognize. It can be observed from Figure 11 that the proposed system was able to recognize 144 samples out of 150 samples of various emotional body expressions during the sitting action. During walking action, 135 out of 150 samples were correctly recognized. In an action-independent scenario, the problem becomes significantly more complicated and therefore, the performance of the system slightly degraded. Out of 300 samples, 260 samples were correctly recognized by the proposed system in an action-independent scenario. From Figure 11, it is noticeable that the samples obtained during the neutral emotion were most frequently confused with other emotion categories during an action-independent scenario.

VI. COMPARISON WITH OTHER STATE-OF-THE-ART METHODS

To compare the proposed emotion recognition system with state-of-the-art research in this domain, features defined in two recent papers: method [12] and method [28], were implemented and tested on our proprietary dataset. For both methods, experiments were conducted using original features introduced in those works with both LDA and SVM classifiers. Standard normalization was used to normalize the extracted features in both methods. In method [28], researchers manually selected nine features related to accelerations, distances, angles of different joints in upper body and arm regions. In method [12], researchers extracted a total of eighty-seven features based on LMA components. Comparisons of the proposed system against the work presented in method [12] and [28] are shown in Table 8. The proposed system was tested for sitting, walking, and action-independent scenarios.

TABLE 8. The table shows a comparison of the proposed emotion recognition system with other state-of-the-art systems.

Method	Sitting	Walking	Action Independent
	Emotion Recognition Rate (%)	Emotion Recognition Rate (%)	Emotion Recognition Rate (%)
Method [12] features + SVM	70.0	62.0	61.0
Method [12] features + LDA	63.33	52.0	62.0
Method [12] features + ANOVA + SVM	73.0	64.0	64.0
Method [12] features + ANOVA + LDA	72.0	58.67	64.66
Method [28] features + SVM	56.67	44.0	53.0
Method [28] features + LDA	58.0	48.67	48.67
Method [28] features + ANOVA + SVM	66.67	50.7	53.33
Method [28] features + ANOVA + LDA	62.67	49.33	49.67
Proposed System	***96.0***	***90.0***	***86.67***

In Table 8, emotion recognition accuracy for method [12] and [28] was in the range of 56.66% to 70% during walking and sitting action, respectively. In action-independent cases, the performance of both methods dropped to a range of 48.67% to 66%. When ANOVA feature selection algorithm was applied to the computed features for method [12], which was our original modification to improve the performance of that method, the recognition rate improved significantly. The recognition rate improved by 2.0% to 8.67% during sitting, 2.0% to 6.67% during walking, and 2.66% to 3.0% during action-independent cases. The recognition rate of method [28] also improved after using our proposed ANOVA method during walking and sitting actions. Highest improvement of 10.0% was observed when selected features using ANOVA were used with SVM during sitting. The emotion recognition rate during walking was also improved from 0.66% to 6.0%.

In summary, we can observe a significant performance improvement when the two-layer feature selection framework was applied in the proposed emotion recognition system. The discriminative ability of the motion features and original two-layer feature selection framework, jointly contributed to achieving emotion recognition accuracy of 90.0%, 96.0%, and 86.67% during walking, sitting, and action-independent cases, respectively. The proposed system achieved 26.0%, 23.0%, and 22.0% improvement over the state-of-the-art methods during sitting, walking, and action-independent scenarios on the proprietary dataset.

VII. PERFORMANCE OF THE PROPOSED SYSTEM ON A PUBLIC DATASET

Experimental results presented thus far have primarily focused on the proprietary dataset. In addition, our proposed system was also tested on a publicly available dataset [22]. Our proposed emotion recognition system significantly improved the overall recognition rate. The expert models: LDA, SVM, DT, NB, and KNN achieved 75.53%, 76.37%, 67.93%, 62.87%, and 67.93% recognition accuracy after computation of a comprehensive list of motion features and applying the proposed two-layer feature selection framework (see Table 9). The recognition accuracy was further boosted when score and rank-level fusion were applied to combine the expert models. Score-level fusion improved emotion recognition accuracy by 4.88% and rank-level fusion improved it by 4.46% over single expert model. After applying score and rank-level fusion, the final recognition accuracy achieved was 81.25% and 80.83%.

TABLE 9. The emotion recognition rate of the proposed emotion recognition system.

Method	Emotion Recognition Rate (%)
Linear discriminant analysis (LDA) + Two-layer feature selection	75.53
Support vector machine (SVM) + Two-layer feature selection	76.37
K-nearest neighbor (KNN) + Two-layer feature selection	67.93
Decision Tree (DT) + Two-layer feature selection	67.93
Naive Bayes (NB) + Two-layer feature selection	62.87
Two-layer feature selection + Rank-level fusion	80.83
Two-layer feature selection + Score-level fusion	81.25

TABLE 10. Comparison of methodologies applied to the dataset.

Method	Emotion Recognition Rate (%)
Two-fold PCA + 1st Eigenwalkers + Naive Bayes (Person-independent) [48]	65.0
Two-fold PCA + Best Eigenwalkers + Naive Bayes (Person-independent) [48]	72.0
KPCA (quadratic kernel) + SVM (Person-independent) [48]	52.0
Motion segmentation + SVM with polynomial kernel (Person-independent) [50]	56.0
Motion segmentation + SVM with polynomial kernel (Person-dependent) [50]	77.0
Proposed emotion recognition system (Person-independent)	81.25

Table 10 compares the proposed system with the state-of-the-art methods on the publicly available dataset [22]. In [48], researchers used a combination of principal component analysis (PCA) and Fourier transformation (FT) for feature reduction. Then, they applied Naive Bayes (NB) to achieve a maximum recognition rate of 65.0% with a single eigenwalk and 72.0% with all the best eigenwalks. The emotion recognition rate was 52% when kernel PCA replaced the two-fold PCA and NB was replaced with SVM. The researchers in [49] used SVM with a polynomial kernel to achieve 56.0% accuracy in person-independent scenarios and 77.0% in person-dependent scenarios. In person-dependent scenarios, personal idiosyncrasies were removed from each motion descriptor. For this reason, their system would only be applicable for recognizing emotion for known subjects, whose personal bias information is available. However, estimating personal bias for unknown subjects raises a severe problem, significantly limiting the applicability of their system. Our system is developed for person-independent cases and therefore, much more robust in addition to being much more accurate.

VIII. OPEN PROBLEMS

Presented research naturally leads to a number of interesting open problems, answering which will pave the way for the

future research. In this section, we decided to focus on three main application domains, where identifying emotion from body movements might have the most impact:

A Humanoid Robots.
B Emergency Response.
C Medical Rehabilitation.

A Within the field of humanoid robots, as well as the general area of human-computer interaction, the need to understand and to imitate the behavior of a human is paramount [7], [30], [51], [52]. Some of the practical applications are enterprise greeting robots (at hotel receptions, university research centers, shopping centers), special needs robots, senior assistant robots, interaction with hearing impaired population, etc. [53]–[57]. Open questions in this domain are:
  - Is there a subset of specific movements of the human that is responsible for specific emotion that will be most beneficial for a humanoid robot?
  - Are there automated tools that can be developed that will enhance quality of human to human or human to robot communication through recognizing human emotion from motion?
  - When is it more beneficial to use multi-modal gait/face emotion recognition system instead of face only/gait only emotion recognition?

B Emergency response is another domain where accurate recognition of a human emotional state can be crucial for a success of a rescue operation or for mitigation of an immediate public risk [31], [58]. The areas for future investigations in this important domain are multiple:
  - Are there any other biometric modalities (such as EEG, heart rate etc.) that can be remotely observed with the intent of emotional state assessment?
  - Which motion features correspond to distress or anxiety emotional state?
  - Can emotion-based features achieve high accuracy while recognized in real-time?

C Athlete performance and medical rehabilitation are areas where mental state plays almost as important role as physical fitness [32]–[34]. From that perspective, observing specific traits on body posture, joint movements and corresponding emotional expressions are paramount for successful outcome of an athlete's training program or a physical rehabilitation process [35], [36], [58]. Open problems arising from our research in relationship to those areas are:
  - Is there a correlation that can be detected by observing athlete training routine and their emotion state and can it be utilized for the development of a better training program?
  - Are there specific poses and/or joint movements that facilitate a more successful physio rehabilitation program with better patient outcomes?
  - What are the major challenges in capturing emotional state of athletes and are those different/similar to general population?

We believe that answers to the above open problems will allow not only to discover new insights on how human emotions are expressed through body movements, but to enhance quality of live and provide meaningful public service in the smart society of the future.

Another research area that has the potential for further exploration is the use of multimodal approaches such as the face, voice, and whole-body expression. Combining multiple modalities has a number of benefits: it can improve overall system accuracy, mitigate missing data and prevent unwanted interference from adversaries [44], [59]. However, the downside of such an approach is additional sensors to sample data, a more complex architecture, and a higher computational complexity. Combining the developed system with face emotion module, for instance, will be a natural future extension of this research.

On the other hand, deployment of a specific application using introduced method will require addressing legal and regulatory challenges. As with any new technology, there is a potential for its misuse. To mitigate this risk, in the area of biometrics, extensive research has been devoted to the revocability of user information and additional means of biometric data protection [60], [61]. When such system is implemented, there is a possibility to store only limited amount of information which can be revoked if compromised [62]. In addition, the rules for sharing biometric data can be restricted according to research presented in [63]. In addition, skeleton sequences can be encrypted using biometric encryption technologies for ensuring data privacy and security [64], [65]. Thus, complying with the regulatory challenges while deploying biometric systems is a very important open research question that needs to be addressed.

IX. CONCLUSION AND FUTURE WORK

Emotion recognition using body movement is an emerging area of research. Significant benefits can be achieved for biometric security, patient behavior monitoring, gaming, and robotics with the creation of a movement-based emotion-aware computer system. Body movement information can provide valuable cues related to the emotional state of a person. Despite showing great potential to be an essential indicator of perceived emotions, body movement information is one of the least explored modalities for emotion recognition. This paper has addressed the problem of creation of a complete system that can accurately recognize five basic emotions: happiness, sadness, anger, fear, and neutral based on body movement features. The experimental results showed that it is possible to build a computer system capable of recognizing human emotion only based on body movement information. Univariate and multivariate analysis of the motion features provided important cues regarding perceived emotion. Experiment results provided critical information regarding the perceived emotion in walking and sitting action scenarios. During walking action, the quantity of movement in the arm and the upper body region were essential indicators. On the other hand, body space utilization, elbow angle, and spatial

extension were essential cues to recognize emotion during the sitting action. During action-independent cases, motion features important during all action scenarios need to be considered to maximize the emotion recognition rate.

The human movement style is often affected by gender, culture, and other idiosyncrasies. For this reason, many researchers considered movement bias resulting from individual movement style during the computation of the motion features. Although this technique yielded improvement of the emotion recognition rate, the approach mentioned severely limits the robustness of the system. The proposed emotion recognition system achieved a very high recognition rate while not considering person-specific bias during body movement. Therefore, our proposed system can be considered more practical and robust. The proposed emotion recognition system outperformed the state-of-the-art methods on both public and proprietary datasets. Our proposed emotion recognition system achieved significant improvement over state-of-the-art methods tested on both proprietary and public datasets during all action scenarios.

In our experiments, we noticed a slight decrease in performance in action-independent cases. For this reason, in real world applications, the expert model might need to be trained based on the actions required for specific applications. It is worth noting that the scope of the current research is confined to emotions that are expressed through nonverbal body movements in a controlled setting. Contextual information may be considered for emotion recognition in the future. The multi-modal system can be designed by fusing emotion from body motion recognition with other biometric modalities, such as emotions from facial expression or voice. While this will likely lead to enhanced system performance, special consideration must be given to privacy and confidentiality of user data. As always, benefits of the novel technologies must be carefully balanced with public security and privacy in real-world applications.

REFERENCES

[1] M. Coulson, "Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence," *J. Nonverbal Behav.*, vol. 28, no. 2, pp. 117–139, 2004.
[2] F. E. Pollick, H. M. Paterson, A. Bruderlin, and A. J. Sanford, "Perceiving affect from arm movement," *Cognition*, vol. 82, no. 2, pp. B51–B61, Dec. 2001.
[3] W. H. Dittrich, T. Troscianko, S. E. G. Lea, and D. Morgan, "Perception of emotion from dynamic point–light displays represented in dance," *Perception*, vol. 25, no. 6, pp. 727–738, Jun. 1996.
[4] H. G. Wallbott, "Bodily expression of emotion," *Eur. J. Soc. Psychol.*, vol. 28, no. 6, pp. 879–896, Nov. 1998.
[5] F. Noroozi, C. A. Corneanu, D. Kamińska, T. Sapiński, S. Escalera, and G. Anbarjafari, "Survey on emotional body gesture recognition," 2018, *arXiv:1801.07481*. [Online]. Available: https://arxiv.org/abs/1801.07481
[6] M. Sultana, P. P. Paul, and M. Gavrilova, "Social behavioral biometrics: An emerging trend," *Int. J. Patt. Recogn. Artif. Intell.*, vol. 29, no. 8, Dec. 2015, Art. no. 1556013.
[7] Y. Tahir, J. Dauwels, D. Thalmann, and N. M. Thalmann, "A user study of a humanoid robot as a social mediator for two-person conversations," *Int. J. Soc. Robot.*, pp. 1–14, Apr. 2018.
[8] S. N. Yanushkevich, A. Stoica, S. N. Srihari, V. P. Shmerko, and M. Gavrilova, "Simulation of biometric information: The new generation of biometric systems," in *Proc. Int. Workshop Model. Simulation Biometric Technol.*, 2004, pp. 87–98.
[9] Y. Luo, M. L. Gavrilova, and P. S. P. Wang, "Facial metamorphosis using geometrical methods for biometric applications," *Int. J. Patt. Recogn. Artif. Intell.*, vol. 22, no. 3, pp. 555–584, May 2008.
[10] N. Fragopanagos and J. G. Taylor, "Emotion recognition in human–computer interaction," *Neural Netw.*, vol. 18, no. 4, pp. 389–405, 2015.
[11] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, "Emotion recognition using facial expressions," *Procedia Comput. Sci.*, vol. 108, pp. 1175–1184, Jan. 2017.
[12] S. Senecal, L. Cuel, A. Aristidou, and N. Magnenat-Thalmann, "Continuous body emotion recognition system during theater performances," *Comput. Animation Virtual Worlds*, vol. 27, nos. 3–4, pp. 311–320, May 2016.
[13] B. De Gelder, "Why bodies? Twelve reasons for including bodily expressions in affective neuroscience," *Biol. Sci.*, vol. 364, no. 1535, pp. 3475–3484, Dec. 2009.
[14] C. Larboulette and S. Gibet, "A review of computable expressive descriptors of human motion," in *Proc. 2nd Int. Workshop Movement Comput. (MOCO)*, 2015, pp. 21–28.
[15] D. Glowinski, N. Dael, A. Camurri, G. Volpe, M. Mortillaro, and K. Scherer, "Toward a minimal representation of affective gestures," *IEEE Trans. Affective Comput.*, vol. 2, no. 2, pp. 106–118, Apr. 2011.
[16] M. Kapadia, I.-K. Chiang, T. Thomas, N. I. Badler, and J. T. Kider, "Efficient motion retrieval in large motion databases," in *Proc. ACM SIGGRAPH Symp. Interact. 3D Graph. Games (I3D)*, 2013, pp. 19–28.
[17] A. Gelman, "Analysis of variance: Why it is more important than ever," *Ann. Statist.*, vol. 33, no. 1, pp. 1–53, Feb. 2005.
[18] B. G. Tabachnick, L. S. Fidell, and J. B. Ullman, *Using Multivariate Statistics*, vol. 5. Boston, MA, USA: Pearson, 2007.
[19] B. C. Ross, "Mutual information between discrete and continuous data sets," *PLoS ONE*, vol. 9, no. 2, Feb. 2014, Art. no. e87357.
[20] M. Robnik-Sikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, nos. 1–2, pp. 23–69, Oct. 2003.
[21] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[22] Y. Ma, H. M. Paterson, and F. E. Pollick, "A motion capture library for the study of identity, gender, and emotion perception from biological motion," *Behavior Res. Methods*, vol. 38, no. 1, pp. 134–141, Feb. 2006.
[23] F. Ahmed and M. L. Gavrilova, "Two-layer feature selection algorithm for recognizing human emotions from 3D motion analysis," in *Proc. Comput. Graph. Int. Conf.* Cham, Switzerland: Springer, 2019, pp. 53–67.
[24] N. Bianchi-Berthouze and A. Kleinsmith, "A categorical approach to affective gesture recognition," *Connection Sci.*, vol. 15, no. 4, pp. 259–269, Dec. 2003.
[25] A. Camurri, I. Lagerlöf, and G. Volpe, "Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques," *Int. J. Hum.-Comput. Stud.*, vol. 59, nos. 1–2, pp. 213–225, Jul. 2003.
[26] F. Durupinar, M. Kapadia, S. Deutsch, M. Neff, and N. I. Badler, "Perform: Perceptual approach for adding OCEAN personality to human motion using laban movement analysis," *TOGACM Trans. Graph.*, vol. 36, no. 4, p. 1, Jul. 2017.
[27] D. Kollias, P. Tzirakis, M. A. Nicolaou, A. Papaioannou, G. Zhao, B. Schuller, I. Kotsia, and S. Zafeiriou, "Deep affect prediction in-the-wild: Aff-wild database and challenge, deep architectures, and beyond," 2018, *arXiv:1804.10938*. [Online]. Available: https://arxiv.org/abs/1804.10938
[28] S. Saha, S. Datta, A. Konar, and R. Janarthanan, "A study on emotion recognition from body gestures using Kinect sensor," in *Proc. Int. Conf. Commun. Signal Process.*, Apr. 2014, pp. 56–60.
[29] J. Hua, Z. Xiong, J. Lowey, E. Suh, and E. R. Dougherty, "Optimal number of features as a function of sample size for various classification rules," *Bioinformatics*, vol. 21, no. 8, pp. 1509–1515, Apr. 2005.
[30] Z. Zhang, A. Beck, and N. Magnenat-Thalmann, "Human–like behavior generation based on head–arms model for robot tracking external targets and body parts," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1390–1400, Aug. 2015.
[31] A. Al-Kaff, F. M. Moreno, A. De La Escalera, and J. M. Armingol, "Intelligent vehicle for search, rescue and transportation purposes," in *Proc. IEEE Int. Symp. Saf., Secur. Rescue Robot. (SSRR)*, Oct. 2017, pp. 110–115.
[32] I. L. Khanin, *Emotions in Sport*. Champaign, IL, USA: Human Kinetics, 2000.
[33] S.-C. Yeh, W.-Y. Hwang, T.-C. Huang, W.-K. Liu, Y.-T. Chen, and Y.-P. Hung, "A study for the application of body sensing in assisted rehabilitation training," in *Proc. Int. Symp. Comput., Consum. Control*, Jun. 2012, pp. 922–925.

[34] N. Stanger, M. Kavussanu, A. Willoughby, and C. Ring, "Psychophysiological responses to sport-specific affective pictures: A study of morality and emotion in athletes," *Psychol. Sport Exercise*, vol. 13, no. 6, pp. 840–848, Nov. 2012.

[35] A. K. Roy, Y. Soni, and S. Dubey, "Enhancing effectiveness of motor rehabilitation using Kinect motion sensing technology," in *Proc. IEEE Global Hum. Technol. Conf., South Asia Satell. (GHTC-SAS)*, Aug. 2013, pp. 298–304.

[36] H. Zacharatos, C. Gatzoulis, and Y. L. Chrysanthou, "Automatic emotion recognition based on body movement analysis: A survey," *IEEE Comput. Graph. Appl.*, vol. 34, no. 6, pp. 35–45, Nov. 2014.

[37] K. Hachimura, K. Takashina, and M. Yoshimura, "Analysis and evaluation of dancing movement based on LMA," in *Proc. IEEE Int. Workshop Robot Hum. Interact. Commun. (ROMAN)*, Oct. 2006, pp. 294–299.

[38] F. Ahmed, P. P. Paul, and M. L. Gavrilova, "DTW-based kernel and rank-level fusion for 3D gait recognition using Kinect," *Vis Comput*, vol. 31, nos. 6–8, pp. 915–924, Jun. 2015.

[39] F. Ahmed, A. S. M. H. Bari, B. Sieu, J. Sadeghi, J. Scholten, and M. L. Gavrilova, "Kalman filter-based noise reduction framework for posture estimation using depth sensor," in *Proc. 18th Int. Conf. Cogn. Inform. Cogn. Comput.*, Jul. 2019, pp. 150–158.

[40] A. Mehrabian, *Nonverbal Communication*. Evanston, IL, USA: Routledge, 2017.

[41] P. Cordier, M. M. France, J. Pailhous, and P. Bolon, "Entropy as a global variable of the learning process," *Hum. Movement Sci.*, vol. 13, no. 6, pp. 745–763, Dec. 1994.

[42] L. Booker, D. Goldberg, and J. Holland, "Classifier systems and genetic algorithms," *Artif. Intell.*, vol. 40, nos. 1–3, pp. 235–282, Sep. 1989.

[43] O. Kramer, *Genetic Algorithm Essentials*, vol. 679. Cham, Switzerland: Springer, 2017.

[44] M. L. Gavrilova and M. Monwar, *Multimodal Biometrics and Intelligent Image Processing for Security Systems*. Hershey, PA, USA: IGI Global, 2013.

[45] F. Herrera, M. Lozano, and J. L. Verdegay, "Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis," *Artif. Intell. Rev.*, vol. 12, no. 4, pp. 265–319, 1998.

[46] H. Wang, T. M. Khoshgoftaar, and J. Van Hulse, "A comparative study of threshold–based feature selection techniques," in *Proc. IEEE Int. Conf. Granular Comput.*, Aug. 2010, pp. 499–504.

[47] F. Ahmed, B. Sieu, and M. L. Gavrilova, "Score and rank–level fusion for emotion recognition using genetic algorithm," in *Proc. IEEE 17th Int. Conf. Cognit. Informat. Cognit. Comput. (ICCI*CC)*, Jul. 2018, pp. 46–53.

[48] M. Karg, R. Jenke, K. Kühnlenz, and M. Buss, "A two-fold PCA-approach for inter-individual recognition of emotions in natural walking," in *Proc. MLDM Posters*, 2009, pp. 51–61.

[49] D. Bernhardt and P. Robinson, "Detecting affect from non-stylised body motions," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.* Berlin, Germany: Springer, 2007, pp. 59–70.

[50] D. Bernhardt, "Emotion inference from human body motion," Ph.D. dissertation, Computer Lab., Univ. Cambridge, Cambridge, U.K., 2010.

[51] N. Magnenat-Thalmann and D. Thalmann, *Handbook of Virtual Humans*. Hoboken, NJ, USA: Wiley, 2005.

[52] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," *Artif. Intell. Rev.*, vol. 43, no. 1, pp. 1–54, Jan. 2015.

[53] K. Yamazaki, R. Ueda, S. Nozawa, M. Kojima, K. Okada, K. Matsumoto, M. Ishikawa, I. Shimoyama, and M. Inaba, "Home–assistant robot for an aging society," *Proc. IEEE*, vol. 100, no. 8, pp. 2429–2441, Aug. 2012.

[54] N. Magnenat-Thalmann and Z. Zhang, "Social robots and virtual humans as assistive tools for improving our quality of life," in *Proc. 5th Int. Conf. Digit. Home*, Nov. 2014, pp. 1–7.

[55] N. M. Thalmann, L. Tian, and F. Yao, "Nadine: A social robot that can localize objects and grasp them in a human way," in *Proc. Frontiers Electron. Technol.* Singapore: Springer, 2017, pp. 1–23.

[56] P. A. Mitkas, "Assistive robots as future caregivers: The rapp approach," in *Progress in Automation, Robotics and Measuring Techniques*. Cham, Switzerland: Springer, 2015, pp. 171–179.

[57] A. Ioannou and A. Andreva, "Play and learn with an intelligent robot: Enhancing the therapy of hearing-impaired children," in *Proc. IFIP Conf. Hum.-Comput. Interact.* Cham, Switzerland: Springer, 2019, pp. 436–452.

[58] M. L. Gavrilova, F. Ahmed, A. H. Bari, R. Liu, T. Liu, Y. Maret, B. K. Sieu, and T. Sudhakar, "Multi-modal motion-capture-based biometric systems for emergency response and patient rehabilitation," in *Design and Implementation of Healthcare Biometric Systems*. Hershey, PA, USA: IGI Global, 2019, pp. 160–184.

[59] A. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*, 1st ed. Boston, MA, USA: Springer, 2011.

[60] P. P. Paul and M. Gavrilova, "Multimodal Cancelable Biometrics," in *Proc. IEEE 11th Int. Conf. Cognit. Informat. Cogn. Comput.*, Aug. 2012, pp. 43–49.

[61] V. M. Patel, N. K. Ratha, and R. Chellappa, "Cancelable biometrics: A review," *IEEE Signal Process. Mag.*, vol. 32, no. 5, pp. 54–65, Sep. 2015.

[62] J. D. Woodward, "Biometrics: Privacy's foe or privacy's friend?" *Proc. IEEE*, vol. 85, no. 9, pp. 1480–1492, Sep. 1997.

[63] N. Robinson, H. Graux, M. Botterman, and L. Valeri, "Review of EU data protection directive: Summary," in *Proc. Inf. Commissioner's Office*, 2009.

[64] M. Salem, S. Taheri, and J.-S. Yuan, "Utilizing transfer learning and homomorphic encryption in a privacy preserving and secure biometric recognition system," *Computers*, vol. 8, no. 1, p. 3, Dec. 2018.

[65] V. Andronikou, D. S. Demetis, and T. Varvarigou, "Biometric implementations and the implications for security and privacy," *J. Future Identity Inf. Soc.*, vol. 1, no. 1, pp. 20–35, 2005.

FERDOUS AHMED received the B.Sc. degree in computer science from the Islamic University of Technology, in 2014, and the master's degrees in computer science including research areas on machine learning, computer vision, computer graphics, cognitive computing, and behavior modeling. He was a Lecturer with the Islamic University of Technology, Bangladesh. He is currently a Graduate Research Assistant with the University of Calgary, Canada. He has published several peer-reviewed articles in reputed conferences.

A. S. M. HOSSAIN BARI received the B.Sc. degree in computer science and information technology from the Islamic University of Technology, in 2010. He is currently pursuing the M.Sc. degree in computer science with the University of Calgary, Canada, under the supervision of Prof. M. L. Gavrilova. He worked at Samsung Research and Development Institute Bangladesh (SRBD), from November 2010 to August 2018. He has authored over ten international journal and conference papers. He has secured U.S. patents while working at SRBD. His research interests are behavioral biometrics, computer vision, and machine learning.

MARINA L. GAVRILOVA is currently a Full Professor and an Associate Head with the Department of Computer Science and an international expert in the area of biometric security, machine learning, pattern recognition, data analytics, and information fusion. He is also the Co-Founder of the Biometric Technologies Laboratory and the SPARCS Laboratory for interdisciplinary computational sciences research. List of publications includes three coauthored books, over 30 books of conference proceedings, and over 200 peer-reviewed articles on machine learning, biometric security, and multimodal cognitive system architectures. He is also the Founding Editor-in-Chief of *Transactions on Computational Sciences* journal (Springer). He serves on the Editorial Board of the IEEE Transactions on Computational Social Sciences, IEEE Access, *The Visual Computer*, the *International Journal of Biometrics*, and the *International Journal of Cognitive Biometrics*, and on the IEEE Transactions on Biometrics, Behavior, and Identity Science's Steering Committee. His professional excellence and international stature was recognized by Senior ACM and Senior IEEE membership status, and the prestigious Canada Foundation for Innovation, Killam Foundation, and University of Calgary U Make a Difference Awards.