REAL TIME HUMAN ACTIVITY IDENTIFICATION SYSTEM TO AID ELDERLY CARE

Udaykiran.K1, H.Chidananda2, Ramya BR3

1B.E. 8th semester, Department of computer science and engineering, RYMEC, Karnataka, India.
2Department of computer science and engineering, RYMEC, Karnataka, India.
3B.E. 8th semester, Department of computer science and engineering, RYMEC, Karnataka, India

Abstract-Human activity recognition has been a tremendous growth in the last decade playing a major role in the field of pervasive computing. This emerging popularity can be attributed to its myriad of real life applications primarily dealing with human-centric problems like healthcare and eldercare. Many research attempts with data mining and machine learning techniques have been undergoing to accurately recognize human activities for e-health systems. Aim of this project is to recognize the activity under care using a cost effective camera rather than multiple costly sensors by using adaptive background subtraction for removal of the background, optical flow model for feature extraction and iterative querying heuristic (IQH) and multiple instance learning model for recognizing the activities and to send a sms using GSM module to the caretakers in real time.

Keywords: Human Activity, Iterative Querying Heuristic (IQH) optimization algorithm, healthcare, eldercare, Multiple Instance Learning Algorithm.

1. INTRODUCTION

Human activity identification based on pc vision is of great scientific and practical importance. Its various important theoretic values and wide potential applications, like real-time video surveillance system, interpretation of sport events, and human pc interactions. In recent years, human activity identification has become a search hot spot. Tremendous amount of researches had been made in the field of human activity identification from video sequence. There are many human activity identification strategies are already present. Common methodology to identify human activity aims at both still images and video sequences.

The human activity recognition continues to be an immature technology and a challenging problem in pc vision to attain robust human activity identification from image sequences due to occlusion, human postures, illumination conditions, dynamic backgrounds, camera movements. With moving target and cameras, non-stationary background, few vision algorithms will categorize and recognize human activity well. It’s essential for intelligent and natural human pc interaction to recognize human actions automatically.

In this paper, we tend to present a unique technique for human activity recognition from video sequences. This approach includes two steps of feature extraction and human activity identification. In the extracting feature, global silhouette features of human are extracted by using adaptive background subtraction algorithm; local optical flow features are extracted by using optical flow model. Then fusing global silhouette feature vector and local optical flow feature vector form a hybrid feature vector.

Finally an optimized Multiple Instance Learning algorithmic program is used for human activity recognition. Additionally, in order to improve the recognition accuracy, an Iterative Querying Heuristic (IQH) optimization algorithmic program is employed to train the Multiple Instance Learning model.

Figure1: Flow diagram of human activity reorganization based on fusion features extraction.

Given unlabeled human activity video sequence, it’s our goal to automatically learn totally different human activity categories in the data and apply the Multiple Instance Learning model for human actions categorization and recognition in the new video sequences. Our approach is illustrated in Figure 1. The rest of this paper is organized as follows. In Section 3, we tend to describe human activity features presentation based on adaptive background subtraction and optical flow models. In Section 4, we tend to explain about the algorithmic rule for human activity identification using an
optimized Multiple Instance Learning algorithmic program. Section 5 shows experimental results, and also the conclusions are given in the final section.

2. Existing System

Vision based strategies are often used with none restrictions upon the user. Hand tracking is a significant section for gesture recognition, within which the signer’s or user’s hands should be detected and localized in image frames. Wherever detection needs correct lightening if not the gesture can’t be captured accurately.

3. Proposed System

The method of identifying a specific activity with the help of camera, the proper preprocessing of their provided info and therefore the learning/reasoning using this info. If choosing of the sensors and the data processing strategies are wrongly performed, the entire activity detection method could fail, resulting in the consequent failure of the entire application. The target of this project is to classify the most activities considered in smart home situation that are with respective to elderly people’s independent living, as well as their characterization and formalized context presentation and is to assist researchers and developers in these lower-level technical aspects that are nevertheless fundamental for the success of the entire application.

3.1 Human Activity Features Illustration

Recognizing human activities from video sequences is both a difficult task and an interesting experimenting area. Generally, two vital queries are concerned in human activity recognition. One way is to represent expeditiously human activity, that is the key for a human activity recognition methodology, and also the other way is to model reference movements, which expeditiously cope with variations at spatial and temporal scales among similar motion categories.

Extracting helpful motion info from raw video information is crucial for human activity recognition. The selection of the motion features affects the results of the human activity recognition methodology directly. Several factors usually influence the only feature differently, such as appearance of human body, atmosphere, and video camera. Therefore the accuracy of action recognition is restricted.

After completely considering the benefits and downsides of various features, we tend to propose a hybrid feature methodology on the idea of studying the representation and recognition of human activity that is fusing global silhouette feature and local optical flow feature.

3.1.1. Global Silhouette Feature Illustration.

Human silhouettes in single frame image can be used to describe the data of which overall shape of a human body movement changes. Adaptive background subtraction could be a quite excellent methodology for global silhouette feature extraction. We tend to use it to work out the general area of the movement and the human body shadow. Assume that every actions that are performed before the static background. We tend to use the adaptive background subtraction to determine the motion area and extract the human body silhouette.

The step for adaptive background subtraction algorithmic rule is as follows.

**Step 1** (initialize the background model). Initially Taking $T$ background image continuously, then through the image, a single Gaussian distribution is created to set the initial background statistical model $H(u_i,v_i^2)$. Forming the Gaussian distribution needs the mean and standard deviation, which might be calculated using

$$u_i = \frac{1}{T} \sum_{t=1}^{T} u_{it}$$

$$v_i^2 = \frac{1}{T} \sum_{t=1}^{T} (u_{it} - u_i)^2$$

Where $u_{it}$ is color value of the point $i$ in the $t^{th}$ image.

**Step 2** (extracting foreground area). Assumptive that color value of the point $i$ is in current image, the image binarization is expressed as

$$B_i = \begin{cases} 1, & \text{if} \ r_i - u_i > 3\sigma_i \\ 0, & \text{else}, \end{cases}$$

Where points of all signs “1” are foreground area and points of all signs “0” are background region.

**Step 3** (update background model). Assume that $(t)$ is color expectations of the point $i$ at time $t$, $V^2(t)$ is color variance of the point $i$ at time $t$, and $r(t)$ is color value of the point $i$ in time $t$ which collect pictures. At time $t + 1$, there is

$$u_i(t + 1) = \begin{cases} au_i(t) + (1 - \alpha) r_i(t), & \text{flag}_i = 0 \\ u_{i(t)}, & \text{flag}_i = 1, \end{cases}$$

$$v_i^2(t + 1) = \begin{cases} \alpha v_i^2(t) + (1 - \alpha) (r_i(t) - u_i(t))^2, & \text{flag}_i = 0 \\ v_i^2(t), & \text{flag}_i = 1 \end{cases}$$

Where is used to control the background update rate, $0 < \alpha < 1$. Each point in the binary image is marked by flag $i$, if value of the flag $i$ is “1”, that is foreground, and flag $i$ is “0”, that is the background. Thus, the area marked as “0” constitutes the background region and the area marked as “1” constitutes the foreground area.
So we will get the foreground of the binary image. Figure 2 shows an example of adaptive background subtraction to extract the moving human body silhouette. Silhouette features have the subsequent benefits. Silhouette features can be used to describe the shape of human movement information simply and visually. Silhouette features are easy to be extracted.

Binary silhouette figure isn’t sensitive to texture and color of the foreground image. Suppose that a video sequence contains T frame image I—namely, $V = \{I_1, I_2, \ldots, I_T\}$—and $S$ is the sequence of motion silhouette corresponding to video $V$—namely, $S = \{s_1, s_2, \ldots, s_T\}$. We tend to use the contour vector to explain the overall human silhouette and shape info.

The step is as follows.

**Step 1.** Use canny operator for edge profile of every frame silhouette and calculate the coordinates of the edge profile. Such human body contour can be set with $n$ points, namely, $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

**Step 2.** Calculate the center of mass regarding human body contour, which is expressed as

$$
(x_c, y_c) = \left( \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n}}, \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n}} \right)
$$

(4)

Where $(x_c, y_c)$ is the center of mass and $(x_i, y_i)$ is edge points of the contour. $n_i$ is the number of edge points in the $t$th image.

**Step 3.** Calculate the distance from the center to the edge points, that is expressed as

$$
d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}, \quad i = 1,2,3,\ldots,n_t
$$

(5)

Where $d_i$ is the distance from the center of mass to edge points, which the $i$th edge points correspond to.

**Step 4.** (Normalize contour vector). We tend to use 2-norm for contour vector normalization process. 2-norm is the most popular general approaches for computing approximations of the condition number. It’s the incremental condition estimation (ICE) of a triangular matrix and might be viewed as a special case of the framework. The strategy is of course connected to adaptative techniques and contains clearly visible links to matrix decompositions. After the normalization process of contour vector, we perform interval such as resampling. We obtain the fixed point $N$, where $N$ is set as 180.

**3.1.2. Local Optical Flow Feature illustration.**

Silhouette image extraction of incorrect information might lead to the contour vector characteristics which cannot accurately express movement characteristic. Optical flow characteristics are effective and accurate motion info representation in video sequences. Optical flow-based human actions representation has attracted much attention. The human activity could be a dynamic event; it can be represented by the motion info of the human. The local optical flow features (optical flow vector) are estimated by optical flow model.

Figure 2: The processing pipeline of the adaptative background subtraction: (top) the original picture, (middle) the object extraction background subtraction and, and (bottom) the body silhouettes.

Firstly, we tend to calculate the optical flow vector of human actions, namely $\mu = (\mu_L, \mu_H)$ at every frame using optical flow equation, that is expressed as

$$
I_{t+1} \mu_L + I_{t+1} \mu_H + I_t = 0,
$$

(6)

Where $L = \frac{\partial I}{\partial x}, H = \frac{\partial I}{\partial y}$, $L_t = \frac{dx}{dt}, H_t = \frac{dy}{dt}$, and $\mu_L, \mu_H$ are the horizontal and vertical velocities in pixel $(x, y)$.

We can acquire $\mu = (\mu_L, \mu_H)$ by minimizing the objective function:

$$
L=\int_{T} \left[ \sqrt{2} \langle \nabla \mu \rangle^2 + (\nabla \mu + I_t)^2 \right] dxdy
$$

(7)

We use the iterative algorithm to calculate the optical flow; optical flow is decomposed into longitudinal and transverse two components, namely, the longitudinal optical flow and transversal optical flow, that is expressed as

$$
\mu^k_{L+1} = \mu^k_{L} - \frac{I_t |I_{t+1} \mu_L^k + I_{t+1} \mu_H^k|^2}{I_{t+1}^2 + I_{t+1}^2},
$$

$$
\mu^k_{H+1} = \mu^k_{H} - \frac{I_t |I_{t+1} \mu_L^k + I_{t+1} \mu_H^k|^2}{I_{t+1}^2 + I_{t+1}^2}
$$

(8)
Where \( k \) is the number of iterations, \( \mu_{L,k} \) and \( \mu_{H,k} \) are the average velocity of the neighborhood of point \((x,y)\), and \( \mu_{L}^{0} = \mu_{H}^{0} = 0 \) is initial value of velocity. After standardization, optical flow diagrams are divided into \( 2 \times 2 \) sub frames, where they are set S1, S2, S3, S4, respectively. The dimension of sub frame is set \( 60 \times 60 \), and the centers of sub frame are set G1s shown in Figure 3, where \( i = 1, 2, 3, 4 \). Then each sub frame is split into 18 subareas in the centers of sub frame and degree of each central angle is \( 20^\circ \) as shown in Figure 4, where they’re set \( S_{i}^{j} \) and where \( j = 1, 2, \ldots, 18 \).

**Figure 3:** \( 2 \times 2 \) sub frame figure of optical flow diagrams.

![Figure 3](image)

**Figure 4:** 18 subareas of each sub frame in the centers of sub frame.

Thus each frame image of optical flow graph is split into 72 subareas.

We calculate the sum of the longitudinal optical flow \( O_{L,i} \) and the sum of the transverse optical flow \( O_{H,i} \) at subarea:

\[
O_{L,i} = \sum_{k=1}^{K} \mu_{L,k+1}^{i} \in S_{i}^{j},
\]

\[
O_{H,i} = \sum_{k=1}^{K} \mu_{H,k+1}^{i} \in S_{i}^{j},
\]

where \( S_{i}^{j} \) is the subarea, \( i = 1, 2, 3, 4 \) and \( j = 1, 2, \ldots, 18 \), and \( k \) is the number of the longitudinal optical flows or the transverse optical flows. Optical flow information of the image can be represented by the sum of longitudinal optical flow and transverse optical flow in 72 subareas, which is expressed as

\[
O_{L} = [O_{L,1,1}, O_{L,1,18}, O_{L,4,1}, \ldots, O_{L,4,18}],
\]

\[
O_{H} = [O_{H,1,1}, O_{H,1,18}, O_{H,4,1}, \ldots, O_{H,4,18}].
\]

Where \( O_{L} \) is the sum of longitudinal optical flow in 72 subareas and \( O_{H} \) is the sum of transversal optical flow in 72 subareas. We calculate local optical flow vector \( O_T \):

\[
O_T = [O_L, O_H],
\]

Where we use 2-norm for \( O_T \) normalization process and get the local optical flow vector of the current frame image.

**3.1.3. Human Activity Hybrid Feature presentation.**

In order to enhance the accuracy of human activity recognition, fusing the contour vector and local optical flow vector forms hybrid feature vector, which is expressed as

\[
F_T = [O_T, D_T],
\]

Where \( F_T \) is the hybrid feature vector of each frame image, \( O_T \) is local optical flow vector, and \( D_T \) is contour vector. Human actions are presented by the hybrid feature vectors of contour vector and local optical flow vector. As a result human actions may be regarded as motion, the local optical flow features will describe human actions effectively. Additionally, silhouette features will describe the shape of human movement information simply and visually. Thus, the hybrid features are shown to perform reliably with noisy image sequences and are applied in various tasks, such as action classification and action recognition.

**4. Human Activity Identification**

After characterizing human activities, there are several strategies to identify human activities. Because human actions recognition is innately a Multiple Instance Learning problem, we use the Multiple Instance Learning algorithmic rule to learn and recognize human actions. The Multiple Instance Learning model has been applied to numerous pc visions, such as object recognition, action recognition, and human detection. So as to extend recognition potency while not compromising with the accuracy, an Iterative Querying Heuristic(IQH) optimization algorithmic rule is employed to train the Multiple Instance Learning model, and later the improved Multiple Instance Learning framework is used to learn a unified classifier rather than of individual classifiers for all categories.

**4.1. The Multiple Instance Learning Rule Analysis.**

The Multiple Instance Learning is one among the foremost efficient machine learning algorithms nowadays. This idea
is known as an Integrated Segmentation and Recognition (ISR), and its key idea is to produce a special means in constituting training samples. Training samples are in "bags," they’re not singletons, and all of the samples in a bag share a label. Every bag contains a large variety of instances. Samples are organized into positive bags or negative bags of instances, where a minimum of one instance is positive in a positive bag, however all instances are negative in a negative bag. In the Multiple Instance Learning, it should learn at the same time the samples in a negative bag, which is expressed as $P^m$.

The step for recognizing human actions based on the Multiple Instance Learning is as follows.

**Step 1.** Given dataset \(\{X_i, C_i\}_{i=1}^N\), where \(X_i = \{x_{i1}, x_{i2}, \ldots, x_{in}\}\) is on behalf of training bags, that contains at least one positive sample in a positive bag, \(C_i = \max_j(C_{ij})\) represents the score of the sample, and \(C_{ik} \in \{0, 1\}\), where \(C_i = \{C_{i1}, C_{i2}, \ldots, C_{in}\}\). \(N\) is the number of all weak classifiers.

**Step 2.** Update all weak classifiers \(k\) with the info \(\{x_{ij}, C_i\}\).

**Step 3.** Initialize all strong classifiers := 0 for all \(i, j\), where strong classifiers are composed of all the weak classifier.

**Step 4.** For \(k = 1\) to \(K\), for \(m = 1\) to \(N\).

In the \(i^{th}\) bag, when the \(j^{th}\) sample is positive, the probability is calculated as follows:

\[
P_i^m = \sigma(H_{ij} + h_m(x_{ij})),
\]

(13)

Where \(P_i^m = p(C_i|x_{ij}) = 1/(1 + \exp(-C_{ij}))\).

In the positive bag, the probability is calculated as follows:

\[
P_i^m = 1 - \prod_j(1 - P_i^m),
\]

(14)

Where \(P_i^m = P(C_i|X_i)\).

The probability that is allotted to a group of training bags is expressed as

\[
L^m = \sum(C_i \log(p^m) + (1 - C_i)\log(1 - P_i^m))
\]

End for.

Find the atmost of \(m^*\) from \(N\) and obtain the current optimal weak classifier as follows:

\[
m^* = \arg max_{m}(L^m).
\]

(16)

The strong classifier is produced by the \(m^*\) as follows:

\[
h_k(x) \leftarrow h_{m^*}(x),
\]

\[
H_{ij} = H_{ij} + h_k(x).
\]

(17)

End for.

**Step 5.** \(K\) weak classifiers constitute a strong classifier, which is expressed as

\[
H(x) = \sum_k h_k(x).
\]

(18)

Where \(h_k\) could be a weak classifier, which can create binary predictions by using sign \((H(x))\). In the Multiple Instance Learning, samples come into positive bags and negative bags of instances. Every instance \(x_{ij}\) is indexed by two indices, where \(i\) stands for the bag and \(j\) stands for the instance in the bag. All instances in a bag share a bag label \(C_i\). The weight of every sample consists of the weight of bags and the weight of samples in the bag, where the amount of the samples can be interpreted as a likelihood ratio. \(P^m_{ij}\) is the probability of positive instances in the bags, so the weight of samples is \(P^m\). We calculate \(w_{ij} = \partial \log L^m/\partial y_{ij}\) and obtain the weight of the bags \(w_i\).

### 4.2. Optimizing the Multiple Instance Learning Algorithmic rules for Human Activity identification.

The Multiple Instance Learning algorithmic rules present a general paradigm for a more relaxed form of supervised learning. In the Multiple Instance Learning, the learner gets unordered sets of occurrences or baggage rather than receiving example or label pairs, and labels are provided for every bag instead for each occurrence, where a positive baggage contains a minimum of one positive occurrence. In the initial training stages, training and evaluating have a direct impact on both the features and the appropriate thresholds selected, and it's the key to a fast and effective classifier. The samples have high score in positive bagage. The final classifier labels these samples to be positive. The remaining samples have a less score in the positive bagage, which are assigned a low weight. The final classifier classifies these samples as negative samples. Setting the detection threshold and training a complete classifier, we tend to acquire the required false positive rates and false negative rates. Retrain the initial weak classifier to get a zero false negative rate on the positive samples. Repeat the process to train the second classifier and yield a zero false negative rate on the remaining samples.

During the inference stage, given a testing image, we are able to treat every aspect in the Multiple Instance Learning model as a single class of human activity. Human activity identification requires a huge amount of training information; therefore it'll lead in long training time. In this paper, we adopt an Iterative Querying Heuristic (IQRH) algorithm to train the Multiple Instance Learning model.

The main step is as follows.

**Input:** Training baggage \(\{x_1, x_2, \ldots, M\}\), labels \(\{y_1, \ldots, y_m\}\), and parameters \(\tau, w, \text{ and } p\), where \(T\) is iterations times, \(p\) is instances per iteration, and \(w\) controls how many new instances, are considered in every iteration.

**Step 1.** Initialize \(H_{ij} = 0\), \(h_k\) being any classifier in \(H(x)\).

**Step 2.** For \(\tau = 1, \ldots, \beta\).

Find the atmost of \(m^*\) from \(N\) and obtain the current optimal weak classifier as follows:

\[
H(x) = \sum_k h_k(x).
\]

(18)
5.1. Experiment on Weizmann Data Set.

The correct recognition rate
\[
\frac{\text{the number of times of correct recognition}}{\text{total number of the samples}} \times 100%. \tag{19}
\]

In Table 1 we can see that our method can correctly recognize most of human actions. The recognition rate is as high as 100% for “bending,” “jumping-jack,” “jumping-in-place-on-two-legs,” “galloping-sideways,” “waving-two-hands,” and “waving-one-hand.” Our method achieves 98.5% average recognition rate. A few mistakes were confusions between “jumping-forward-on-two-legs” and “jumping-in-place-on-two-legs” because these two kinds of actions were similar to each other.

Table 1: The results of human activity on Weizmann database.

<table>
<thead>
<tr>
<th>The action Categories</th>
<th>Total number of samples</th>
<th>The number of the times of correct recognition</th>
<th>The correct recognition rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bending</td>
<td>200</td>
<td>200</td>
<td>100.0</td>
</tr>
<tr>
<td>Running</td>
<td>200</td>
<td>190</td>
<td>95.0</td>
</tr>
<tr>
<td>Walking</td>
<td>200</td>
<td>199</td>
<td>99.5</td>
</tr>
<tr>
<td>Jack</td>
<td>196</td>
<td>196</td>
<td>100.0</td>
</tr>
<tr>
<td>Jump</td>
<td>195</td>
<td>180</td>
<td>92.3</td>
</tr>
<tr>
<td>Pjump</td>
<td>192</td>
<td>192</td>
<td>100.0</td>
</tr>
<tr>
<td>Sideways</td>
<td>186</td>
<td>186</td>
<td>100.0</td>
</tr>
<tr>
<td>Wave2</td>
<td>192</td>
<td>192</td>
<td>100.0</td>
</tr>
<tr>
<td>Wave1</td>
<td>180</td>
<td>180</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Conclusion
In this paper, we present a new method for human activity identification in video sequences, which focuses on two key problems: extracting more useful and discriminating features and improving the accuracy of classifier. The main contribution can be concluded as follows. In feature extraction and representation, we extracted global silhouette features using adaptive background subtraction method, and optical flow model was used for extracting motion features. Later on fusing these two types of features formed a hybrid feature vector. In activity modeling and recognition, we proposed the improved Multiple Instance Learning algorithm for human activity identification using Iterative Querying Heuristic (IQH) algorithmic rule, so that the recognition efficiency can be increased without compromising accuracy. Experiments were performed on Weizmann human activity data set. Experiments evaluated the proposed method. Experimental results revealed that our proposed method performed better than previous ones. Our algorithmic rule can also recognize multiple actions in complex motion sequences.

References


