Hierarchical Relaxed Partitioning System for Activity Recognition

Faisal Azhar, *Student Member, IEEE*, and Chang-Tsun Li, *Senior Member, IEEE*

***Abstract*—A hierarchical relaxed partitioning system (HRPS) is proposed for recognizing similar activities which has a feature space with multiple overlaps. Two feature descriptors are built from the human motion analysis of a 2-D stick figure to represent cyclic and noncyclic activities. The HRPS first discerns the pure and impure activities, i.e., with no overlaps and multiple overlaps in the feature space, respectively, then tackles the multiple overlaps problem of the impure activities via an innovative majority voting scheme. The results show that the proposed method robustly recognizes various activities of two different resolution data sets, i.e., low and high (with different views). The advantage of HRPS lies in the real-time speed, ease of implementation and extension, and nonintensive training.**

***Index Terms*—Activity recognition, decision tree (DT), hierarchical relaxed partition, model.**

I. Introduction

Human activity recognition is important due to potential applications in video surveillance, assisted living, animation, etc. [1], [2]. In general, a standard activity recognition framework consists of feature extraction, feature selection (dimension reduction), and pattern classification. Feature extraction can be broadly categorized into the holistic (shape or optical flow) [3]–[6], local feature (descriptors of local regions) [7]–[10], and model-based (prior model) or model-free (no prior model) approaches. Techniques such as principal component analysis (PCA) [11] or linear discriminant analysis [12] are commonly used to select the most prominent features. Decision tree (DT) [3] or support vector machines (SVMs) [2] are used for efficient classification.

Recognizing similar activities in real-time speed still remains a challenge for numerous human activity recognition methods (see Section II). The local feature and holistic approaches are computationally expensive and require intensive training while the model-based/model-free approach is efficient but less accurate. Therefore, the robust and efficient implicit body model-based approach for significant body point (SBP) detection described in [13] is used for feature extraction. In this context, we extend the work in [14] to extract the leg frequency, torso inclination, leg power, and torso power. Also, the SBP detection method is augmented to extract features (similar to [6]) that extract variations in the movement of different body parts at different directions, i.e., up, down, right, and left, during an activity. These features are used to create two feature descriptors.

Most researchers use off-the-shelve classifier such as SVM and DT but with a tradeoff of performance. For example, SVM struggles due to the lack of generalized information, i.e., each test activity is compared with the training activity of one subject [6]. On the other hand DT imposes hard constraint that lead to separation problems when the number of categories increases or when categories are similar, i.e., a lack of clear separation boundary [15]. Similar to DT, hierarchical methods [16], [17] are also used at lower levels for feature-wise classification. The relaxed hierarchy (RH) method in [15] focuses on building high-level class hierarchies and look into the problem of class-wise partitioning. To achieve high accuracy while being fast the RH [15] uses a relaxed constraint, i.e., postpone decisions on confusing classes, to tackle the increased number of categories but still remains inadequate to accurately discern similar categories. The hierarchical strategy (HS) method in [18] uses the RH and groups together easily confused classes to improve the classification performance. RH and HS have only been applied to the spatial domain. We are motivated from the work of RH and HS to perform class-wise partitioning for recognizing similar activities accurately.

We propose a hierarchical relaxed partitioning system (HRPS) (see Section III for details) that classifies and organizes activities in a hierarchical manner according to their type, i.e., pure activities (easily separable) and impure activities (easily confused). Subsequently, it applies relaxed partitioning to all the easily confused activities by postponing the decisions on them until the last level of the hierarchy, where they are labeled by using a novel majority voting scheme (MVS). As opposed to a conventional multiclass classifier as in [18] that can distinguish between only two similar activities, i.e., two classes overlapping simultaneously, the proposed MVS is able to discern between three or more similar activities, i.e., three classes overlapping concurrently. Thus, making the HRPS more robust and suitable for identifying activities in real world scenarios.

The major contributions of this paper are: 1) feature descriptors that represent changes in figure shape characteristics during an activity; 2) expert knowledge at the root node to

Manuscript received May 30, 2015; revised December 12, 2015; accepted January 26, 2016. Date of publication July 12, 2016; date of current version February 14, 2017. This work was supported by the Warwick PostGraduate Research Scholarship. This paper was recommended by Associate Editor J. Basak.

The authors are with the Department of Computer Science, University of Warwick, Coventry CV4 7AL, U.K. (e-mail: faisal.azhar@warwick.ac.uk; c-t.li@warwick.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2016.2526970

split activities into two groups, i.e., significant and no significant translation; and 3) HRPS with a novel MVS to efficiently recognize similar activities.

This paper is organized as follows. Section II reviews related methods. Sections III and IV present the foundation of HRPS and its application to activity recognition, respectively. Experiments are shown in Section V.

II. Literature Review

A. Holistic and Local Feature Approaches

Several human activity recognition methods (see [3], [7], [8], [19]–[24]) verified on the benchmark data sets (see [25] for data sets) struggle in correctly classifying similar activities of the Weizmann data set. The methods in [3], [5], [6], and [10] that are able to correctly classify similar activities of the Weizmann data set are either computationally expensive or require intensive training or need to learn a large set of features. Also, these methods require tuning of parameters with respect to the data set. Therefore, they require extensive retraining for new activities. The bag of words or bag of feature-based methods [7], [8] have high computational cost, requires intensive training, and confuses similar activities.

B. Model-Free and Model-Based Approaches

In model-free methods no prior model is used to determine the SBPs. For example, the method in [14] creates a one-star (a shape that is formed by connecting the center of mass of a human silhouette contour to the extreme boundary points) by using a local maximum on the distance curve of the human contour to locate the SBPs which are at the extremities. It uses two motion features, i.e., leg frequencies and torso angles, to recognize only the *Walk* and *Run* activities. A two star method [26] extends [14] by adding the highest contour point as the second star. It uses a 5-D feature descriptor with a hidden Markov model (HMM) to detect the fence climbing activity. The method in [23] extends [26] by using the medial axis [27] to generate the junction points from which variable star models are constructed. It is compared with [14] and [26] on the fence climbing activity, and evaluated on the Weizmann data set. In [28], multiple cues such as the skin color, principal and minor axes of the human body, the relative distances between convex points, convex point curvature, etc., are used to enhance the method in [14] for the task of posture estimation. It does not provide quantitative results, and uses a nonstandard and nonpublicly available data set. Thus, it requires extensive further work to validate and apply it to activity recognition. The method in [24] assumes that SBPs are given and uses the chaotic invariant for activity recognition on the Weizmann data set. It uses the trajectories of SBPs to reconstruct a phase space, and applies the properties of this phase space such as the Lyapunov exponent, correlation integral and dimension, to construct a feature vector, for activity recognition. The above-described distance curve-based methods are sensitive to the silhouette contour, occlusion, resolution, etc., which affects their accuracy for activity recognition. The methods in [23] and [24] confuse similar activities while only two features of the method in [14] are not sufficient for recognizing more than two similar activities.

In model-based methods a predefine body model is use to determine SBPs. The model-based method in [29] uses the Poisson equation to obtain the torso, and negative minimum curvature to locate extremities which are labeled as SBPs using a 2-D body model. An 8-D feature descriptor from the articulated model is used with the HMM to recognize six activities. In [30], the dominant points along the convex hull of a silhouette contour are used with the body ratio, appearance, etc., to fit a predefined model. It is extended in [31] for activity recognition. These methods are evaluated on nonstandard and publicly unavailable data sets. The method in [32] uses the convex hull with a topological body model to identify the SBPs. However, it is designed to be used for surveillance purposes. In [13] implicit body models are used with the convex hull of a human contour to label SBPs. It tracks the SBPs by using a variant of the particle filter. This method works in real-time by fitting the knowledge from the implicit body models. It outperforms most of the cutting edge methods that use the distance curve method. Thus, we are motivated to extend and apply it for activity recognition.

III. Foundation of Proposed Method-HRPS

Methods like DT and random forest assume that at each node the feature-space can be partitioned into disjoint subspaces, however, as mentioned in [15] this does not hold when there are similar classes or when there are a large number of classes. In this case, finding a feature-space partitioning that reflects the class-set partitioning is difficult as observed in [15]. Therefore, similar to [15] and [18] the goal of this paper is to establish a class hierarchy and then train a classifier such as simple binary classifier at each node of the class hierarchy to perform efficient and accurate classification. This allows us to define different set of rules for classifying different types of activities. This is important as different feature sets are useful for discerning different types of activities [33].

Let us demonstrate the concept of creating an HRPS using a simple example with three overlapping classes (A–C) that represent similar categories as shown in Fig. 1(a). It can be seen from Fig. 1(a) that it is not possible to clearly distinguish between only two overlapping classes by using the RH method as it assumes that only two classes overlap simultaneously. This is because now the overlap is among three classes concurrently, i.e., the overlap between the two classes A and B also contain some overlap with the third class C. Similar phenomena occurs for B and C, and A and C classes. In addition, a combined overlap occurs, i.e., $A \cap B \cap C \neq \emptyset$. Hence, the RH method is not capable of tackling the multiple overlaps class separation problem.

The proposed HRPS method addresses this deficiency in the RH method by splitting the set of classes $K = A' \cup B' \cup C' \cup X$, where $X = X_{AB} \cup X_{BC} \cup X_{AC}$ and $X_{AB} = A \cap B - A \cap B \cap C$, $X_{BC} = B \cap C - A \cap B \cap C$, $X_{AC} = A \cap C - A \cap B \cap C$, and $X_{ABC} = A \cap B \cap C$. X contains samples from two or more overlapping classes. First, at each level of the hierarchy the clearly separable samples of each class are partitioned into

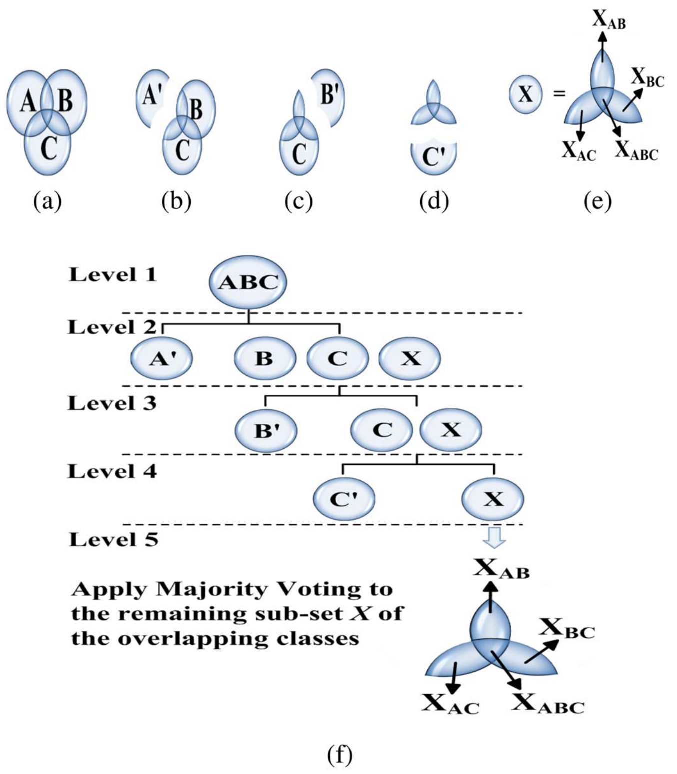


Fig. 1. (a) Example of three classes to illustrate multiple overlaps class separation problem. (b)–(d) HRPS: partition nonoverlapping samples from class A, B, and C, respectively. (e) HRPS: remaining overlapping samples of all the three classes discerned using the MVS (see Section IV-B for details). (f) Corresponding class hierarchy structure.

the A' or B' or C' as shown in Fig. 1(b)–(d)

$$A' = A - X_{AB} - X_{AC} - X_{ABC} \tag{1}$$

$$B' = B - X_{AB} - X_{BC} - X_{ABC} \tag{2}$$

$$C' = C - X_{AC} - X_{BC} - X_{ABC}. \tag{3}$$

Next, the overlapping samples of each class as shown in Fig. 1(e) are partitioned into A or B or C via an MVS (see Section IV-B). The class hierarchy structure for HRPS method is shown in Fig. 1(f). Note that at each level one class is partitioned from the remaining group of easily confused classes [1], [18].

IV. HRPS for Activity Recognition

We present HRPS for the Weizmann data set [34] containing multiple similar activities such as *Walk*, *Run*, *Side*, *Skip*, etc., that can easily confuse the activity recognition methods in the literature. Application of HRPS to the multicamera human action video (MuHAVi) data set [35] containing similar activities, e.g., *Walk*, *Run*, *Turn*, etc., is also described in order to establish its generality, i.e., adaptability to work on a different data set. The work flow of the proposed activity recognition is shown in Fig. 3.

A. Feature Extraction

Distinguishing between the cyclic and noncyclic activities is vital for activity recognition [36]. Thus, we augment our earlier work in [13] to build two feature descriptors D_i, $i = 1, 2$. The 2-D stick figure shown in Fig. 2(a) is used to describe

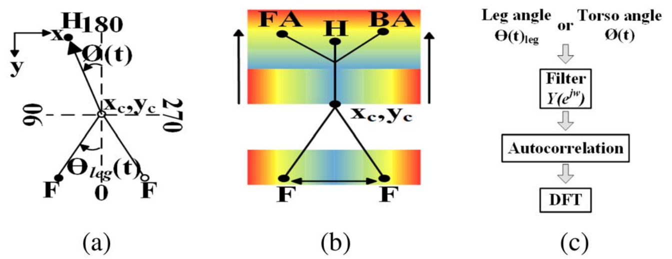


Fig. 2. Feature extraction. (a) 2-D stick figure analysis for cyclic activities. (b) Upper and lower body analysis based on the arm and feet movement. (c) Process of acquiring D_1 for the cyclic activities. The SBPs are labeled as head (H), front arm (FA), back arm (BA), and feet (F).

$$D_1 = [V_1 \;\; V_2 \;\; V_3 \;\; V_4 \;\; V_5] \tag{4}$$

for cyclic activities, while the 2-D stick figure shown in Fig. 2(b) is utilized to build

$$D_2 = [V_6 \;\; V_7 \;\; V_8 \;\; V_9 \;\; V_{10} \;\; V_{11} \;\; V_{12} \;\; V_{13}] \tag{5}$$

for noncyclic activities. The V_i, $i = 1, 2, \ldots, 13$ represents the feature elements of the descriptors as explained later. In Fig. 2, the SBPs are labeled as the head (H), front arm (FA), back arm (BA), and feet (F). Each SBP abbreviation can be considered as a vector which has a 2-D position, e.g., $FA = (x^{FA}, y^{FA})$, $F = (x^F, y^F)$. Here, the superscripts denote the abbreviations of SBP.

The 2-D stick figure motion analysis method in [14] uses two motion-based features, i.e., the leg power and torso inclination angle, to discern between the *Walk* and *Run* activities. This method is suitable for only classifying the cyclic activities with less interclass similarity, i.e., the activities are not similar to each other. Therefore, we propose two more features, i.e., the torso angle and torso power, to strengthen the method in [14]. Given the global angle from contour moments $V_6 = \theta(t)$ at time t, center (x_c, y_c), and SBPs from [13], we extend the method in [14] to acquire D_1 which contains four motion-based features, i.e., the leg cyclic frequency (V_1) and leg power (V_2), the torso inclination angle $V_3 = \phi(t) = |90-(\theta(t)3.14/180)|$, and torso power V_4 for the cyclic activities. The foot point $x^F > x_c$ is used for computing

$$\theta_{\text{leg}}(t) = \tan^{-1}\left(\frac{x^F - x_c}{y^F - y_c}\right). \tag{6}$$

The computed torso angle $V_3 = \phi(t)$ and leg angle $\theta(t)_{\text{leg}}$ are converted into radians. A highpass digital filter $Y(e^{jw})$ is applied to $\theta(t)_{\text{leg}}$

$$Y(e^{jw}) = b(1) - b(2)e^{-jw}. \tag{7}$$

Here, $b(1) = 1$, $b(2) = -0.9$ as in [14]. The filtered leg angles $\theta(t)_{\text{leg}}$ are then autocorrelated in order to emphasize the major cyclic components. The discrete Fourier transform (DFT) is applied to the autocorrelated leg angles to quantify the leg frequency V_1 and magnitude expressed as leg power V_2 in decibels [14] as shown in Fig. 2(c). The proposed activity recognition system also applies the high pass digital filter

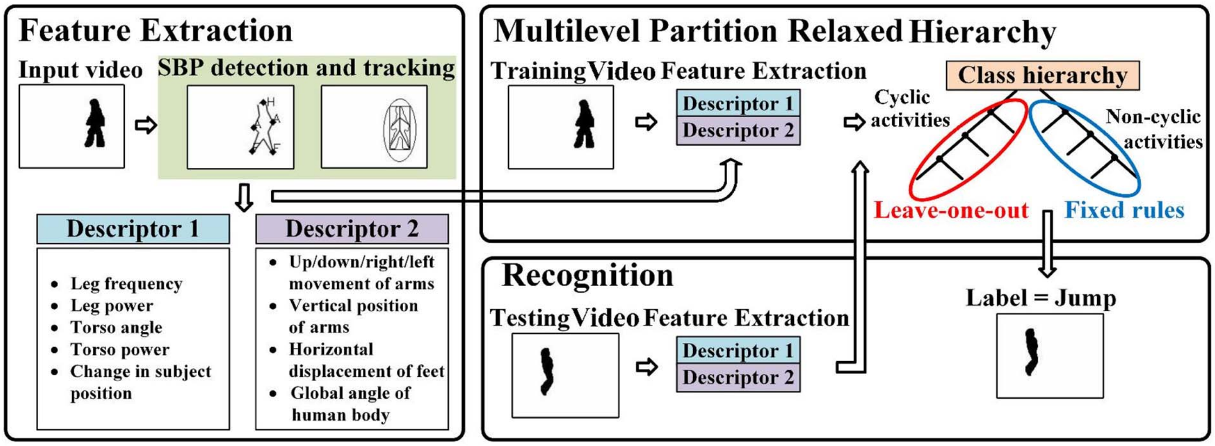


Fig. 3. Main components and work flow of the proposed human activity recognition.

TABLE I
ACRONYMS FOR ACTIVITIES

Type	Activities (α)
1	Walk
2	Run
3	Skip
4	Side
5	Jump
6	Turn

Type	Activities (β)
7	Jump-in-place-on-Two-Legs/Pause Jump
8	Bend
9	One Hand Wave
10	Two Hand Wave
11	Jack
12	Standup
13	Collapse
14	Kick
15	Punch
16	Guard-to-Kick
17	Guard-to-Punch

$Y(e^{jw})$ to the torso angle V_3 (in radians) in order to remove the low frequency components in contrast to [14] where this filter is only applied to the leg angle $\theta(t)_{\text{leg}}$. This high pass filter helps to remove the noise (which appears as large peaks in the low frequency) produced by the autocorrelation process. Next, the autocorrelation and DFT steps in Fig. 2(c) are performed on the filtered torso angle to compute a new feature, i.e., the torso magnitude expressed as torso power V_4 in decibels. The change in direction of movement or position is incorporated as

$$V_5 = \min\left(x_c^{t+1} - x_c^t\right) \tag{8}$$

$\forall\ t \in [1, N-1]$, where N is the total number of frames and min gives the minimum value. A positive and negative value of V_5, respectively, indicate whether subject moved in the same direction or changed the direction (turn around) of movement during an activity.

The feature descriptor D_2 characterizes the upper body (torso and arms) and lower body (legs) movements as a proportion of the mean height μ_h at different directions during an activity as shown in Fig. 2(b) for the noncyclic activities. The interframe displacement (movement) of the front and back arms are described as

$$V_7 = \max\left(\left|x_{t+1}^{FA} - x_t^{FA}\right|\right)\Big/\mu_h,\ V_8 = \max\left(\left|y_{t+1}^{FA} - y_t^{FA}\right|\right)\Big/\mu_h \tag{9}$$

$$V_9 = \max\left(\left|x_{t+1}^{BA} - x_t^{BA}\right|\right)\Big/\mu_h,\ V_{10} = \max\left(\left|y_{t+1}^{BA} - y_t^{BA}\right|\right)\Big/\mu_h \tag{10}$$

$\forall\ t \in 1, [N-1]$, max gives the maximum value. The features V_7–V_{10} do not contain information with respect to the actual positioning of the front and back arm SBPs, i.e., where the arm displacement is being taken place. This information is represented as

$$V_{11} = \min\left(y_t^{FA}\right), \quad V_{12} = \min\left(y_t^{BA}\right), \ \ \forall\ t \in [1, N] \tag{11}$$

which uses the vertical position of the front and back arms to represent their maximum height (as the minimum y location of the front and back arms). The variation in the lower body movement due to the leg can be represented by computing the maximum interframe horizontal displacement between the two feet as

$$V_{13} = \max\left(\left|x_{t+1}^F - x_t^F\right|\right)/\mu_h, \ \ \forall\ t \in [1, N-1]. \tag{12}$$

B. Classification: HRPS for the Weizmann Data Set

The Weizmann data set contain ten activities, i.e., *Walk* (α1), *Run* (α2), *Skip* (α3), *Side* (α4), *Jump* (α5), *Jump-in-Place-on-Two-Legs* or *Pause Jump* (β7), *Bend* (β8), *One Hand Wave* (β9), *Two Hand Wave* (β10), and *Jack* (β11) (see Table I). In [37], a binary DT splits the activities into still and moving categories at the root node in order to obtain better classification. Therefore, motivated by Arbab-Zavar *et al.* [37], we add an expert knowledge at the root node level 1 to automatically split the above-mentioned ten activities in two groups, i.e., significant translation (α) and no significant translation (β) by using

$$\begin{aligned} \alpha &= w_1 I_w > x_c \text{ or } x_c > w_2 I_w \\ \beta &= w_1 I_w < x_c \text{ or } x_c < w_2 I_w \end{aligned} \tag{13}$$

as shown in level 2 of Fig. 4. I_w is the frame width and I_h is the frame height. The weights w_1 and w_2 have been empirically determined as 0.25 and 0.75, respectively. These weights allow us to define a range that is used to determine whether the subject's initial position x_c is within or outside this range

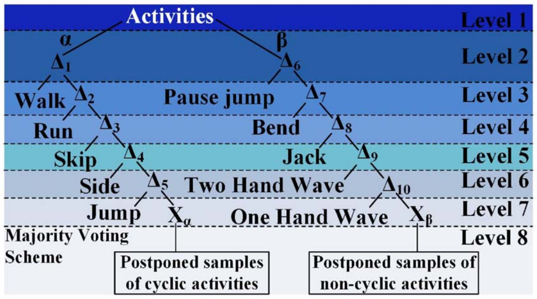


Fig. 4. HRPS for the Weizmann data set. $\Delta_i, i = 1, 2, \ldots, 10$ are the decision rules, and X_α and X_β are the unassigned impure cyclic and noncyclic activities, respectively, with significant multiple overlaps.

by using (13). When the subject's initial position is outside this range the subject is likely to perform an activity with significant movement/translation across the frame, otherwise the subject might perform an activity with no significant translation. Thus, based on (13) most cyclic activities, i.e., *Walk* ($\alpha 1$), *Run* ($\alpha 2$), *Skip* ($\alpha 3$), *Side* ($\alpha 4$), and *Jump* ($\alpha 5$), which have significant translation of the subject and repetitive nature are grouped together under α. The cyclicity of activities with significant translation is captured by using (6) to compute the leg cyclic frequency (V_1) as explained in Section IV-A. The activities, i.e., the *Pause Jump* ($\beta 7$), *Bend* ($\beta 8$), *One Hand Wave* ($\beta 9$), *Two Hand Wave* ($\beta 10$), and *Jack* ($\beta 11$), which have no significant translation of the subject are grouped under β.

An HRPS with eight levels is created with decision rules $\Delta_i, i = 1, 2, \ldots, 10$ as shown in Fig. 4. The decision rules Δ_i, $i = 1, 2, \ldots, 5$ for cyclic activities are learned by using algorithm cyclic activity learning algorithm (CAL) on the training data set that contains the activities performed by eight subjects. The last subject is used as the testing data set in a leave-one-out cross validation approach to determine the performance of the HRPS for cyclic activities. Algorithm CAL postpones decisions on those samples of an activity that are closer to the samples of all the remaining activities by updating the decision rules Δ_i, $i = 1, 2, \ldots, 5$ according to variable adjustment κ. In [13], SBPs were accurately detected by using implicit body models (IBMs) that are based on the human kinesiology and anthropometric studies, and observed human body characteristics. This inspired us to define decision rules Δ_i, $i = 6, 8, \ldots, 10$ that are fixed based on the human kinesiology (torso flexion or extension V_6) [38] and anthropometric studies (upper body motion V_7–V_{10} and leg motion V_{13}) [39], and individual arm location V_{11} and V_{12}), observed human body characteristics and experimental cues for noncyclic activities. The *Pause Jump* ($\beta 7$) is a cyclic activity with no significant translation but has repetitive nature. Thus, it is first separated using (14) from the noncyclic activities, i.e., *Bend* ($\beta 8$), *One Hand Wave* ($\beta 9$), *Two Hand Wave* ($\beta 10$), and *Jack* ($\beta 11$). This knowledge will assure an increase in the accuracy and reliability of the activity classification

$$\Delta_6 = \begin{cases} \beta 7 & \text{if } |90 - V_6| < 9 \\ \Delta 7 & \text{Otherwise.} \end{cases} \tag{14}$$

Algorithm 1 Cyclic Activity Learning Algorithm (D_1)

Input: Training sequences $S_1, \ldots, S_M$
Corresponding labels $y_1, \ldots, y_M$
Feature descriptor $D_1 = [V_1 \ V_2 \ V_3 \ V_4 \ V_5]$
Output: Decision rules Δ_i, $i = 1, 2, \ldots, 5$

1- For each activity, determine the mean μ_i and standard deviation σ_i of feature elements V_i, $i = 1, \ldots, 5$ from K training subjects/samples as

$$\mu_i = \sum_{k=1}^{K} V_i^k / K \quad , \ \sigma_i = \sqrt{1/K \sum_{k=1}^{K} (V_i^k - \mu_i)^2}.$$

2- Learn decision rules as one standard deviation on either side of the mean

$$\Delta_i = \mu_i - \sigma_i < V_i < \mu_i + \sigma_i, i = 1, 2, \ldots, 5.$$

3- Update decision rules by using a variable adjustment κ to separate clearly separable samples, i.e., pure samples, of an activity from the samples of all the remaining activities

$$\Delta_i = \mu_i - \sigma_i + \kappa < V_i < \mu_i + \sigma_i + \kappa, i = 1, 2, \ldots, 5$$

4- Accumulate impure samples of an activity that are closer to the samples of all the remaining activities in X_α.

A full flexion of the vertebra in the *Bend* ($\beta 8$) activity causes a large increase in the torso angle [38]. Based on the experimental observation in Section V-A most training subjects have a torso angle variation greater than 9 degrees, thus

$$\Delta_7 = \begin{cases} \beta 8 & \text{if } |90 - (V_6 180/3.14)| > 9 \\ \Delta 8 & \text{Otherwise.} \end{cases} \tag{15}$$

The *Jack* ($\beta 11$) activity which involves a large upper body and lower body movement is determined based on large arm and feet displacement by using

$$\Delta_8 = \begin{cases} \beta 11 & \text{if } V_7 \text{ or } V_8 > 15/\mu_h \text{ and } V_9 \text{ or } V_{10} > 15/\mu_h \\ & \text{and } V_{13} > 20/\mu_h \\ \Delta 9 & \text{Otherwise.} \end{cases} \tag{16}$$

where $\mu_h = 68$ pixels for the Weizmann data set. The human head is one-eighth the human height, i.e., 0.125. Hence, a 15 pixel movement equates to $15/68 = 0.22$ that is almost twice of the height of the human head.

The individual arm motion in the *Two Hand Wave* ($\beta 10$) and *One Hand Wave* ($\beta 9$) activities is discerned using the location information. In the *Two Hand Wave* ($\beta 10$) activity there will be significant movement of both arms while in the *One Hand Wave* ($\beta 9$) activity there will be significant movement of only one arm. Therefore, the *Two Hand Wave* ($\beta 10$) and *One Hand Wave* ($\beta 9$) activities are

$$\Delta_9 = \begin{cases} \beta 10 & \text{if } V_{13} < 20/\mu_h \text{ and } V_8 \geq 5/\mu_h \text{ and} \\ & V_{10} \geq 5/\mu_h \text{ and } V_{11} \leq 55 \text{ and } V_{12} < 50 \\ \Delta 10 & \text{Otherwise} \end{cases} \tag{17}$$

$$\Delta_{10} = \begin{cases} \beta 9 & \text{if } V_{13} < 20/\mu_h \text{ and } V_8 \text{ or } V_{10} \leq 8/\mu_h \\ & \text{and } V_{11} \leq 55 \text{ and } V_{12} > 50 \\ X_\beta & \text{Otherwise.} \end{cases} \tag{18}$$

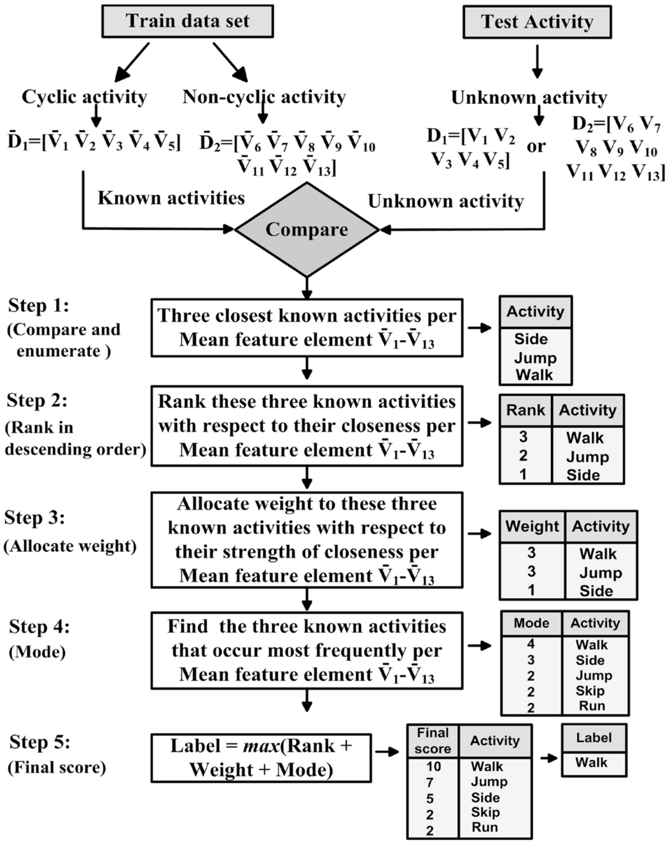

Train data set
Test Activity
Cyclic activity
Non-cyclic activity
Unknown activity
D̄₁=[V̄₁ V̄₂ V̄₃ V̄₄ V̄₅]
D̄₂=[V̄₆ V̄₇ V̄₈ V̄₉ V̄₁₀ V̄₁₁ V̄₁₂ V̄₁₃]
D₁=[V₁ V₂ V₃ V₄ V₅]
or
D₂=[V₆ V₇ V₈ V₉ V₁₀ V₁₁ V₁₂ V₁₃]
Known activities
Unknown activity
Compare
Step 1: (Compare and enumerate)
Three closest known activities per Mean feature element V̄₁-V̄₁₃
Activity
Side
Jump
Walk
Step 2: (Rank in descending order)
Rank these three known activities with respect to their closeness per Mean feature element V̄₁-V̄₁₃
Rank Activity
3 Walk
2 Jump
1 Side
Step 3: (Allocate weight)
Allocate weight to these three known activities with respect to their strength of closeness per Mean feature element V̄₁-V̄₁₃
Weight Activity
3 Walk
3 Jump
1 Side
Step 4: (Mode)
Find the three known activities that occur most frequently per Mean feature element V̄₁-V̄₁₃
Mode Activity
4 Walk
3 Side
2 Jump
2 Skip
2 Run
Step 5: (Final score)
Label = max(Rank + Weight + Mode)
Final score Activity
10 Walk
7 Jump
5 Side
2 Skip
2 Run
Label
Walk

Fig. 5. Proposed MVS for the unassigned impure activities X_α and X_β using the mean $\bar{D}_i$, $i = 1, 2$.

1) Majority Voting Scheme: The justification for using the proposed MVS is based on the fact that its design and accumulated voting criteria is better suited to recognize three or more similar activities, i.e., three classes overlapping simultaneously (see Section III for details). Also, the current state-of-the-art methods, i.e., RH and HS (using the conventional multi-class classifier) can distinguish between only two similar categories/activities.

The HRPS postpones decisions on those samples of an activity that are closer to samples of all the remaining activities, so that they trickle to the bottom where they are captured at the second last level (see Fig. 4). These unassigned activities are supplied to a novel MVS for classification at the last level of the HRPS. The key idea of this scheme is to accumulate votes based on the rank, assigned weight and frequency (mode) value in order to deduce more accurate decisions for the unassigned activities in X_α and X_β (see Fig. 4).

As shown in Fig. 5, given the mean feature descriptors, i.e., $\bar{D}_1 = [\bar{V}_1 \; \bar{V}_2 \; \bar{V}_3 \; \bar{V}_4 \; \bar{V}_5]$ and $\bar{D}_2 = [\bar{V}_5 \; \bar{V}_6 \; \bar{V}_7 \; \bar{V}_8 \; \bar{V}_9 \; \bar{V}_{10} \; \bar{V}_{11} \; \bar{V}_{12}]$, of the known activities of the training data set, the goal is to label an unknown impure activity (which contain significant overlaps in the feature space) by extracting the feature descriptors, i.e., $D_1 = [V_1 \; V_2 \; V_3 \; V_4 \; V_5]$ and $D_2 = [V_6 \; V_7 \; V_8 \; V_9 \; V_{10} \; V_{11} \; V_{12} \; V_{13}]$, in order to calculate the rank, weight, and mode as shown in Fig. 5. D_1 and D_2 are used for cyclic and noncyclic activities, respectively. V_1–V_{13} represent each feature element of the feature descriptors. The label for the unknown impure activity is determined as follows.

1) *Step 1:* Compare each feature element of the feature descriptor, i.e., D_1 or D_2, of one unknown impure activity with the respective mean feature elements of the feature descriptor, i.e., $\bar{D}_1$ or $\bar{D}_2$, for each of the known activities in order to enumerate three closest known activities per mean feature element.
2) *Step 2:* Assign a score (rank) $\nu = 3, 2, 1$ to the three activities enumerated in step 1 based on their closeness to each of the mean feature elements of $\bar{D}_1$ or $\bar{D}_2$. Next, arrange them in the descending order of their ranks.
3) *Step 3:* Allocate a weight $\omega = 3, 2, 1$ to the three ranked activities in step 2 based on their strength of closeness to the mean feature elements of $\bar{D}_1$ or $\bar{D}_2$.
4) *Step 4:* Find the three known activities that occur most frequently (i.e., mode ϖ) per mean feature element of $\bar{D}_1$ or $\bar{D}_2$.
5) *Step 5:* Calculate the final score to find the label of the unknown activity. The known activity of the training data set whose rank, weight, and mode yield the maximum score with respect to the unknown activity is assigned as the label for the unknown activity, i.e., Label $= \max(\varpi + \nu + \omega)$. This metric has been selected based on empirical analysis on the training data set to obtain an optimal decision.

Fig. 5 also shows how the label for an unknown impure activity, e.g., *Walk*, is determined using the MVS. According to step 1, each feature element of unknown activity is compared with the mean feature elements to enumerate the three closest known activities as *Side*, *Jump*, and *Walk*. In step 2, their rank value is computed with respect to their closeness per mean feature element. In step 3, the weights associated to each activity is found to represent the strength of closeness of these activities to the unknown activity. In step 4, the known activities that occur most frequently are counted per mean feature element. In step 5, the final scores for each known activity is calculated as an accumulation of the rank, weight and mode values. The known activity with the maximum score is the correct label for the unknown activity.

C. Classification: HRPS for the MuHAVi Data Set

The generality of the proposed HRPS method is further validated by applying it with the same feature descriptors D_i, $i = 1, 2$ and expert knowledge on the MuHAVi dataset [35]. The MuHAVi data set contains eight activities, i.e., *Walk* ($\alpha 1$), *Run* ($\alpha 2$), *Turn* ($\alpha 6$), *Standup* ($\beta 12$), *Collapse* ($\beta 13$), *Kick* ($\beta 14$), *Punch* ($\beta 15$), and *Guard-to-Kick* or *Guard-to-Punch* ($\beta 16/\beta 17$) (see Table I). As in Section IV-B the root node is split into α and β activities by using (13). An HRPS with seven levels is created with decision rules Δ_i, $i = 11, \ldots, 19$ as shown in Fig. 6. Algorithm CAL is used on the seven training samples of the MuHAVi data set to learn the decision rules Δ_i, $i = 11, 12, 13$ for the *Walk* ($\alpha 1$), *Run* ($\alpha 2$), and *Turn* ($\alpha 6$) cyclic activities, respectively. The last sample is used as the testing data in a leave-one-out procedure to determine the performance of the HRPS.

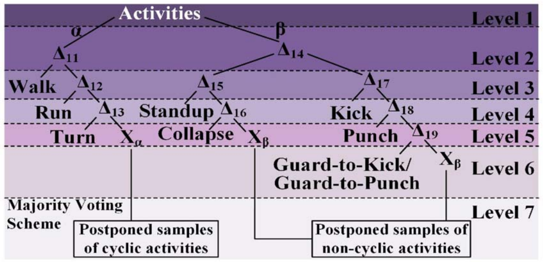


Fig. 6. HRPS for the MuHAVi data set. Δ_i, $i = 11, 12, \ldots, 19$ are the decision rules, and X_α and X_β are the unassigned impure cyclic and noncyclic activities, respectively, with significant multiple overlaps.

Similar to Section IV-B, we define decision rules Δ_i, $i = 14, \ldots, 19$ that are fixed based on the human kinesiology [38], anthropometry [39], and body characteristics for noncyclic activities. Let the reference global angle $V_6 = \theta(t)$ in Stand posture be 90°. Then, based on biomechanical analysis [40] of human spine the maximum flexion of torso is 60°, i.e., $(90 - 60 = 30$ or $90 + 60 = 150)$, which causes a significant change in posture. Thus

$$\Delta_{14} = \begin{cases} \Delta 15 & \text{if } 30 \geq V_6 \geq 150 \\ \Delta 17 & \text{Otherwise} \end{cases} \tag{19}$$

is used to determine whether a transition occurred $\forall t \in [1, N]$ frames of the activity video. The transition Δ_{15} includes *Standup* ($\beta 12$) and *Collapse* ($\beta 13$) activities which contain significant change in posture while the nontransition Δ_{16} contain *Kick* ($\beta 14$), *Punch* ($\beta 15$), and *Guard-to-Kick* or *Guard-to-Punch* ($\beta 16/\beta 17$) which do not have significant change in posture. The decision rules for the *Standup* ($\beta 12$) and *Collapse* ($\beta 13$), i.e., Δ_{15} and Δ_{16}, respectively, are defined as

$$\Delta_{15} = \begin{cases} \beta 12 & \text{if } 30 \geq V_6 \geq 150, \text{ at } t = 1 \\ & \text{and } 65 \leq V_6 \leq 125, \forall\ t \in 2, N \\ \Delta 16 & \text{Otherwise} \end{cases} \tag{20}$$

$$\Delta_{16} = \begin{cases} \beta 13 & \text{if } 65 \leq V_6 \leq 125, \text{ at } t = 1 \\ & \text{and } 30 \geq V_6 \geq 150, \forall\ t \in 2, N \\ X_\beta & \text{Otherwise.} \end{cases} \tag{21}$$

The range $125 - 65 = 60°$ [40] is selected as it corresponds to the flexion and extension range of the human body while maintaining a somewhat Stand posture. The decision rules Δ_{17} to Δ_{19} are defined based on the empirical analysis of the body characteristics in [13]. Hence, for the *Kick* ($\beta 14$) and *Punch* ($\beta 15$) activities

$$\Delta_{17} = \begin{cases} \beta 14 & \text{if } 2 \leq 90 - V_6 \leq 15 \\ \Delta 18 & \text{Otherwise} \end{cases} \tag{22}$$

$$\Delta_{18} = \begin{cases} \beta 15 & \text{if } 90 - V_6 > 15 \\ \Delta 19 & \text{Otherwise.} \end{cases} \tag{23}$$

Note that in *Punch* ($\beta 15$), the arm moves across the body in a diagonal manner and as a result the angle of the body from the vertical is quite large. The *Guard-to-Punch* and *Guard-to-Kick* are considered as one class because both primarily have a guard activity with minimal movement of the arms and legs. In *Guard-to-Kick* or *Guard-to-Punch* ($\beta 16/\beta 17$), the human remains in the Stand posture with the least angle of the body from the vertical. Hence

$$\Delta_{19} = \begin{cases} \beta 16/\beta 17 & \text{if } 90 - V_6 < 2 \\ X_\beta & \text{Otherwise.} \end{cases} \tag{24}$$

The unassigned impure activities X_α and X_β are given a label by using the MVS (see Section IV-B).

V. Experimental Results

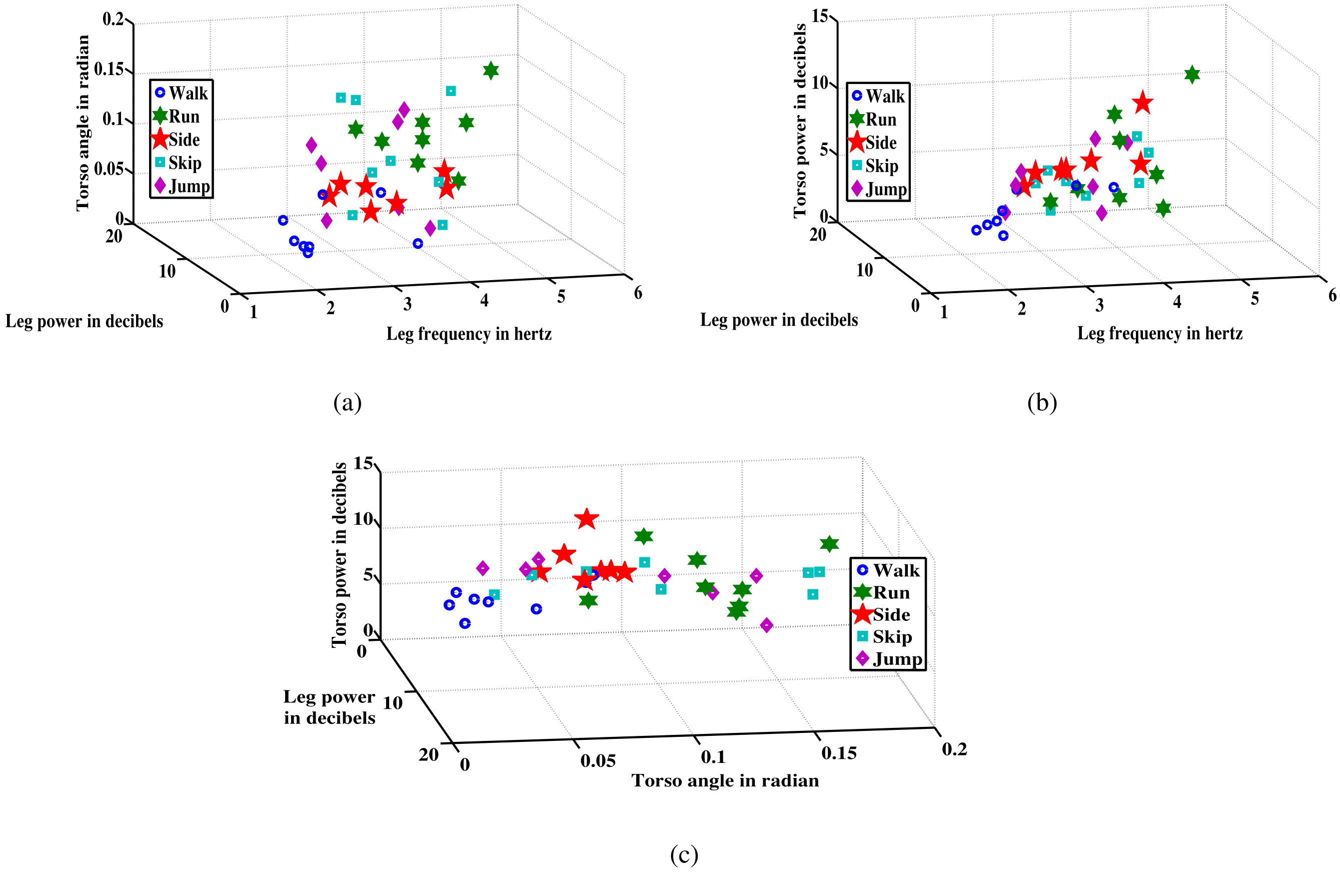
We have used two standard publically available data sets, i.e., Weizmann and MuHAVi, with a standard leave-one-out cross validation method to ensure a correct comparative study, i.e., with same environment and data set. The Weizmann data set [34] contains low resolution videos 180 × 144, imperfect silhouettes and ten routine activities performed by nine subject. In contrast, the MuHAVi data set [35] contains high resolution videos 720 × 576, perfect silhouettes and nine routine and nonroutine activities of two actors with two samples with two different views (cameras 3 and 4), i.e., in total eight samples, per activity. The activities and their acronyms are shown in Table I.

The main challenges of the Weizmann data set are as follows: 1) low resolution videos make it challenging to detect body parts; 2) rapid limb movements make it difficult to track body parts in self occlusion; and 3) very similar activities are difficult to recognize. The main challenges of the MuHAVi data set are as follows: 1) rapid change of posture including mild-to-severe occlusion makes it difficult to detect and track body parts and 2) similar activities and activities with similar postural changes are difficult to recognize. In addition, both the data sets contain background illumination variation and subjects of different height and built. Therefore, the proposed HRPS method has been verified on two extremely challenging data sets.

A. Feature Descriptors Evaluation

The 3-D scatter plots of the selected features are shown in Figs. 7 and 8 to visualize the distribution of the activities of the input data set. It can be seen from Fig. 7(a) that the *Walk* activity has the least leg frequency (most blue circles are between 2–3 Hz) and the *Run* activity has the maximum leg frequency (green pentagons lie between 4–6 Hz onward). Similarly, it can be seen in Fig. 7(b) that the torso power of the *Walk* activity is much less than the remaining cyclic activities. In Fig. 7(c) it can be seen that the torso angle of most of the *Run* (green pentagons), *Jump* (purple diamonds), and *Skip* (light blue square) activities is greater than the *Walk* (blue circles) and *Side* (red stars) activities. It can be observed from Fig. 7(c) that the *Walk* activity has the least torso angle (blue circles between 0–0.05 radian) while the torso angle for the *Side* (red stars) activity is concentrated between 0.05–0.1 radian.

Fig. 8(a) shows the 3-D scatter plots of the selected features for the *Bend*, *Jack*, *One Hand Wave*, and *Two Hand Wave* activities of the Weizmann data set. It can be seen that the *Jack* activity has the maximum displacement of the feet as a proportion of the mean height of the subject. Also, it

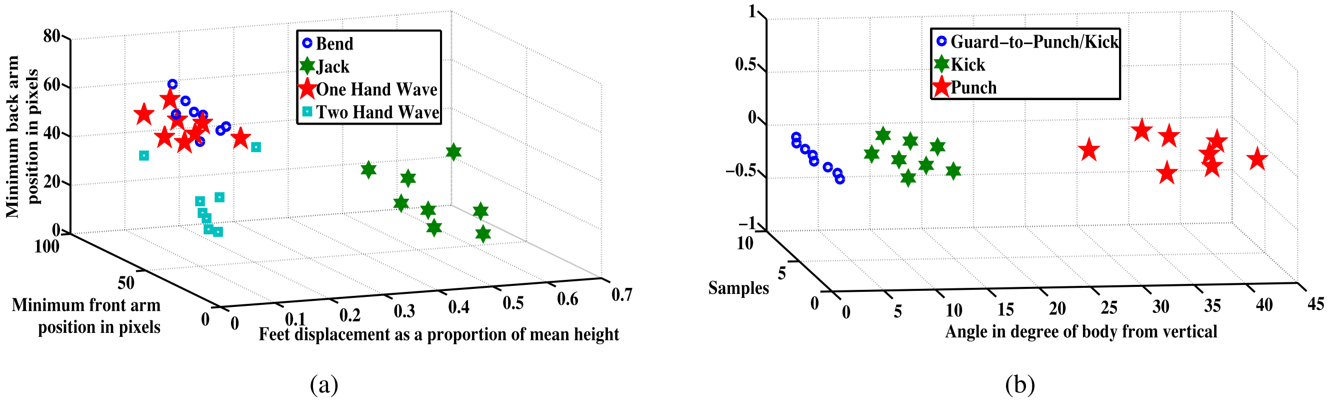


(a)

(b)

(c)

Fig. 7. 3-D scatter plots of the selected features that show the distribution of the cyclic activities for the input Weizmann data set. (a) Leg power, leg frequency, and torso angle, (b) Leg power, leg frequency, and torso power. (c) Leg power, torso angle and torso power.



(a)

(b)

Fig. 8. 3-D scatter plots of the selected features that show the distribution of the activities for the input Weizmann and MuHAVi data sets. (a) Front and back arm position, and feet displacement. (b) Leg power, leg frequency and torso power.

can be seen that in the *Two Hand Wave* (light blue square) activity both front and back arm have minimum position in pixels, and is well separate from the *One Hand Wave* (red star) activity. Fig. 8(b) shows the 3-D scatter plots of a selected feature for the *Guard-to-Punch* or *Guard-to-Kick*, *Kick*, and *Punch* activities of the MuHAVi data set. It can be seen that the *Guard-to-Punch* or *Guard-to-Kick* has the least variation in the angle of the body from the vertical and the *Punch* has the maximum angle of the body from the vertical. The angle of the body from the vertical for the *Kick* activity lies in between the *Guard-to-Punch* or *Guard-to-Kick* and *Punch* activities.

In Fig. 9, we illustrate the ability of some of the features from D_i, $i = 1, 2$ to discern various human activities of the Weizmann and MuHAVi data sets. The error bars show 95% confidence intervals on selected features with two standard deviations as an error metric. Although the leg frequency, i.e., V_1, of the *Walk* ($\alpha 1$) and *Run* ($\alpha 2$) activities is dissimilar based on the speed of the leg movement, anomalies like some subjects walking faster causes misclassification. However, it can be seen from Fig. 9(a) that the torso angle $V_3 = \phi(t)$ provides a good separation to discern the *Walk* ($\alpha 1$) and *Run* ($\alpha 2$) activities. Similarly, the newly introduced torso power feature V_4 provides a reasonable distinction between the *Side* ($\alpha 4$)

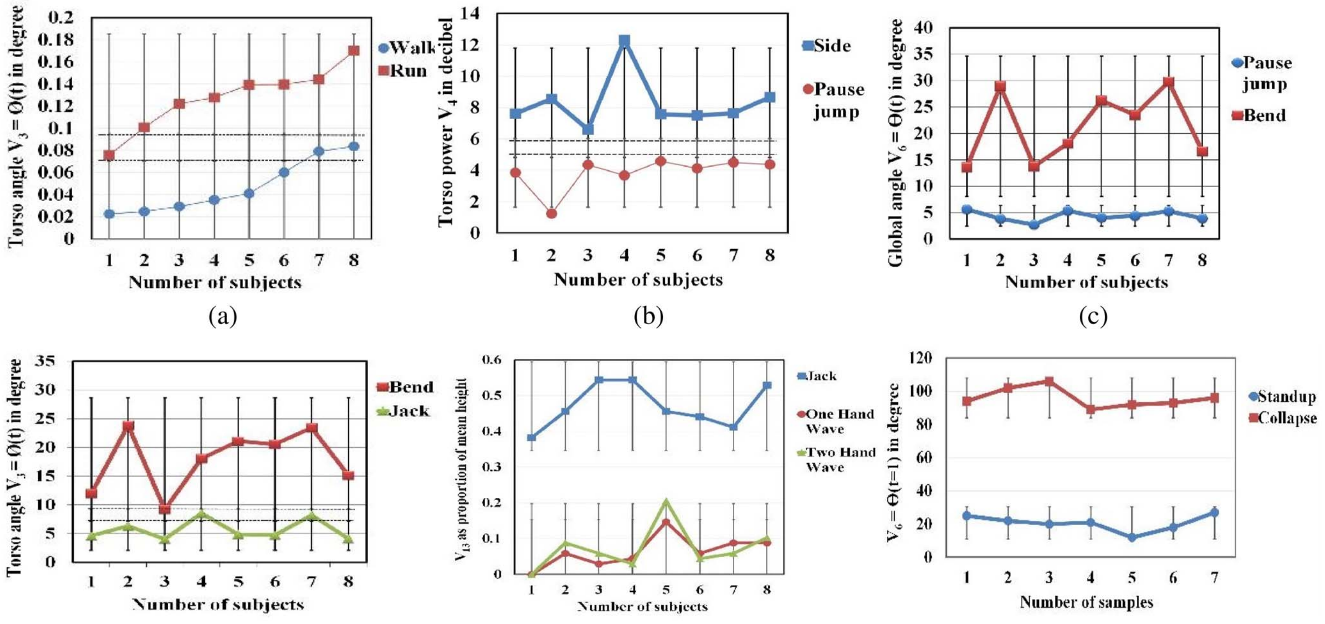


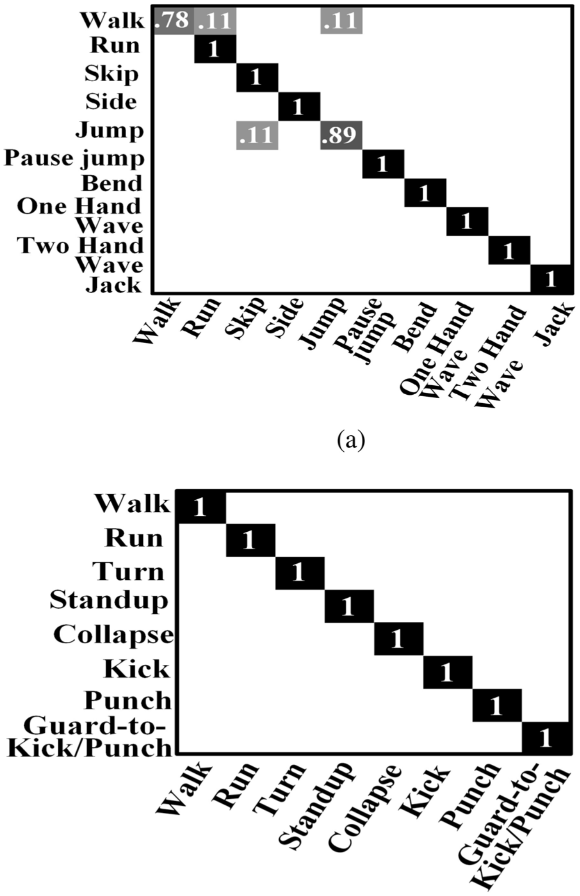
Fig. 9. Significance of the extracted features for discerning activities. Error bars show 95% confidence intervals on selected features with two standard deviations as an error metric. (a)–(e) Weizmann data set. (f) MuHAVi data set.

and *Pause Jump* ($\beta 7$) activities as shown in Fig. 9(b). In Fig. 9(c), the global angle $V_6 = \theta(t)$ provides clear separation between the *Pause Jump* ($\beta 7$) and *Bend* ($\beta 8$) activities while in Fig. 9(d) the torso angle $V_3 = \phi(t)$ provides sufficient discerning ability between the *Bend* ($\beta 8$) and *Jack* ($\beta 11$) activities. It can be observed from Fig. 9(e) that the distance between the legs, i.e., V_{13}, gives a very good separation among the *Jack* ($\beta 11$), *One Hand Wave* ($\beta 9$), and *Two Hand Wave* ($\beta 10$) activities. Finally, in Fig. 9(f) the global angle $V_6 = \theta(t = 1)$ easily discern the *Standup* ($\beta 12$) and *Collapse* ($\beta 12 = 3$) activities. Thus, the D_i, $i = 1, 2$ acquires meaningful information. However, there is a slight overlap in the confidence intervals of some of the features, e.g., Fig. 9(a), (b), and (d). This illustrate the importance of using HRPS to postpone decisions on such samples that lie closer to the samples of another activity. Also, for these samples the MVS is better suited because it takes into account multiple criteria based on the average values of all the feature elements obtained from the training data set to assign a label to an unknown activity. As stated in [6] the average features provide more generalized information about the movement pattern of the body during an activity.

B. Classification Evaluation

The confusion tables for the HRPS method on the Weizmann and MuHAVi data set are shown in Fig. 10(a) and (b), respectively. We obtained a mean classification accuracy of 96.7% for ten activities of the Weizmann data set (see Table II and details below for significance in comparison to other methods). This shows that our method robustly recognizes activities that have significant multiple overlaps in the feature space. In particular, our method recognizes four activities, i.e., *Run* ($\alpha 2$), *Side* ($\alpha 4$), *Jump* ($\alpha 5$), and *Pause Jump* ($\beta 13$), out of the six cyclic activities with a mean classification accuracy of 100%. Thus, our method robustly discerns similar cyclic activities. It obtains a mean classification accuracy of 94.5% for all the six cyclic activities, i.e., *Walk* ($\alpha 1$), *Run* ($\alpha 2$), *Side* ($\alpha 4$), *Jump* ($\alpha 5$), *Skip* ($\alpha 3$), and *Pause Jump* ($\beta 13$). The decomposition of the *Walk* ($\alpha 1$) into the *Run* ($\alpha 2$) and *Jump* ($\alpha 5$) activities is reasonable due to similar motion. Also, the *Skip* ($\alpha 3$) and *Jump* ($\alpha 5$) activities are similar in the way the subject bounces across the video. The noncyclic activities, i.e., *Bend* ($\beta 14$), *Jack* ($\beta 11$), *Two Hand Wave* ($\beta 10$) and *One Hand Wave* ($\beta 15$) are robustly classified with a mean classification accuracy of 100%. This proves that the decision rules based on human kinesiology and body characteristics work well. We obtained a mean classification accuracy of 100% for eight activities of the MuHAVi data set as shown in Fig. 10(b). The results demonstrate that the proposed HRPS method can robustly distinguish various activities in two different (low and high) resolution data sets. It also shows that our method performs well under different views, i.e., cameras 3 and 4, for the MuHAVi data set. A high accuracy on the *Standup* ($\beta 12$), *Collapse* ($\beta 13$), *Kick* ($\beta 14$), *Punch* ($\beta 15$), and *Guard-to-Kick* or *Guard-to-Punch* ($\beta 16/\beta 17$) activities demonstrates the importance of decision rules based on human kinesiology and body characteristics.

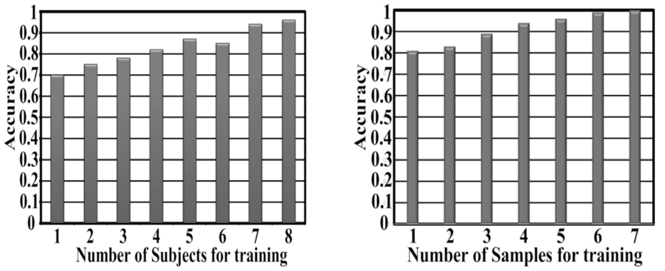
Fig. 11(a) shows the HRPS's classification performance with respect to the training subjects of the Weizmann data set. It can be seen that the classification accuracy of the proposed method is about 70% with only one training subject. However, as the number of training subjects increase the classification accuracy also improves. The best performance is achieved with eight training subjects. The classification performance with respect to the training samples of the MuHAVi data set



(a)

(b)

Fig. 10. Confusion table (see Table I for α and β). (a) Weizmann data set. (b) MuHAVi data set.



(a)

(b)

Fig. 11. Classification performance. (a) Weizmann data set. (b) MuHAVi data set.

TABLE II
COMPARISON ON THE WEIZMANN DATA SET

Method	Accuracy%	Real-time	Intensive training	Year
Vrigkas, et al. [5]	100	**No**	**Yes**	2014
Alcantara et al. [22]	96.7	Yes	No	2014
Mahbub, et al. [6]	100	**No**	No	2014
Ma, at al. [10]	100	**No**	**Yes**	2013
Tavenard, et al. [8]	82.79	**No**	**Yes**	2013
Jiang, et al. [3]	100	Yes	**Yes**	2012
Ali, et al. [7]	95.75	**No**	**Yes**	2010
Yu, et al. [23]	93.6	Yes	No	2009
Ali, et al. [24]	92.6	-	No	2007
Our method	**96.7**	**Yes**	**No**	2014

TABLE III
COMPARISON ON THE MUHAVI DATA SET

Method	Accuracy%	Real-time	Intensive training	Year
Chaaraoui, et al. [42]	100	Yes	No	2014
Alcantara, et al. [22]	100	Yes	No	2014
Chaaraoui, et al. [41]	97.1	Yes	No	2013
Eweiwi, et al. [43]	98.5	**No**	No	2011
Singh, et al. [35]	97.8	Yes	No	2010
Martinez, et al. [44]	98.4	**No**	**Yes**	2009
Our method	**100**	**Yes**	**No**	2014

is shown in Fig. 11(b). It can be seen that the classification performance increases steadily till it reaches 100% with seven samples used for training.

Table II compares the HRPS with relevant state-of-the-art methods (see Section II) for activity recognition on the Weizmann data set. It shows that our method outperforms the methods in [7], [8], [23], and [24] in terms of accuracy. Ali *et al.* [24] only deal with nine activities. The method in [5]–[8] and [10] are not real-time since they require intensive training for learning the vocabulary. Jiang *et al.* [3] required both shape and motion features to achieve 100% accuracy. On a similar basis, i.e., using motion features, they obtain 88.89% accuracy while our method obtains 96.7%. Their method is reported to be fast but requires intensive training and uses optical flow which is usually computationally expensive. Hence, these methods are not suitable for real-world applications. In contrast, our method operates in real-time, avoid intensive training, and it is simple to implement and extend for new activity categories (i.e., for each new category new features can be added to the HRPS). This makes it more suitable for real world applications. The model-free method in [14] recognizes only two activities, i.e., *Walk* and *Run* with 97% accuracy. On similar activities, i.e., *Walk* ($\alpha1$), *Run* ($\alpha2$), and *Jump* ($\alpha5$), the method in [29] has mean classification accuracy of 82.4% while we obtain 92.7% mean classification accuracy. Although, the method in [41] can work in real-time, it achieves only 90.32% on the Weizmann data set.

In Table III, our HRPS method is compared with recent methods on the MuHAVi data set. Our method achieved better recognition rate than most of the methods and works in real-time with no intensive training. On both data sets our method is comparable to the method in [22].

In order to avoid blaming heavy training for high accuracy of the HRPS one can either perform another experiment with a new data set or alternatively use the HRPS trained on one data set to recognize same activities present in the other data set. The alternative approach might be more appropriate because a heavily trained HRPS on one data set will not work well for another data set due to overfitting. If the HRPS works well then one cannot blame heavy training for its high accuracy. The alternative approach also allows to verify whether the decision rules learned on one data set are generic enough to recognize same activities of another data set. Table IV shows result of

TABLE IV
RECOGNITION ACCURACY OF HRPS ON SAME ACTIVITIES WITH DIFFERENT TRAINING DATA SET

Training dataset	Testing data set	Activity	Accuracy %
Weizmann	MuHAVi	Walk	100
		Run	100
MuHAVi	Weizmann	Walk	100
		Run	100

recognizing the *Walk* and *Run* activities of the MuHAVi data set using the HRPS trained on the Weizmann data set, and vice versa. The 100% recognition accuracy shows that the proposed HRPS is generic and heavy training cannot be blamed for its high accuracy.

C. Computational Complexity

The feature extraction of our HRPS method computes convex hull using the Sklanskys algorithm which has a computational complexity of $O(N)$, where N in the number of convex points. The contour moments algorithm is based on the Green theorem which has a computational complexity of $O(L)$, where L is the length of the boundary of the object. The particle filter with $N = 100$ particles has an approximate complexity of $O(N)$. DFT has $O(N \log N)$ complexity.

The optical flow Lucas–Kanade method has a time complexity of $O(n^2 N + n^3)$, where n is the number of warp parameters and N is the number of pixels. k-mean clustering is $O(nkdi)$ complex, where n is the number of d-dimensional vectors, k the number of clusters and i the number of iterations needed until convergence. The computational complexity of the expectation maximization algorithm for Gaussian mixture models (GMMs) is $O(iND^2)$, where i is the number of iterations performed, N is the number of samples, and D is the dimensionality of the state. Time complexity of PCA is $O(p^2 n + p^3)$, where n is the number of data points and each point is represented with p features. Locality preserving projection (LPP) algorithm is $O((n+k)m^2 + (n+d)n^2)$, where n is dimensions, m is data points, d is dimension of subspace, and k is the number of nearest neighbor. For k nearest neighbor search, the complexity is $O((n+k)m^2)$. The complexity of singular value decomposition (SVD) is $O(n^3)$. A standard DT has a time complexity of $\mathrm{O}(MN^2)$, where M is the size of the training data and N is the number of attributes. Time complexity of SVM is $O(n^3)$, where n is number of pattern. HMM has a time complexity of $O(N^T T)$, where N is state paths and T is the length of paths.

The method in [3] and [5]–[7] uses optical flow method. The method in [3] and [5], respectively, use k-means and GMMs for clustering. Also, the method in [43] uses k-means clustering. The method in [5] and [7] uses PCA for dimension reduction. The method in [10] uses LPP. The method in [3] uses DT, the method in [22] uses SVM or k-nearest neighbor, and the method in [23] and [44] uses HMM for activity recognition.

On Intel Core i7 2.93 GHz with 4 GB RAM and Windows 7, the feature extraction in OpenCV 2.4.6 takes 0.031 and 0.071 s per image frame on the Weizmann and MuHAVi data sets, respectively. The classification in MATLAB takes 0.183 s for all activities. Alcantara *et al.*'s [22] method takes 4.85 and 2859.29 s for feature extraction on the Weizmann and MuHAVi data sets, respectively. This demonstrates that the HRPS method works in real-time.

VI. CONCLUSION

In light of the inadequacy of existing activity recognition methods, we proposed an HRPS to efficiently and robustly recognize activities. Our method first discerns the pure activities from the impure activities, and then tackles the multiple overlaps problem of the impure activities via an innovative MVS. The results proved that our method not only accurately discerns similar activities but also obtains real-time recognition on two (low and high) resolution data sets, i.e., Weizmann and MuHAVi, respectively. It also performs well under two different views of the MuHAVi data set. These attributes make our method more suitable for real-world applications in comparison to the state-of-the-art methods.

REFERENCES

[1] P. C. Ribeiro and J. Santos-Victor, "Human activity recognition from video: Modeling, feature selection and classification architecture," in *Proc. Int. Workshop Human Activity Recognit. Model.*, 2005, pp. 61–70.

[2] H. Qian, Y. Mao, W. Xiang, and Z. Wang, "Recognition of human activities using SVM multi-class classifier," *Pattern Recognit. Lett.*, vol. 31, no. 2, pp. 100–111, 2010.

[3] Z. Jiang, Z. Lin, and L. Davis, "Recognizing human actions by learning and matching shape-motion prototype trees," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 533–547, Mar. 2012.

[4] M. Lucena, N. P. de la Blanca, and J. M. Fuertes, "Human action recognition based on aggregated local motion estimates," *Mach. Vis. Appl.*, vol. 23, no. 1, pp. 135–150, 2012.

[5] M. Vrigkas, V. Karavasilis, and C. Nikou, "Matching mixtures of trajectories for human action recognition," *Comput. Vis. Image Understand.*, vol. 19, pp. 27–40, Jan. 2014.

[6] U. Mahbub, H. Imtiaz, and M. A. R. Ahad, "Action recognition based on statistical analysis from clustered flow vectors," *Signal Image Video Process.*, vol. 8, no. 2, pp. 243–253, 2014.

[7] S. Ali and M. Shah, "Human action recognition in videos using kinematic features and multiple instance learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 2, pp. 288–303, Feb. 2010.

[8] R. Tavenard, R. Emonet, and J. M. Odobez, "Time-sensitive topic models for action recognition in videos," in *Proc. IEEE Int. Conf. Image Process.*, Melbourne, VIC, Australia, 2013, pp. 2988–2992.

[9] I. Jargalsaikhan, S. Little, C. Direkoglu, and N. E. O'Connor, "Action recognition based on sparse motion trajectories," in *Proc. IEEE Int. Conf. Image Process.*, Melbourne, VIC, Australia, 2013, pp. 3982–3985.

[10] A. J. Ma, P. C. Yuen, W. W. W. Zou, and J.-H. Lai, "Supervised spatio-temporal neighborhood topology learning for action recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 8, pp. 1447–1460, Aug. 2013.

[11] K. Schindler and L. van Gool, "Action snippets: How many frames does human action recognition require?" in *Proc. IEEE Int. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.

[12] K. Mikolajczyk and H. Uemura, "Action recognition with motion-appearance vocabulary forest," in *Proc. IEEE Int. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.

[13] F. Azhar and T. Tjahjadi, "Significant body point labeling and tracking," *IEEE Trans. Cybern.*, vol. 44, no. 9, pp. 1673–1685, Sep. 2014.

[14] H. Fujiyoshi, A. J. Lipton, and T. Kanade, "Real-time human motion analysis by image skeletonization," *IEICE Trans. Inf. Syst. E Series D*, vol. 87, no. 1, pp. 113–120, 2004.

[15] M. Marszałek and C. Schmid, "Constructing category hierarchies for visual recognition," in *Proc. Eur. Conf. Comput. Vis.*, Marseille, France, 2008, pp. 479–491.

[16] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, 2006, pp. 2161–2168.

[17] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, USA, 2007, pp. 1–8.

[18] G. Griffin and P. Perona, "Learning and using taxonomies for fast visual categorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, 2008, pp. 1–8.

[19] A. Klaeser, M. Marszałek, and C. Schmid, "A spatio-temporal descriptor based on 3D-gradients," in *Proc. Brit. Mach. Vis. Conf.*, Leeds, U.K., 2008, pp. 99.1–99.10.

[20] J. Liu, S. Ali, and M. Shah, "Recognizing human actions using multiple features," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.

[21] X. Sun, M. Y. Chen, and A. Hauptmann, "Action recognition via local descriptors and holistic features," in *Proc. IEEE Int. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, 2009, pp. 58–65.

[22] M. F. Alcantara, T. P. Moreira, and H. Pedrini, "Real-time action recognition based on cumulative motion shapes," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Florence, Italy, May 2014, pp. 2917–2921.

[23] E. Yu and J. K. Aggarwal, "Human action recognition with extremities as semantic posture representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Miami, FL, USA, 2009, pp. 1–8.

[24] S. Ali, A. Basharat, and M. Shah, "Chaotic invariants for human action recognition," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Rio de Janeiro, Brazil, Oct. 2007, pp. 1–8.

[25] J. M. Chaquet, E. J. Carmona, and A. Fernández-Caballero, "A survey of video datasets for human action and activity recognition," *Comput. Vis. Image Underst.*, vol. 117, no. 6, pp. 633–659, Jun. 2013.

[26] E. Yu and J. K. Aggarwal, "Detection of fence climbing from monocular video," in *Proc. 18th Int. Conf. Pattern Recognit.*, vol. 1. Hong Kong, 2006, pp. 375–378.

[27] A. Telea and J. J. van Wijk, "An augmented fast marching method for computing skeletons and centerlines," in *Proc. Symp. Data Visual.*, Barcelona, Spain, 2002, pp. 251–259.

[28] C. F. Juang, C. M. Chang, J. R. Wu, and D. Lee, "Computer vision-based human body segmentation and posture estimation," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 39, no. 1, pp. 119–133, Jan. 2009.

[29] C.-C. Yu, Y.-N. Chen, H.-Y. Cheng, J.-N. Hwang, and K.-C. Fan, "Connectivity based human body modeling from monocular camera," *J. Inf. Sci. Eng.*, vol. 26, no. 2, pp. 363–377, 2010.

[30] W. Lao, J. Han, and P. H. de With, "Fast detection and modeling of human-body parts from monocular video," in *Proc. 5th Int. Conf. Articulated Motion Deformable Objects*, Heidelberg, Germany: Springer-Verlag, 2008, pp. 380–389.

[31] W. Lao, J. Han, and P. H. N. de With, "Flexible human behavior analysis framework for video surveillance applications," *Int. J. Digit. Multimedia. Broadcast.*, vol. 2010, Jan. 2010, Art. no. 920121.

[32] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: Real-time surveillance of people and their activities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 809–830, Aug. 2000.

[33] M. Bregonzio, S. Gong, and T. Xiang, "Action recognition with cascaded feature selection and classification," in *Proc. Int. Conf. Imag. Crime Detect. Prevent.*, London, U.K., 2009, pp. 1–6.

[34] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2247–2253, Dec. 2007.

[35] S. Singh, S. A. Velastin, and H. Ragheb, "MuHAVi: A multicamera human action video dataset for the evaluation of action recognition methods," in *Proc. 7th IEEE Int. Conf. Adv. Video Signal Based Surv.*, Boston, MA, USA, Sep. 2010, pp. 48–55.

[36] J. Bent, "Data-driven batch scheduling," Ph.D. dissertation, Dept. Comput. Sci., Univ. Wisconsin, Madison, WI, USA, May 2005.

[37] B. Arbab-Zavar, I. Bouchrika, J. N. Carter, and M. S. Nixon, "On supervised human activity analysis for structured environments," in *Proc. 6th Int. Symp. Vis. Comput.*, vol. 6455. Las Vegas, NV, USA, 2010, pp. 625–634.

[38] N. Hamilton, W. Weimar, and K. Luttgens, *Kinesiology Scientific Basis of Human Motion*. New York, NY, USA: McGraw-Hill, 2011.

[39] R. Easterby, K. Kroemer, and D. B. Chaffin, *Anthropometry and Biomechanics*. New York, NY, USA: Plenum Press, 2010.

[40] J. Hamill and K. M. Knutzen, *Biomechanical Basis of Human Movement*. Philadelphia, PA, USA: Lippincott Williams and Wilkins, Wolters Kluwer, 2009.

[41] A. A. Chaaraoui, P. Climent-Pérez, and F. Flórez-Revuelta, "Silhouette-based human action recognition using sequences of key poses," *Pattern Recognit. Lett.*, vol. 34, no. 15, pp. 1799–1807, 2013.

[42] A. A. Chaaraoui and F. Flórez-Revuelta, "A low-dimensional radial silhouette-based feature for fast human action recognition fusing multiple views," *Int. Scholarly Res. Notices*, vol. 2014, Jul. 2014, Art. no. 547069.

[43] A. Eweiwi, S. Cheema, C. Thurau, and C. Bauckhage, "Temporal key poses for human action recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Barcelona, Spain, Nov. 2011, pp. 1310–1317.

[44] F. Martinez-Contreras, C. Orrite-Urunuela, E. Herrero-Jaraba, H. Ragheb, and S. A. Velastin, "Recognizing human actions using silhouette-based HMM," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surv.*, Genoa, Italy, Sep. 2009, pp. 43–48.

Faisal Azhar (S'14) received the B.Sc. degree in biomedical engineering from Sir Syed University, Karachi, Pakistan, the M.S. degree in electrical engineering from the National University of Sciences and Technology, Karachi, and the Ph.D. degree in marker-less human body part detection, labeling and tracking for human activity recognition from the University of Warwick, Coventry, U.K.

He is currently a Research Associate with HP Labs, Bristol, U.K. His current research interest include computer vision, machine learning, and 3-D reconstruction.

Chang-Tsun Li (SM'97) received the B.Eng. degree in electrical engineering from National Defense University, Taoyuan, Taiwan, in 1987, the M.Sc. degree in computer science from U.S. Naval Postgraduate School, Monterey, CA, USA, in 1992, and the Ph.D. degree in computer science from the University of Warwick, Coventry, U.K., in 1998.

He is a Professor with the Department of Computer Science, University of Warwick, Coventry, U.K. His current research interests include multimedia forensics and security, biometrics, data mining, machine learning, data analytics, computer vision, image processing, pattern recognition, bioinformatics, and content-based image retrieval.

Prof. Li is currently an Associate Editor of the *EURASIP Journal of Image and Video Processing*, an Associate Editor-in-Chief of the *International Journal of Biometrics and Bioinformatics*, and an Associate Editor of *Human-Centric Computing and Information Sciences*. He was the Coordinator and the PI of the international joint project entitled Digital Image and Video Forensics funded through the Marie Curie Action under the EU's Seventh Framework Programme from 2010 to 2014. He is currently the Coordinator and the PI of the EU Horizon 2020 project, entitled Computer Vision Enabled Multimedia Forensics and People Identification (IDENTITY). The IDENTITY project has a consortium consisting of 16 institutions from 12 countries and will be running for four years, from 2016 to 2019.