Research Article

Human Activity Recognition Using Gaussian Mixture Hidden Conditional Random Fields

Muhammad Hameed Siddiqi,1 Madallah Alruwaili,1 Amjad Ali,2 Saad Alanazi,1 and Furkh Zeshan2

1College of Computer and Information Sciences, Jouf University, Sakaka, Saudi Arabia
2Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Lahore, Pakistan

Correspondence should be addressed to Muhammad Hameed Siddiqi; siddiqi1984@gmail.com

Received 12 May 2019; Accepted 10 July 2019; Published 18 August 2019

1. Introduction

In real-life environments, there are some fascinating applications in which the analysis of human activities plays a significant role. Some applications include human/object detection and recognition based on vision object analysis and processing areas such as tracking and detection [1, 2], computer engineering [3], physical sciences [4], health-related issues, natural sciences, and industrial academic areas [5]. Most of the authors [6–11] recognized the human activities in indoor environments based on different methodologies. However, in their respective systems, they used stable environment like fixed camera setting and prelighting setting, and most of the activities were performed by the instructions provided by the instructor. Similarly, the authors of [10, 12–14] proposed different methods to recognize the human daily activities in outdoor environments. However, in most of the used datasets, they used static background and this is one of the common drawbacks in their systems. Similarly, different sensors were utilized by the authors of [15–17] in order to classify indoor and outdoor human activities.

Moreover, in telemedicine and healthcare, human activity recognition (HAR) can be explained by helping physically disabled persons’ scenario. A paralyzed patient with half of the body critically disturbed by stroke is completely unable to walk and the one way to recover him is through daily exercises. Normally, the daily exercises (activities) are recommended by the doctors to the stroke patients for getting better improvements in their health. A human activity recognition (HAR) system can correctly train and identify the activities performed by the stroke patients, through which the doctors easily can monitor the improvement scale in the patients’ health.

There are four modules in a typical HAR system: pre-processing (segmentation), feature extraction, feature
selection, and recognition as shown in Figure 1. Most of the existing works [18–23] focused on feature extraction and selection; however, very limited works have been done for the recognition module. Some studies exploited conventional techniques [24–28]. Among them, HMM is one of the best candidates for the activity recognition; however, HMM is generative in nature and less precise than its matching part like HCRF model [29].

The inspiration behind the recognition stage is the lack of enhancement in the learning method. Therefore, we have made the following contribution:

(i) The existing HCRF model is inadequate by independence assumptions, which may reduce classification accuracy. Therefore, the first objective of this study is to propose a recognition model that presents a new algorithm to relax the assumption, allowing our model in order to use full-covariance distribution.

(ii) Another objective of the work is to prove that computation wise our method has very much lower complexity against the existing methods. In this method, our goal was to find some parameters to maximize the conditional probability of the training data at the training phase. Therefore, in our work, we utilize limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) method to search for the optimal point. However, instead of repeating the forward and backward algorithms to compute the gradients as others did [30], we run the forward and backward algorithms only when calculating the conditional probability, and then we reuse the result to compute the gradients. As a result, the computation time is significantly reduced.

(iii) A comprehensive set of experiments which yielded a weighted average classification rate 97% that is better improvement in the performance against the state-of-the-art methods.

The rest of the paper is organized as follows: Section 2 presents related works with their limitations. Section 3 provides the proposed recognition model with its advantages. Section 4 describes the experimental setup for the proposed model against four datasets. Based on the setup, a series of experiments are presented in Section 5. Finally, Section 6 describes conclusion with some future directions.

2. Related Works

In a typical HAR system, different types of latest segmentation methods were used in preprocessing module in order to extract the human body from the activity frame. This process helps to improve the performance of the activity recognition system. Therefore, in the literature, the authors of [31–36] utilized the latest methods to segment the human body from the video frames. Similarly, for the feature extraction, different latest methodologies have been employed which help the classifiers to accurately classify the human activities (as the workflow shown in Figure 1) [37–42]. They showed better performance on different datasets, and most of them achieved average accuracy between 70 and 90%.

Regarding recognition, the researchers have proposed diverse systems which exploit various classifiers such as Gaussian mixture model (GMM) [43, 44], artificial neural network (ANN) [45, 46], and support vector machine (SVMs) [47–50]. These classifiers were principally employed for frame-based classification. Contrarily, in many HAR systems [37, 51, 52], the eminent hidden Markov model (HMM) has extensively been utilized for sequence-based classification. In the case of frame-level features, HMMs are benefited over vector-based classifiers like SVM, GMM, and ANN in terms of effectively handling the sequential data. However, the Markovian property implied in the traditional HMM assumes that the current state is a function of the past state only. This causes the labels of two adjacent states in the observation sequence to hypothetically appear in succession. But in practical implementation, this assumption often does not meet satisfaction. Besides, the generative characteristic of HMM and independence presumptions between observations and states also limit its performance [29]. To get rid of these limitations, the maximum entropy Markov model (MEMM) had been proposed which comparatively performs better than HMM [53]. However, MEMM is associated with the well-known disadvantage termed as "label bias problem".
Two generalized models of MEMM known as conditional random fields (CRFs) [29] and HCRF [54] were developed to fix the shortcoming of “label bias problem” [29]. For learning the hidden structure of the sequential data, HCRF facilitates the effectiveness of CRF with hidden states. However, in both models, the per-state normalization is replaced with global normalization, permitting the weighted scores which in turn result in larger parameter spaces as compared to HMM and MEMM.

For example, the CRFs achieved in the HAR system having the observed frames from a video are represented by feature vector $U$, resultant label $V$, and unknown state label $K$.

Suppose, the problem image labeling is assumed by original labels $K$ with image features $U$ and parameter of the model is $\Lambda$, then the later probability (post $\Lambda$) is the vector factor, and $\Lambda$ having the observed frames from a video are represented by $\sum_{K} f(K, U, \Lambda)$ maximized by CRF is given as

$$
\text{post}(K \mid U; \Lambda) = \frac{e_{x} \cdot f(K, U, \Lambda)}{z(U, \Lambda)},
$$

(1)

where the normalization factor is

$$
\begin{align*}
    z(U, \Lambda) &= \sum_{K} e_{x} \cdot f(K, U, \Lambda). \\
    &= \sum_{K} e_{x} \cdot f(K', U, \Lambda).
\end{align*}
$$

Some issues in HCRF implementation are reviewed and analyzed in the following description. The later probability of CRF in (1) has been updated by the post $\text{post}(K \mid U; \Lambda)$ in a HCRF model that is the addition of exponentials of latent functions with all expected labels $L$ as given below

$$
\begin{align*}
    \text{post}(V \mid U; \Lambda) &= \frac{\sum_{L} e_{x} \cdot f(V, \Lambda, U)}{z(U, \Lambda)}, \\
    &= \sum_{V \in \Lambda} e_{x} \cdot f(V, \Lambda, U).
\end{align*}
$$

The above equations are used to warranty the sum to one rule of the conditional probability. $V'$ is the possible tag for the series of frames, and $\Lambda = \{l_{1}, l_{2}, \ldots, l_{T}\}$ is a series of hidden states $l_{t}$, $i = 1, 2, \ldots, T$, and equations (1) and (2) have constant values from 1 to Q (the number of states), $\Lambda$ is the vector factor, and $f(V, \Lambda, U)$ is a feature vector that will yield a decision which parameter will be educated by the model. Then, the feature vector concludes the addition of the existing HCRF model. For example, the underneath selections will create a Markov restraint HCRF with a Gaussian distribution at every state:

$$
\begin{align*}
    f_{t}^{\text{Prior}}(V, \Lambda, U) &= \delta(l_{t} = l) \cdot v, \\
    f_{t}^{\text{Transition}}(V, \Lambda, U) &= \sum_{t=1}^{T} \delta(l_{t-1} = l') \delta(l_{t} = l') \cdot v, \\
    f_{t}^{\text{Occurrence}}(V, \Lambda, U) &= \sum_{t=1}^{T} \delta(l_{t} = l) \cdot v,
\end{align*}
$$

(4)

where each $v \in V$ is the expected tag and every $u \in U$ is a predicted vector. The per-component square of the observation vector $v$ at state $t$ (i.e., $v_{t}$) is given as

$$
\begin{align*}
    f_{t}^{M1} &= \sum_{t=1}^{T} \delta(l_{t} = l) \cdot u, \\
    &= \sum_{t=1}^{T} \delta(l_{t} = l) \cdot u_l, \\
    f_{t}^{M2}(V, \Lambda, U) &= \sum_{t=1}^{T} \delta(l_{t} = l) \cdot u_l^2 v.
\end{align*}
$$

(5)

It can be seen that along with certain set of parameters $\Lambda$, the HCRF addition is similar to the hidden Markov model, for instance along with the abovementioned feature vector, if we choose

$$
\begin{align*}
    \text{Prior} \Lambda &= \log(b(l)) \\
    \text{Transition} \Lambda &= \log(C(l, f)), \\
    \text{Occurrence} \Lambda &= -\frac{1}{2} \left( \log \left( 2 \pi \sigma_{l}^{2} \right) + \frac{L^{2}_{l} \pi_{l}}{\sigma_{l}^{2}} \right),
\end{align*}
$$

(6)

(7)

(8)

(9)

(10)

where $b$ in (6) is an earlier dissemination of Gaussian HMM and $C$ in (7) is an evolution matrix; then conditional possibility numerator might be explained as

$$
\sum_{L} e_{x} \cdot f(V, \Lambda, U) = \sum_{L} b(l) \prod_{t=1}^{T} C(l_{t-1}, l_{t}) N(u_{l}, \mu_{l}, \sigma_{l}).
$$

(11)

In the above equation, $N$ represents Gaussian distribution. Equation (11) is the conditional probability of $U$, given $V$ is calculated along with a Gaussian HMM through equation (11) which has an earlier distribution $b$ with a conversion matrix $C$.

Moreover, the authors of [30] proposed a comprehensive form of the HCRF model to tackle composite scatterings utilizing a linear combination of Gaussian distribution functions, which is explained as

$$
\begin{align*}
    \text{post}(V \mid U; \Lambda) &= \frac{\sum_{L} b(l) \prod_{t=1}^{T} C(l_{t-1}, l_{t}) N(u_{l}, \mu_{l}, \sigma_{l})}{z(U, \Lambda)}.
\end{align*}
$$

(12)

In equation (12), $M$ indicates the number of components in Gaussian mixture.

Lots of works have been developed which showed better performance based on the usage of the abovementioned HCRF [55, 56]; however, most of them did not consider the limitations of the model. It is obvious from the aforementioned equations that the existing model employed diagonal (sloping)-covariance Gaussian distribution, which means that the variables (columns of $u_{l}$, $i = 1, 2, \ldots, N$) were presumed to be couples independent. On the other hand, equations (8)–(10) suggest that with a specific set of value, each state observation density will congregate to Gaussian procedure. Unluckily, there is no training method
3. Proposed Methodology

3.1. Feature Extraction. In our previous work, we utilized symlet wavelet [37] for extracting various features from the activity frames. There are number of reasons for using the symlet wavelet which produces relatively better classification results. These include its capability to extract the conspicuous information from the activity frames in terms of frequency and its support to the characteristics of the grayscale images like orthogonality, biorthogonality, and reverse biorthogonality. For a certain provision size, the symlet is characterized with the highest number of vanishing moments and has the least asymmetry.

3.2. Proposed Hidden Conditional Random Fields (HCRFs) Model. As described earlier, the current Gaussian mixture HCRF model does not have the capability of utilizing full-covariance distributions and also does not guarantee the conjunction of its factors to certain values upon which the conditional probability is demonstrated as a combination of the normal density functions.

To address these limitations, we explicitly involve a mixture of Gaussian distributions in the feature functions as illustrated in the following forms:

\[
\text{Prior} \quad f_1^\text{Prior} (V, T, U) = \delta(l_1 = l) \cdot \nu, \quad \forall \nu \in \mathcal{I},
\]

\[
\text{Transition} \quad f_{ll'}^\text{Transition} (V, T, U) = \sum_{i=1}^{T} \delta(l_{i-1} = l) \delta(l_i = l') \cdot \nu, \quad \forall \nu \in \mathcal{I},
\]

then,

\[
N(u_i^2, \mu_{l,m}, \Sigma_{l,m}) = \frac{1}{(2\pi)^{D/2} |\Sigma_{l,m}|^{1/2}} e^{-\frac{1}{2}(u_i^2 - \mu_{l,m})^T \Sigma_{l,m}^{-1} (u_i^2 - \mu_{l,m})},
\]

where \( N \) represents the number of density functions, Gamma "\( \Gamma \)" considers the appropriate information of the entire observations, \( D \) indicates the dimension of the observation, and \( \Gamma_{l,m}^{\text{Obs}} \) presents the partying weightiness for the \( m \)th constituent along with mean \( \mu_{l,m} \) and covariance matrix \( \Sigma_{l,m} \).

As indicated in equation (14), when we change some of the parameters such as \( \Gamma, \mu, \) and \( \Sigma \), then we may build a combination of the standard densities. The resultant conditional probability might be written as

\[
\text{post}(V | U; \Lambda, \Gamma, \mu, \Sigma) = \frac{\sum \phi(x)(P(T) + T(T) + O(T))}{z(U, \Lambda, \Gamma, \mu, \Sigma)},
\]

\[
P(T) = \sum_{l} \Lambda_{l}^\text{Prior} f_{l}^\text{Prior} (V, T, U),
\]

\[
T(T) = \sum_{ll'} \Lambda_{ll'}^\text{Transition} f_{ll'}^\text{Transition} (V, T, U),
\]

\[
O(T) = \sum_{l} \int_{l}^\text{Observation} (V, T, U),
\]

therefore,

\[
\text{post}(V | U; \Lambda, \Gamma, \mu, \Sigma) = \frac{\sum \phi(x)(P(T) + T(T) + O(T))}{z(U, \Lambda, \Gamma, \mu, \Sigma)},
\]

\[
\text{post}(V | U; \Lambda, \Gamma, \mu, \Sigma) = \frac{\sum \phi(x)(P(T) + T(T) + O(T))}{z(U, \Lambda, \Gamma, \mu, \Sigma)},
\]

\[
\text{post}(V | U; \Lambda, \Gamma, \mu, \Sigma) = \frac{\sum \phi(x)(P(T) + T(T) + O(T))}{z(U, \Lambda, \Gamma, \mu, \Sigma)},
\]

The forward and backward algorithms are used to calculate the conditional probability based on equations (19) and (20) that can be written as
\[ \alpha_z = \sum_{l_i, l_{i-1}, \ldots, l_1 \neq z} \exp \left( \Lambda_i^\text{Prior} + \sum_{t=1}^{T} \Lambda_{t-i}^\text{Transition} \right) + \log \left( \sum_{m=1}^{M} \Gamma_{l_m} N \cdot \left( u_t^z, \mu_{l,m}, \Sigma_{l,m} \right) \right) \]

\[ = \sum_{l_i} \alpha_{z-1}(l') \exp \left( \Lambda_{i-1}^\text{Transition} + \log \sum_{m=1}^{M} \Gamma_{l_m} N \cdot \left( u_t^z, \mu_{l,m}, \Sigma_{l,m} \right) \right), \tag{21} \]

\[ \beta_z(l) = \sum_{l_{i+1}, \ldots, l_T} \gamma_{l_{i+1}, \ldots, l_T} \exp \left( \Lambda_{i+1}^\text{Prior} + \sum_{t=1}^{T} \Lambda^\text{Transition}_{t} \right) + \log \left( \sum_{m=1}^{M} \Gamma_{l_m} N \cdot \left( u_t^z, \mu_{l,m}, \Sigma_{l,m} \right) \right) \]

\[ = \sum_{l_i} \beta_{z+1}(l') \exp \left( \Lambda_{i+1}^\text{Transition} + \log \left( \sum_{m=1}^{M} \Gamma_{l_m} N \left( u_t^z, \mu_{l,m}, \Sigma_{l,m} \right) \right) \right), \tag{22} \]

Score \((V \mid V; \Lambda, \Gamma, \mu, \Sigma) = \sum_{l} \alpha_{T}(l) = \sum_{l} \beta_{1}(l). \tag{23} \]

In the training data, to maximize the conditional probability, we initially focused on calculating the parameters \((\Lambda, \Gamma, \mu, \text{and } \Sigma)\). In the proposed approach, limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) has been implemented in order to search the optimum point. Unlike the other models [30], both the forward and backward algorithms are used to compute the conditional probability and the results were reused for finding the gradients. This makes the algorithm more significant in reducing the computation time.

At the observation level, we particularly incorporated the full-covariance matrix in the feature function as shown in (16). Equation (17) may be used for getting the normal distribution which is further elaborated in the following equations:

\[ \frac{d \text{Score}(V \mid V; \Lambda, \Gamma, \mu, \Sigma)}{d\Gamma_{l,m}^\text{Observation}} = \sum_{t} \frac{d g(V, \bar{V}, U)}{d\Gamma_{l,m}^\text{Observation}} \gamma_{l_{i+1}, \ldots, l_T} \exp \left( \Lambda_{i+1}^\text{Transition} \right) + \log \left( \sum_{m=1}^{M} \Gamma_{l_m} N \left( u_t^z, \mu_{l,m}, \Sigma_{l,m} \right) \right) \]

\[ = \sum_{t} \delta (l_t = l) \gamma_{l_{i+1}, \ldots, l_T} \exp \left( \Lambda_{i+1}^\text{Transition} \right) + \log \left( \sum_{m=1}^{M} \Gamma_{l_m} N \left( u_t^z, \mu_{l,m}, \Sigma_{l,m} \right) \right) \]

\[ = \sum_{t} \alpha_{t} \gamma_{l_{i+1}, \ldots, l_T} \exp \left( \Lambda_{i+1}^\text{Transition} \right) + \log \left( \sum_{m=1}^{M} \Gamma_{l_m} N \left( u_t^z, \mu_{l,m}, \Sigma_{l,m} \right) \right) \]
The \( d \) Score function is a gradient function for a Gaussian mixture weight variable. Here, a function \( V(t) \) can be determined as

\[
y(t) = \sum_l \beta(t, l),
\]

\( (27) \)

\[
d \text{Score}(V \mid U; \Lambda, \Gamma, \mu, \Sigma) \bigg|_{\text{d} \mu_{lm}} = \sum_{l=1}^{L} \sum_{m=1}^{M} \Gamma_{lm}^{(U)} \frac{dN(u_{t}, \mu_{lm}, \Sigma_{lm})}{d \mu_{lm}} \cdot \alpha(t, l)y(t + 1). \]

\( (28) \)

The \( d \) Score function is a gradient function for the Gaussian distribution mean:

\[
d \text{Score}(V \mid U; \Lambda, \Gamma, \mu, \Sigma) \bigg|_{\text{d} \Sigma_{lm}} = \sum_{l=1}^{L} \sum_{m=1}^{M} \Gamma_{lm}^{(U)} \frac{dN(u_{t}, \mu_{lm}, \Sigma_{lm})}{d \Sigma_{lm}} \cdot \alpha(t, l)y(t + 1). \]

\( (29) \)

The \( d \) Score function is a gradient function for the covariance of the Gaussian distributions.

Equations (24)–(27) presented above describe an analysis method algorithm for calculating values of gradients for a feature function, the mean of Gaussian distributions, and the covariance of the prior probability vector, the transition probability vector, and the observation probability vector obtained from the existing HCRF.

In our model, the recognition of a variety of real-time activities can be divided into two steps: a training step and an inference step. In the first step, data with known labels are inputted for recognizing the target as well as training the hidden conditional field model. In the inference step, the inputs to be actually estimated are ordered dependent on parameters determined in the training phase.

If the activity frame is acting as an input in the training step, then, in the preprocessing step, the applied distinctive lighting effects are decreased for detecting and extracting faces from the activity frames. At that point, the movable features are extracted from the various facial parts for creating the feature vector. After that, the feature vector obtained serves as an input to a full-covariance Gaussian-mixed hidden conditional random field model of the suggested recognition model.

As mentioned in the earlier discussion, a feature gradient is generally determined by LBFG approach in the training phase of the HCRF model. Nonetheless, in the current gradient calculation technique, a forward and backward iterative execution algorithm is iteratively called upon, which needs an exceptionally high computational time and thus leads to reduction in the computational speed. Another analysis approach has been formulated that reduces the invoking of the forward and backward iterative execution algorithm using five gradient functions determined by equations (24)–(28). Using
this analysis, the real-time computation can be carried out at a higher speed resulting in an enormous decrease in the computational time compared to a known analysis approach. The overall workflow of the proposed model is shown in Figure 2.

4. Model Validation

4.1. Datasets Used. In this work, we employed four open-source standard action datasets like Weizmann action datasets [57], KTH action dataset [58], UCF sports dataset [59], and IXMAS action dataset [60] for corroborating the proposed HCRF model performance. All the datasets are explained below.

4.1.1. Weizmann Action Dataset. This dataset consisted of 10 actions such as bending, running, walking, skipping, place jumping, side movement, jumping forward, two hand waving, and one hand waving that were performed by total 9 subjects. This dataset comprised of 90 video clips with average of 15 frames per clip where the frame size is 144 × 180.

4.1.2. KTH Action Dataset. KTH dataset employed for activity recognition comprised of 25 subjects who performed 6 activities like running, walking, boxing, jogging, hand clapping, and hand waving in four distinctive scenarios. Using a static camera, in the homogenous background, a total of 2391 sequences were taken with a frame size of 160 × 120.

4.1.3. UCF Sports Dataset. In this dataset, there were 182 videos which were evaluated by n-fold cross-validation rule. This dataset has been taken from different sports activities in broadcast television channels. Some of the videos had high intraclass similarities. This dataset was also collected using a static camera. This dataset covers 9 activities like running, diving, lifting, golf swinging, skating, kicking, walking,

### Table 1: Confusion matrix of the proposed recognition model using Weizmann action dataset (unit: %).

| Activities | Bend | Jack | Pjump | Run | Side | Skip | Walk | Wave 1 | Wave 2 |
|------------|------|------|-------|-----|------|------|------|--------|--------|
| Bend       | 98   | 0    | 1     | 0   | 0    | 1    | 0    | 0      | 0      |
| Jack       | 1    | 96   |       | 0   | 2    | 0    | 1    | 0      | 0      |
| Pjump      | 0    | 1    | 97    | 0   | 1    | 0    | 0    | 1      | 0      |
| Run        | 0    | 0    | 0     | 99  | 0    | 0    | 0    | 1      | 0      |
| Side       | 1    | 2    | 0     | 0   | 95   | 0    | 0    | 2      | 0      |
| Skip       | 0    | 0    | 0     | 0   | 0    | 100  | 0    | 0      | 0      |
| Walk       | 1    | 0    | 0     | 1   | 0    | 0    | 96   | 0      | 1      |
| Wave 1     | 0    | 1    | 1     | 0   | 0    | 0    | 98   | 0      | 0      |
| Wave 2     | 0    | 1    | 0     | 0   | 1    | 1    | 1    | 1      | 95     |
| Average    |      |      |       |     |      |      |      | 97.11  |        |

### Table 2: Confusion matrix of the proposed recognition model using KTH action dataset (unit: %).

| Activities | Walking | Jogging | Running | Boxing | Hand-wave | Handclap |
|------------|---------|---------|---------|--------|-----------|----------|
| Walking    | 100     | 0       | 0       | 0      | 0         | 0        |
| Jogging    | 0       | 98      | 1       | 1      | 0         | 0        |
| Running    | 2       | 1       | 95      | 1      | 0         | 0        |
| Boxing     | 0       | 2       | 1       | 97     | 0         | 0        |
| Hand-wave  | 1       | 0       | 1       | 0      | 98        | 0        |
| Handclap   | 0       | 0       | 1       | 0      | 0         | 99       |
| Average    |         |         |         |         | 97.83     |          |

### Table 3: Confusion matrix of the proposed recognition model using UCF sports dataset (unit: %).

| Activities | Diving | GS | Kicking | Lifting | HBR | Run | Skating | BS | Walk |
|------------|--------|----|---------|---------|-----|-----|---------|----|------|
| Diving     | 95     | 1  | 2       | 1       | 0   | 1   | 0       | 0  | 0    |
| GS         | 1      | 94 | 0       | 0       | 2   | 1   | 1       | 1  | 0    |
| Kicking    | 0      | 2  | 98      | 0       | 0   | 0   | 0       | 0  | 0    |
| Lifting    | 1      | 1  | 1       | 94      | 1   | 0   | 0       | 1  | 1    |
| HBR        | 0      | 0  | 2       | 0       | 96  | 1   | 0       | 1  | 0    |
| Running    | 0      | 3  | 0       | 0       | 0   | 97  | 0       | 0  | 0    |
| Skating    | 1      | 0  | 1       | 1       | 0   | 1   | 95      | 0  | 1    |
| BS         | 0      | 0  | 0       | 1       | 0   | 1   | 0       | 97 | 1    |
| Walking    | 0      | 0  | 0       | 0       | 0   | 0   | 0       | 100|      |
| Average    |         |    |         |         |     |     |         | 96.22|      |

Gs: golf swinging, HBR: horseback riding, and BS: baseball swinging.
Table 4: Confusion matrix of the proposed recognition model using IXMAS action dataset (unit: %).

| Activities | CA | SD | GU | TA | Walk | Wave | Punch | Kick |
|------------|----|----|----|----|------|------|-------|------|
| CA         | 97 | 0  | 0  | 1  | 2    | 0    | 0     | 0    |
| SD         | 0  | 99 | 1  | 0  | 0    | 0    | 0     | 0    |
| GU         | 1  | 2  | 94 | 3  | 0    | 0    | 0     | 0    |
| TA         | 0  | 0  | 1  | 95 | 2    | 1    | 1     | 0    |
| Walk       | 0  | 1  | 0  | 98 | 0    | 0    | 0     | 0    |
| Wave       | 0  | 0  | 2  | 0  | 1    | 97   | 0     | 0    |
| Punch      | 0  | 1  | 0  | 1  | 0    | 2    | 96    | 0    |
| Kick       | 0  | 0  | 0  | 0  | 1    | 0    | 0     | 99   |
| Average    |    |    |    |    | 96.88|      |       |      |

CA: cross arm, SD: sit down, GU: get up, and TA: turn around.

Table 5: Classification results of the proposed system on Weizmann action dataset (A) using ANN, (B) using SVM, (C) using HMM, and (D) using existing HCRF [30], while removing the proposed HCRF model (unit: %).

| Activities | Bend | Jack | Pjjump | Run | Side | Skip | Walk | Wave 1 | Wave 2 |
|------------|------|------|--------|-----|------|------|------|--------|--------|
| (A)        |      |      |        |     |      |      |      |        |        |
| Bend       | 70   | 4    | 5      | 3   | 3    | 2    | 5    | 5      | 3      |
| Jack       | 4    | 68   | 3      | 6   | 7    | 2    | 3    | 3      | 4      |
| Pjjump     | 2    | 4    | 75     | 6   | 3    | 2    | 2    | 4      | 4      |
| Run        | 4    | 2    | 3      | 72  | 6    | 2    | 5    | 3      | 3      |
| Side       | 5    | 3    | 5      | 4   | 65   | 6    | 4    | 6      | 2      |
| Skip       | 4    | 6    | 4      | 3   | 5    | 67   | 3    | 2      | 6      |
| Walk       | 4    | 2    | 4      | 7   | 3    | 4    | 70   | 3      | 3      |
| Wave 1     | 2    | 1    | 3      | 3   | 4    | 5    | 7    | 71     | 4      |
| Wave 2     | 2    | 5    | 3      | 6   | 4    | 4    | 3    | 5      | 68     |
| Average    |      |      |        |     |      |      |      | 69.55  |        |
| (B)        |      |      |        |     |      |      |      |        |        |
| Bend       | 69   | 3    | 4      | 4   | 6    | 4    | 2    | 3      | 5      |
| Jack       | 2    | 72   | 3      | 4   | 3    | 5    | 4    | 5      | 4      |
| Pjjump     | 1    | 4    | 75     | 2   | 4    | 5    | 4    | 2      | 3      |
| Run        | 2    | 4    | 3      | 78  | 2    | 2    | 4    | 2      | 3      |
| Side       | 2    | 4    | 5      | 3   | 70   | 4    | 3    | 5      | 4      |
| Skip       | 2    | 1    | 3      | 2   | 4    | 80   | 3    | 3      | 2      |
| Walk       | 2    | 0    | 3      | 4   | 3    | 2    | 82   | 1      | 3      |
| Wave 1     | 2    | 2    | 3      | 4   | 3    | 2    | 3    | 77     | 4      |
| Wave 2     | 2    | 1    | 2      | 3   | 1    | 3    | 4    | 83     |        |
| Average    |      |      |        |     |      |      |      | 76.22  |        |
| (C)        |      |      |        |     |      |      |      |        |        |
| Bend       | 82   | 3    | 0      | 2   | 2    | 3    | 1    | 5      | 2      |
| Jack       | 3    | 80   | 1      | 2   | 3    | 2    | 3    | 4      | 2      |
| Pjjump     | 3    | 4    | 85     | 3   | 0    | 0    | 1    | 2      | 2      |
| Run        | 5    | 4    | 2      | 79  | 0    | 2    | 1    | 3      | 4      |
| Side       | 0    | 1    | 5      | 4   | 81   | 3    | 1    | 2      | 3      |
| Skip       | 3    | 1    | 2      | 2   | 3    | 88   | 0    | 0      | 1      |
| Walk       | 0    | 2    | 3      | 2   | 1    | 2    | 83   | 3      | 4      |
| Wave 1     | 1    | 3    | 2      | 2   | 4    | 2    | 3    | 78     | 5      |
| Wave 2     | 1    | 2    | 2      | 2   | 2    | 3    | 1    | 0      | 87     |
| Average    |      |      |        |     |      |      |      | 82.56  |        |
| (D)        |      |      |        |     |      |      |      |        |        |
| Bend       | 80   | 2    | 3      | 1   | 4    | 0    | 5    | 2      | 3      |
| Jack       | 1    | 88   | 0      | 2   | 0    | 3    | 2    | 3      | 1      |
| Pjjump     | 0    | 2    | 90     | 1   | 0    | 3    | 0    | 2      | 2      |
| Run        | 2    | 1    | 2      | 85  | 2    | 3    | 0    | 0      | 5      |
| Side       | 4    | 1    | 2      | 3   | 80   | 4    | 1    | 2      | 3      |
| Skip       | 1    | 4    | 0      | 5   | 1    | 84   | 0    | 3      | 2      |
| Walk       | 2    | 1    | 0      | 1   | 2    | 89   | 2    | 3      |        |
| Wave 1     | 3    | 0    | 1      | 2   | 0    | 2    | 0    | 91     | 1      |
| Wave 2     | 4    | 1    | 3      | 0   | 2    | 3    | 0    | 2      | 85     |
| Average    |      |      |        |     |      |      |      | 85.78  |        |
horseback riding, and baseball swinging. Each frame has a size of 720 × 480.

4.1.4. IXMAS Action Dataset. IXMAS (INRIA Xmas motion acquisition sequences) dataset comprised of 13 activity classes which were performed by 11 actors, each 3 times. Every actor opted a free orientation as well as position. IXMAS dataset has provided annotated silhouettes for each person. For our experiments, we have selected only 8 action classes like walk, cross arms, punch, turn around, sit down, wave, get up, and kick. IXMAS dataset is a multiview dataset for a view-invariant human activity recognition where each frame has a size of 390 × 291. This dataset has a major occlusion and that may cause misclassification; therefore, we utilized global histogram equalization [61] in order to resolve the occlusion issue.

4.2. Setup. For a comprehensive validation, we carried out the following set of experiments executed using Matlab.

(i) The first experiment was conducted on each dataset separately in order to show the performance of the proposed model. In this experiment, we employed 10-fold cross-validation rule, which means that data from 9 subjects were utilized for training data, while the data from one subject was picked as a testing data. The procedure was reiterated for 10 times provided each subject data is utilized for both training and testing.

(ii) The second experiment was conducted in the absence of the proposed recognition model on all the four datasets that will show the importance of the developed model. For this purpose, we used the existing eminent classifiers like SVM, ANN, HMM, and existing HCRF [30] as a recognition model rather than utilizing the proposed HCRF model.

(iii) The third experiment was conducted to show the performance of the proposed approach against the state-of-the-art methods.

(iv) In the last experiment, the computational complexity of the proposed HCRF model was compared with forward/backward algorithms.

5. Results and Discussion

5.1. First Experiment. As described before, this experiment validates the performance of the proposed recognition model

| Activities         | Walking | Jogging | Running | Boxing | Hand-wave | Handclap |
|--------------------|---------|---------|---------|---------|-----------|----------|
| (A) Walking        | 79      | 5       | 6       | 4       | 3         | 3        |
| Jogging            | 3       | 81      | 5       | 3       | 4         | 4        |
| Running            | 6       | 4       | 77      | 5       | 5         | 3        |
| Boxing             | 6       | 7       | 6       | 69      | 5         | 7        |
| Hand-wave          | 4       | 7       | 5       | 5       | 73        | 6        |
| Handclap           | 4       | 6       | 5       | 4       | 6         | 75       |
| Average            |         |         |         |         | 75.66     |          |
| (B) Walking        | 82      | 2       | 3       | 5       | 4         | 4        |
| Jogging            | 3       | 86      | 2       | 3       | 2         | 4        |
| Running            | 5       | 3       | 80      | 4       | 5         | 3        |
| Boxing             | 5       | 3       | 3       | 79      | 4         | 6        |
| Hand-wave          | 1       | 4       | 3       | 3       | 89        | 0        |
| Handclap           | 3       | 5       | 2       | 4       | 3         | 83       |
| Average            |         |         |         |         | 83.17     |          |
| (C) Walking        | 86      | 3       | 2       | 4       | 2         | 3        |
| Jogging            | 0       | 88      | 3       | 2       | 4         | 3        |
| Running            | 0       | 3       | 90      | 0       | 4         | 3        |
| Boxing             | 3       | 0       | 4       | 92      | 1         | 0        |
| Hand-wave          | 1       | 3       | 2       | 2       | 91        | 1        |
| Handclap           | 1       | 3       | 4       | 1       | 2         | 89       |
| Average            |         |         |         |         | 89.33     |          |
| (D) Walking        | 90      | 3       | 0       | 3       | 4         | 0        |
| Jogging            | 2       | 88      | 2       | 3       | 3         | 2        |
| Running            | 4       | 2       | 92      | 0       | 0         | 2        |
| Boxing             | 1       | 3       | 2       | 91      | 3         | 0        |
| Hand-wave          | 0       | 1       | 3       | 2       | 93        | 1        |
| Handclap           | 1       | 3       | 2       | 4       | 3         | 87       |
| Average            |         |         |         |         | 90.17     |          |
on an individual dataset. The overall results are shown in Tables 1 (using Weizmann dataset), 2 (using KTH dataset), 3 (using UCF sports dataset), and 4 (using IXMAS), respectively.

As observed from Tables 1–4, the proposed recognition model constantly obtained higher recognition rates on individual dataset. This result shows the robustness of the proposed model which means that the model not only showed better performance on one dataset but also showed better performance across multiple spontaneous datasets.

5.2. Second Experiment. As described before, the second experiment was conducted in the absence of the proposed recognition model, to show the importance of the proposed model using all the four datasets. For this purpose, we used the existing eminent classifiers like SVM, ANN, HMM, and existing HCRF [30] as a recognition model rather than utilizing the proposed HCRF model.

Tables 5–8 show that when the proposed HCRF model was substituted with ANN, SVM, HMM, and existing HCRF [30], the system failed to accomplish higher recognition rates. The better performance of the proposed HCRF model is visualized in Tables 1–4, which show that the proposed HCRF model effectively fix the drawbacks of HMM and existing HCRF that has been extensively utilized for sequential HAR.

### Table 7: Classification results of the proposed system on UCF sports dataset (A) using ANN, (B) using SVM, (C) using HMM, and (D) using existing HCRF [30], while removing the proposed HCRF model (unit: %).

| Activities | Diving | GS | Kicking | Lifting | HBR | Run | Skating | BS | Walking | Average |
|------------|--------|----|---------|---------|-----|-----|---------|----|---------|---------|
| (A) Diving | 68     | 4  | 2       | 5       | 6   | 6   | 4       | 3  | 2       | 67.78   |
| GS         | 2      | 71 | 2       | 4       | 5   | 4   | 6       | 3  | 3       | 75.00   |
| Kicking    | 3      | 4  | 70      | 3       | 5   | 4   | 2       | 3  | 6       | 70.50   |
| Lifting    | 5      | 4  | 3       | 65      | 5   | 6   | 4       | 6  | 2       | 82.67   |
| HBR        | 3      | 4  | 6       | 3       | 66  | 4   | 5       | 5  | 4       | 86.89   |
| Running    | 3      | 3  | 5       | 4       | 6   | 64  | 6       | 4  | 5       |         |
| Skating    | 2      | 5  | 4       | 5       | 3   | 4   | 69      | 3  | 5       |         |
| BS         | 4      | 2  | 5       | 3       | 4   | 6   | 5       | 67 | 4       |         |
| Walking    | 5      | 4  | 2       | 3       | 4   | 3   | 6       | 3  | 70      |         |
| Average    | 67.78  |     |         |         |     |     |         |     |         |         |
| (B) Diving | 71     | 4  | 2       | 3       | 5   | 6   | 3       | 2  | 4       | 75.00   |
| GS         | 3      | 77 | 2       | 4       | 3   | 2   | 5       | 2  | 2       |         |
| Kicking    | 4      | 2  | 74      | 4       | 5   | 3   | 2       | 3  | 3       |         |
| Lifting    | 5      | 6  | 3       | 69      | 4   | 3   | 5       | 3  | 2       |         |
| HBR        | 2      | 3  | 3       | 2       | 80  | 2   | 4       | 2  | 2       |         |
| Running    | 2      | 3  | 2       | 2       | 5   | 75  | 6       | 2  | 3       |         |
| Skating    | 2      | 1  | 2       | 3       | 4   | 4   | 78      | 2  | 4       |         |
| BS         | 3      | 4  | 6       | 3       | 4   | 2   | 3       | 70 | 5       |         |
| Walking    | 4      | 1  | 2       | 4       | 2   | 3   | 0       | 81 |         |         |
| Average    | 75.00  |     |         |         |     |     |         |     |         |         |
| (C) Diving | 79     | 3  | 2       | 2       | 3   | 4   | 3       | 2  | 2       | 82.67   |
| GS         | 0      | 83 | 2       | 4       | 3   | 2   | 1       | 3  | 2       |         |
| Kicking    | 1      | 2  | 85      | 1       | 3   | 3   | 2       | 3  | 0       |         |
| Lifting    | 3      | 0  | 2       | 82      | 3   | 2   | 4       | 2  | 2       |         |
| HBR        | 0      | 2  | 2       | 80      | 0   | 5   | 3       | 4  | 4       |         |
| Running    | 1      | 2  | 1       | 3       | 4   | 84  | 2       | 1  | 2       |         |
| Skating    | 2      | 0  | 3       | 4       | 0   | 1   | 86      | 3  | 1       |         |
| BS         | 1      | 1  | 1       | 2       | 0   | 3   | 0       | 88 | 4       |         |
| Walking    | 1      | 2  | 4       | 2       | 5   | 2   | 4       | 3  | 77      |         |
| Average    | 82.67  |     |         |         |     |     |         |     |         |         |
| (D) Diving | 90     | 3  | 0       | 1       | 0   | 2   | 2       | 1  | 1       | 86.89   |
| GS         | 3      | 84 | 2       | 1       | 3   | 1   | 3       | 2  | 1       |         |
| Kicking    | 3      | 4  | 85      | 0       | 0   | 2   | 3       | 1  | 2       |         |
| Lifting    | 1      | 2  | 1       | 89      | 1   | 1   | 1       | 2  | 2       |         |
| HBR        | 0      | 2  | 1       | 0       | 91  | 2   | 3       | 0  | 1       |         |
| Running    | 2      | 3  | 1       | 2       | 3   | 80  | 4       | 2  | 3       |         |
| Skating    | 2      | 4  | 1       | 2       | 3   | 0   | 84      | 4  | 0       |         |
| BS         | 2      | 1  | 1       | 2       | 1   | 0   | 3       | 88 | 2       |         |
| Walking    | 0      | 2  | 1       | 1       | 0   | 1   | 4       | 0  | 91      |         |
| Average    | 86.89  |     |         |         |     |     |         |     |         |         |

GS: golf swinging, HBR: horseback riding, BS: baseball swinging.
5.3. Third Experiment. In this experiment, a comparative analysis was made between the state-of-the-art methods and the proposed model. All of these approaches were implemented by the instructions provided in their particular articles. A 10-fold cross-validation rule was employed on each dataset as explained in Section 4. The average classification results of the existing methods along with the proposed method across different datasets are summarized in Table 9.

Table 8: Classification results of the proposed system on IXMAS action dataset (A) using ANN, (B) using SVM, (C) using HMM, and (D) using existing HCRF [30], while removing the proposed HCRF model (unit: %).

| Activities | CA | SD | GU | TA | Walk | Wave | Punch | Kick |
|------------|----|----|----|----|------|------|-------|------|
| (A)        |    |    |    |    |      |      |       |      |
| CA         | 65 | 5  | 7  | 6  | 5    | 4    | 3     | 5    |
| SD         | 5  | 72 | 4  | 3  | 5    | 4    | 3     | 4    |
| GU         | 4  | 3  | 75 | 5  | 3    | 2    | 4     | 4    |
| TA         | 6  | 7  | 4  | 68 | 3    | 5    | 4     | 3    |
| Walk       | 3  | 4  | 5  | 4  | 70   | 6    | 4     | 4    |
| Wave       | 4  | 6  | 5  | 3  | 4    | 71   | 3     | 4    |
| Punch      | 3  | 5  | 5  | 6  | 7    | 3    | 67    | 4    |
| Kick       | 4  | 5  | 4  | 6  | 5    | 4    | 3     | 69   |
| Average    |    |    |    |    |      |      | 69.62 |      |
| (B)        |    |    |    |    |      |      |       |      |
| CA         | 77 | 3  | 4  | 2  | 2    | 3    | 5     | 4    |
| SD         | 3  | 79 | 3  | 2  | 4    | 5    | 2     | 2    |
| GU         | 5  | 6  | 69 | 3  | 4    | 5    | 4     | 4    |
| TA         | 2  | 3  | 2  | 80 | 4    | 4    | 3     | 2    |
| Walk       | 3  | 5  | 4  | 2  | 71   | 5    | 6     | 4    |
| Wave       | 2  | 6  | 3  | 5  | 4    | 73   | 4     | 3    |
| Punch      | 1  | 5  | 3  | 4  | 1    | 3    | 81    | 2    |
| Kick       | 3  | 6  | 7  | 4  | 3    | 5    | 4     | 68   |
| Average    |    |    |    |    |      |      | 74.75 |      |
| (C)        |    |    |    |    |      |      |       |      |
| CA         | 79 | 3  | 4  | 1  | 2    | 4    | 3     | 4    |
| SD         | 1  | 84 | 3  | 2  | 3    | 1    | 4     | 2    |
| GU         | 0  | 1  | 88 | 1  | 2    | 3    | 2     | 3    |
| TA         | 5  | 2  | 3  | 79 | 2    | 3    | 2     | 4    |
| Walk       | 1  | 0  | 3  | 1  | 90   | 2    | 3     | 0    |
| Wave       | 2  | 3  | 1  | 0  | 3    | 86   | 2     | 3    |
| Punch      | 1  | 0  | 2  | 3  | 0    | 4    | 89    | 1    |
| Kick       | 3  | 2  | 4  | 1  | 0    | 2    | 4     | 84   |
| Average    |    |    |    |    |      |      | 84.77 |      |
| (D)        |    |    |    |    |      |      |       |      |
| CA         | 90 | 1  | 2  | 0  | 3    | 4    | 0     | 0    |
| SD         | 3  | 85 | 2  | 1  | 3    | 3    | 2     | 1    |
| GU         | 0  | 1  | 91 | 1  | 0    | 2    | 3     | 2    |
| TA         | 1  | 3  | 2  | 87 | 1    | 2    | 2     | 2    |
| Walk       | 1  | 0  | 3  | 3  | 89   | 2    | 1     | 3    |
| Wave       | 0  | 2  | 1  | 0  | 2    | 90   | 3     | 1    |
| Punch      | 1  | 4  | 2  | 1  | 2    | 84   | 3     | 3    |
| Kick       | 1  | 2  | 4  | 1  | 3    | 2    | 4     | 83   |
| Average    |    |    |    |    |      |      | 87.38 |      |

CA: cross arm, SD: sit down, GU: get up, TA: turn around.

Table 9: Weighted average recognition rates of the proposed method with the existing state-of-the-art methods (unit: %).

| State-of-the-art works | Average classification rates | Standard deviation |
|------------------------|-----------------------------|-------------------|
| GMM                    | 63.3                        | ±2.7              |
| SVM                    | 67.5                        | ±4.4              |
| HMM                    | 82.8                        | ±3.8              |
| Embedded HMM           | 85.9                        | ±1.9              |
| [62]                   | 92.1                        | ±3.2              |
| [63]                   | 84.3                        | ±4.9              |
| [18]                   | 93.6                        | ±2.7              |
| [19]                   | 93.0                        | ±1.6              |
| [22]                   | 92.7                        | ±2.5              |
| [64]                   | 80.1                        | ±3.2              |
| Proposed method        | 97.2                        | ±2.8              |

Figure 3: An illustration of gradient computational time (equation (30)) of the previous forward and backward algorithms and the proposed HCRF model. (a) $Q = 1 - 5$, $M = 5$, $T = 90$ and (b), $Q = 5$, $M = 1 - 5$, $T = 90$. 

5.3. Third Experiment. In this experiment, a comparative analysis was made between the state-of-the-art methods and the proposed model. All of these approaches were implemented by the instructions provided in their particular articles. A 10-fold cross-validation rule was employed on each dataset as explained in Section 4. The average classification results of the existing methods along with the proposed method across different datasets are summarized in Table 9.
It is obvious from Table 9 that the proposed method showed a significant performance against the existing state-of-the-art methods. Therefore, the proposed method accurately and robustly recognizes the human activities using different video data.

5.4. Fourth Experiment. In this experiment, we have presented the computational complexity that is also one of the contributions in this paper. The implementations of the previous HCRF are available in literature, which calculate the gradients by reiterating the forward and backward techniques, while the proposed HCRF model executes them once only and cashes the outcomes for the later use. From (21) and (22), it is clear that the forward or backward technique has a complexity of $O(TQ^2M)$, where $T$ represents the input sequence length, $Q$ represents the number of states, and $M$ indicates the number of mixtures. The proposed HCRF model, however, requires a full complexity of $O(TM)$ to calculate gradients as can be seen from (22)–(29).

Figure 3 shows a comparison of the execution time when the gradients are computed by the forward (or backward) algorithm and by our proposed method. The computational time is calculated by running Matlab R2013a with the specification of Intel® Pentium® Core™ i7-6700 (3.4 GHz) with a RAM capacity of 16 GB.

6. Conclusion

In healthcare and telemedicine, the human activity recognition (HAR) can be best explained by helping physically disabled persons’ scenario. A paralyzed patient with half of the body critically attacked by paralysis is completely unable to perform their daily exercises. The doctors recommend specific activities to get better improvement in their health. So, for this purpose, the doctors need a human activity recognition (HAR) system through which they can monitor the patients’ daily routines (activities) on a regular basis.

The accuracy of most of the HAR systems depends upon the recognition modules. For feature extraction and selection modules, we used some of the existing well-known methods, while for the recognition module, we proposed the usage of HCRF model which is capable of approximating a complex distribution using a mixture of Gaussian density functions. The proposed model was assessed against four publicly available standard action datasets. From our experiments, it is obvious that the proposed full-covariance Gaussian density function showed a significant improvement in accuracy than the existing state-of-the-art methods. Furthermore, we also proved that such improvement is significant from statistical point of view by showing value ≤0.2 of the comparison. Similarly, the complexity analysis points out that the proposed computational method strongly decreases the execution time for the hidden conditional random field model.

The ultimate goal of this study is to deploy the proposed model on smartphones. Currently, the proposed model is using full-covariance matrix; however, this might be time-consuming, especially when using on smartphones. Using a lightweight classifier such as K-nearest neighbor (K-NN) could be one possible solution. But K-NN is very much sensitive to environmental factor (like noise). Therefore, in future, we will try to investigate further research to reduce the time and sustain the same recognition rate when employing on smartphones in real environment.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare there are no conflicts of interest.

Acknowledgments

This study was supported by the Jouf University, Sakaka, Aljouf, Kingdom of Saudi Arabia, under the registration no. 39/791.

References

[1] Z. Chang, J. Cao, and Y. Zhang, “A novel image segmentation approach for wood plate surface defect classification through convex optimization,” *Journal of Forestry Research*, vol. 29, no. 6, pp. 1789–1795, 2018.
[2] Z. S. Abdallah, M. M. Gaber, B. Srinivasan, and S. Krishnaswamy, “Adaptive mobile activity recognition system with evolving data streams,” *Neurocomputing*, vol. 150, pp. 304–317, 2015.
[3] M. S. Bakli, M. A. Sakr, and T. H. A. Soliman, “A spatio-temporal algebra in Hadoop for moving objects,” *Geo-Spatial Information Science*, vol. 21, no. 2, pp. 102–114, 2018.
[4] W. Zhao, L. Yan, and Y. Zhang, “Geometric-constrained multi-view image matching method based on semi-global optimization,” *Geo-Spatial Information Science*, vol. 21, no. 2, pp. 115–126, 2018.
[5] H. Gao, X. Zhang, J. Zhao, and D. Li, “Technology of intelligent driving radar perception based on driving brain,” *CAA Transactions on Intelligence Technology*, vol. 2, no. 3, pp. 93–100, 2017.
[6] Y. Wang, X. Jiang, R. Cao, and X. Wang, “Robust indoor human activity recognition using wireless signals,” *Sensors*, vol. 15, no. 7, pp. 17195–17208, 2015.
[7] J. Wang, Z. Liu, Y. Wu, and J. Yuan, “Mining actionlet ensemble for action recognition with depth cameras,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1290–1297, IEEE, Providence, RI, USA, June 2012.
[8] S. Karungaru, M. Daikoku, and K. Terada, “Multi cameras based indoors human action recognition using fuzzy rules,” *Journal of Pattern Recognition Research*, vol. 10, no. 1, pp. 61–74, 2015.
[9] S. Shukri, L. M. Kamarudin, and M. H. F. Rahman, "Device-free localization for human activity monitoring.,” in *Intelligent Video Surveillance*, IntechOpen, London, UK, 2018.
[10] L. Fiore, D. Fehr, R. Bodor, A. Drenner, G. Somasundaram, and N. Papanikolopoulos, “Multi-camera human activity monitoring,” *Journal of Intelligent and Robotic Systems*, vol. 52, no. 1, pp. 5–43, 2008.
[11] C. Zhang and Y. Tian, “RGB-D camera-based daily living activity recognition,” Journal of Computer Vision and Image Processing, vol. 2, no. 4, pp. 1–7, 2012.

[12] M. Leo, T. D’Orazio, and P. Spagnolo, “Human activity recognition for automatic visual surveillance of wide areas,” in Proceedings of the ACM 2nd International Workshop on Video Surveillance and Sensor Networks, pp. 124–130, ACM, New York, NY, USA, October 2004.

[13] C.-B. Jin, S. Li, and H. Kim, “Real-time action detection in video surveillance using sub-action descriptor with multiconn,” 2017, https://arxiv.org/abs/1710.03383.

[14] W. Niu, J. Long, D. Han, and Y.-F. Wang, “Human activity detection and recognition for video surveillance,” in Proceedings of the 2004 IEEE International Conference on Multimedia and Expo (ICME) (IEEE Cat. No. 04TH8763), vol. 1, pp. 719–722, IEEE, Taipei, Taiwan, June 2004.

[15] Q. Li, M. Zheng, F. Li, J. Wang, Y.-A. Geng, and H. Jiang, “Retinal image segmentation using double-scale non-linear thresholding on vessel support regions,” CAAI Transactions on Intelligence Technology, vol. 2, no. 3, pp. 109–115, 2017.

[16] L. M. G. Fonseca, L. M. Namikawa, and E. F. Castejon, “Digital image processing in remote sensing,” in Proceedings of the 2009 Tutorials of the XXII Brazilian Symposium on Computer Graphics and Image Processing, pp. 59–71, IEEE, Rio de Janeiro, Brazil, October 2009.

[17] K. Buys, C. Cagniart, A. Baksheev, T. De Laet, J. De Schutter, and C. Pantofaru, “An adaptable system for RGB-D based human body detection and pose estimation,” Journal of Visual Communication and Image Representation, vol. 25, no. 1, pp. 39–52, 2014.

[18] M. Sharif, M. A. Khan, T. Akram, M. Y. Javed, T. Saba, and A. Rehman, “A framework of human detection and action recognition based on uniform segmentation and combination of Euclidean distance and joint entropy-based features selection,” EURASIP Journal on Image and Video Processing, vol. 2017, no. 1, p. 89, 2017.

[19] K.-P. Chou, M. Prasad, D. Wu et al., “Robust feature-based automated multi-view human action recognition system,” IEEE Access, vol. 6, pp. 15283–15296, 2018.

[20] M. Ullah, H. Ullah, and I. M. Alseadon, “Human action recognition in videos using stable features,” Signal and Image Processing: An International Journal, vol. 8, no. 6, pp. 1–10, 2017.

[21] M. Sharif, M. A. Khan, F. Zahid, J. H. Shah, and T. Akram, “Human action recognition: a framework of statistical weighted segmentation and rank correlation-based selection,” Pattern Analysis and Applications, pp. 1–14, 2019.

[22] J. Zang, L. Wang, Z. Liu, Q. Zhang, G. Hua, and N. Zheng, “Attention-based temporal weighted convolutional neural network for action recognition,” in Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations, pp. 97–108, Springer, Rhodes, Greece, May 2018.

[23] V.-D. Hoang, “Multiple classifier-based spatiotemporal features for living activity prediction,” Journal of Information and Telecommunication, vol. 1, no. 1, pp. 100–112, 2017.

[24] S. Dharmalingam and A. Palanisamy, “Vector space based augmented structural kinematic feature descriptor for human activity recognition in videos,” ETRI Journal, vol. 40, no. 4, pp. 499–510, 2018.

[25] D. Liu, Y. Yan, M.-L. Shyu, G. Zhao, and M. Chen, “Spatio-temporal analysis for human action detection and recognition in uncontrolled environments,” International Journal of Multimedia Data Engineering and Management, vol. 6, no. 1, pp. 1–18, 2015.

[26] C. A. Ronao and S.-B. Cho, “Human activity recognition with smartphone sensors using deep learning neural networks,” Expert Systems with Applications, vol. 59, pp. 235–244, 2016.

[27] A. Tharwat, H. Malhdi, M. Elhoseny, and A. E. Hassanien, “Recognizing human activity in mobile crowdsensing environment using optimized k-NN algorithm,” Expert Systems with Applications, vol. 107, pp. 32–44, 2018.

[28] A. Jalal, Y.-H. Kim, Y.-J. Kim, S. Kamal, and D. Kim, “Robust human activity recognition from depth video using spatio-temporal multi-fused features,” Pattern Recognition, vol. 61, pp. 295–308, 2017.

[29] J. LaFerry, A. McCallum, and F. C. Pereira, “Conditional random fields: probabilistic models for segmenting and labeling sequence data,” in Proceedings of the 18th International Conference on Machine Learning (ICML-2001), pp. 282–289, Williamstown, MA, USA, June-July 2001.

[30] A. Gunawardana, M. Mahajan, A. Acero, and J. C. Platt, “Hidden conditional random fields for phone classification,” in Proceedings of the INTERSPEECH, vol. 2, pp. 1117–1120, Citeseer, Lisbon, Portugal, September 2005.

[31] L. Yang, Q. Song, Z. Wang, and M. Jiang, “Parsing r-cnn for instance-level human analysis,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 364–373, Long Beach, CA, USA, June 2019.

[32] X. Liang, Y. Wei, L. Lin et al., “Learning to segment human by watching youtube,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 7, pp. 1462–1468, 2016.

[33] M. M. Requena, “Human segmentation with convolutional neural networks,” Thesis, University of Alicante, Alicante, Spain, 2018.

[34] H. Gao, S. Chen, and Z. Zhang, “Parts semantic segmentation aware representation learning for person re-identification,” Applied Sciences, vol. 9, no. 6, p. 1239, 2019.

[35] C. Peng, S.-L. Lo, J. Huang, and A. C. Tsoi, “Human action segmentation based on a streaming uniform entropy slice method,” IEEE Access, vol. 6, pp. 16958–16971, 2018.

[36] M. H. Siddiqi, A. M. Khan, and S.-W. Lee, “Active contours level set based still human body segmentation from depth images for video-based activity recognition,” KSII Transactions on Internet and Information Systems (TIIS), vol. 7, no. 11, pp. 2839–2852, 2013.

[37] M. Siddiqi, R. Ali, M. Rana, E.-K. Hong, E. Kim, and S. Lee, “Video-based human activity recognition using multilevel wavelet decomposition and stepwise linear discriminant analysis,” Sensors, vol. 14, no. 4, pp. 6370–6392, 2014.

[38] S. Althloothi, M. H. Mahoor, X. Zhang, and R. M. Voyles, “Human activity recognition using multi-features and multiple kernel learning,” Pattern Recognition, vol. 47, no. 5, pp. 1800–1812, 2014.

[39] A. Jalal, S. Kamal, and D. Kim, “A depth video-based human action recognition and activity recognition using multi-features and embedded hidden markov models for health care monitoring systems,” International Journal of Interactive Multimedia and Artificial Intelligence, vol. 4, no. 4, p. 54, 2017.

[40] T. Dobhal, V. Shitole, G. Thomas, and G. Navada, “Human activity recognition using binary motion image and deep learning,” Procedia Computer Science, vol. 58, pp. 178–185, 2015.

[41] S. Zhang, Z. Wei, J. Nie, L. Huang, S. Wang, and Z. Li, “A review on human activity recognition using vision-based method,” Journal of Healthcare Engineering, vol. 2017, Article ID 3090343, 31 pages, 2017.
[42] Y. Yang, B. Zhang, L. Yang, C. Chen, and W. Yang, "Action recognition using completed local binary patterns and multiple-class boosting classifier," in Proceedings of the 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), IEEE, pp. 336–340, Kuala Lumpur, Malaysia, November 2015.

[43] W. Lin, M. T. Sun, R. Pooivandran, and Z. Zhang, "Human activity recognition for video surveillance," in Proceedings of the 2008 IEEE International Symposium on Circuits and Systems, pp. 2737–2740, IEEE, Seattle, WA, USA, May 2008.

[44] H. Permuter, J. Francos, and I. Jermyn, "A study of Gaussian mixture models of color and texture features for image classification and segmentation," Pattern Recognition, vol. 39, no. 4, pp. 695–706, 2006.

[45] M. Fiaz and B. Ijaz, "Vision based human activity tracking using artificial neural networks," in Proceedings of the 2010 International Conference on Intelligent and Advanced Systems, pp. 1–5, IEEE, Kuala Lumpur, Malaysia, June 2010.

[46] H. Foroughi, A. Naseri, A. Saberi, and H. S. Yazdi, "An eigenspace-based approach for human fall detection using integrated time motion image and neural network," in Proceedings of the 2008 9th International Conference on Signal Processing, pp. 1499–1503, IEEE, Beijing, China, October 2008.

[47] A. K. B. Sonali, "Human action recognition using support vector machine and k-nearest neighbor," International Journal of Engineering and Technical Research, vol. 3, no. 4, pp. 423–428, 2015.

[48] D. Das and S. Sabaria, "Human gait analysis and recognition using support vector machines," International Journal of Computer Science and Information Technology, vol. 6, no. 5, 2004.

[49] K. G. M Chathuramali and R. Rodrigo, "Faster human activity recognition with SVM," in Proceedings of the International Conference on Advances in ICT for Emerging Regions (ICTer2012), pp. 197–203, IEEE, Colombo, Sri Lanka, December 2012.

[50] M. Z. Uddin, D.-H. Kim, J. T. Kim, and T.-S. Kim, "An indoor human activity recognition system for smart home using local binary pattern features with hidden markov models," Indoor and Built Environment, vol. 22, no. 1, pp. 289–298, 2013.

[51] M. Z. Uddin, J. Lee, and T.-S. Kim, "Independent shape component-based human activity recognition via hidden Markov model," Applied Intelligence, vol. 33, no. 2, pp. 193–206, 2010.

[52] S. Kumar and M. Hebert, "Discriminative Fields for Modeling Spatial Dependencies in Natural Images," in Proceedings of the NIPS, Vancouver, Canada, December 2003.

[53] A. Quattoni, S. Wang, L.-P. Morency, M. Collins, and T. Darrell, "Hidden conditional random fields," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 10, pp. 1848–1852, 2007.

[54] M. Mahajan, A. Gunawardana, and A. Acero, "Training algorithms for hidden conditional random fields," in Proceedings of the 2006 IEEE International Conference on Acoustics Speed and Signal Processing, vol. 1, IEEE, Toulouse, France, May 2006.

[55] S. Reiter, B. Schuller, and G. Rigoll, "Hidden conditional random fields for meeting segmentation," in Proceedings of the Multimedia and Expo, 2007 IEEE International Conference, pp. 639–642, IEEE, Beijing, China, July 2007.
