Human Activity Recognition Using Gait Pattern

Jay Prakash Gupta, Systems Engineer, Infosys Limited, Bangalore, Karnataka, India
Nishant Singh, CEA Department, Poornima Group of Colleges, Jaipur, Rajasthan, India
Pushkar Dixit, IT Department, Dr. M.P.S. Group of Institutions, Agra, Uttar Pradesh, India
Vijay Bhaskar Semwal, Senior System Engineer, Siemens Information Systems Limited, Gurgaon, India
Shiv Ram Dubey, GLA University, Mathura, India

ABSTRACT

Vision-based human activity recognition is the process of labelling image sequences with action labels. Accurate systems for this problem are applied in areas such as visual surveillance, human computer interaction and video retrieval. The challenges are due to variations in motion, recording settings and gait differences. Here the authors propose an approach to recognize the human activities through gait. Activity recognition through Gait is the process of identifying an activity by the manner in which they walk. The identification of human activities in a video, such as a person is walking, running, jumping, jogging etc are important activities in video surveillance. The authors contribute the use of Model based approach for activity recognition with the help of movement of legs only. Experimental results suggest that their method is able to recognize the human activities with a good accuracy rate and robust to shadows present in the videos.

Keywords: Activity Recognition, Feature Extraction, Gait Pattern, Human Computer Interaction, Video Surveillance

INTRODUCTION

Nowadays, huge number of images and videos are developed and uploaded on a daily basis in different fields and for different purposes for example news, sport, entertainment, and education. The continued rapid growth in digital visualization makes it increasingly difficult to find, organize, access, and maintain user’s visual information. Video has become a pervasive media type over the last decade. The rapid progress in imaging sensor technology, faster data transmission, larger data storage and increasing computational power, all contribute to the ubiquitous availability of this media type. The fast growing number of video media has led to a significantly increased interest in automatic video analysis in recent years.

The goal of automatic video analysis is to use computer algorithms to automatically extract information from unstructured data such as video frames and generate structured...
description of objects and events that are present in the scene. Among many objects under consideration, humans are of special significance because they play a major role in most activities of interest in daily life. Therefore, being able to recognize basic human actions in an indispensable component towards this goal and has many important applications. For example, detection of unusual actions such as jumping, running can provide timely alarm for enhanced security (e.g. in a video surveillance environment) and safety (e.g. in a life-critical environment such as a patient monitoring system).

The capability of understanding human gestures offers users a new means of human interaction and can bring brand new gaming experience (e.g. using body pose to control the character in the game) to video players. Human activity recognition is also useful in video content indexing which makes searching in large volume of video data more accessible and efficient.

In this paper, we use the concept of Gait for human activity recognition. The definition of Gait is defined as: “A particular way or manner of moving on foot”. Using gait as a biometric is a relatively new area of study, within the realms of computer vision. It has been receiving growing interest within the computer vision community and a number of gait metrics have been developed. Early psychological Gait studies by Murray (1967) suggest that gait is a unique personal characteristic, with cadence and cyclic in nature.

We use the term Gait recognition to signify the identification of an individual from a video sequence of the subject walking. This does not mean that Gait is limited to walking, it can also be applied to running or any means of movement on foot. Gait as a biometric can be seen as advantageous over other forms of biometric identification techniques for the following reasons:

- **Unobtrusive**: The gait of a person walking can be extracted without user knowing that they are being analyzed and without any cooperation from the user in the information gathering stage unlike fingerprinting or retina scans.

- **Distance Recognition**: The gait of an individual can be captured at a distance unlike other biometrics such as fingerprint recognition and face recognition.

- **Reduced Detail**: Gait recognition does not require images that have been captured to be of very high quality unlike other biometrics such as face recognition, which can be easily affected by low resolution images.

- **Difficult to Conceal**: The gait of an individual is difficult to disguise, by trying to do so the individual will probably appear more suspicious. With other biometric techniques such as face recognition, the individuals face can easily be altered or hidden.

Implementing real life activity recognition system is a daunting task considering the challenges at each stage of the system like background clutter, dynamic illumination changes, camera movements etc. in the background subtraction stage, partial occlusions in the tracking and feature extraction stages. The performance of the recognition system depends upon these stages. The activity recognition problem is characterized by large intra class variability introduced by various sources like the changes in camera viewpoint, anthropometry (body shapes and sizes of different actors), different dressing styles, changes in execution rate of activity, individual styles of actors, skin colors, etc. as shown in Figure 1, Figure 2, Figure 3, and Figure 4.

This paper focuses on the design, implementation, and evaluation of activity recognition system through gait in video sequences. It introduces a novel method of identifying activities only on the basis of leg components and waist component. The use of waist below components for recognizing the activities makes it to achieve fast activity recognition over the
Figure 1. Lack of reliable image features: Large variation in clothing and color texture

Figure 2. Lack of reliable image features: Variation in skin color

Figure 3. Lack of reliable image features: There might be ambiguities when implementing pose from body silhouette
large databases of videos and hence improves the efficiency and decreases the complexity of the system. This is a model based approach. To recognize the actions, we establish the features of each action from the parameters of human model. Our aim is to develop a human activity recognition system that must work automatically without human intervention.

The rest of the paper is structured as follows: the next section discusses the trend of activity recognition research area in the past decade which introduces the fundamentals of gait recognition systems and human activity recognition models; the section following presents the proposed work of human activity recognition using Gait; the section after that analyzes and evaluates the empirical results of experiments to validate the proposed framework. Before evaluating the proposed system, some hypotheses are established and the evaluations are conducted against these hypotheses; finally the last section summarizes the novelties, achievements, and limitations of the framework, and proposes some future directions of this research.

**LITERATURE REVIEW**

In recent years, various approaches have been proposed for human motion understanding. These approaches generally fall under two major categories: model-based approaches and model-free approaches. Poppe (2010) has made a survey on vision based human action recognition. When people observe human walking patterns, they not only observe the global motion properties, but also interpret the structure of the human body and detect the motion patterns of local body parts. The structure of the human body is generally interpreted based on their prior knowledge. Model-based gait recognition approaches focus on recovering a structural model of human motion, and the gait patterns are then generated from the model parameters for recognition. Model-free approaches make no attempt to recover a structural model of human motion. The features used for gait representation includes: moments of shape, height and stride/width, and other image/shape templates.

**Model-Based Approaches**

Niyogi and Adelson (1994) made an initial attempt for gait-based recognition in a spatio-temporal (XYT) volume. They first find the bounding contours of the walker, and then fit a simplified stick model on them. A characteristic gait pattern in XYT is generated from the model parameters for recognition advocated...
segmentation over time because of robustness. Their procedure involves finding human silhouettes with deformable contours in $X-T$ space or deformable surfaces in $X-Y-T$ space.

Leung & Yang (1995) reported progress on the general problem of segmenting, tracking, and labeling of body parts from a silhouette of the human. Their basic body model consists of five U-shaped ribbons and a body trunk, various joint and mid points, plus a number of structural constraints, such as support. In addition to the basic 2-D model, view-based knowledge is defined for a number of generic human postures (e.g., “side view kneeling model,” “side horse motion”), to aid the interpretation process. The segmentation of the human silhouette is done by detecting moving edges.

Wren et al. (1997) took a region-based approach. Their real-time person finder system “Pfinder” models and tracks the human body using a set of “blobs”; each blob is described in statistical terms by a spatial ($x; y$) and color ($Y; U; V$) gaussian distribution over the pixels it consists of as compare with the shape–color model used in Heisele et al. (1997). The blobs typically correspond to the person’s hands, head, feet, shirt, and pants. A statistical model is also constructed for the background region; here each pixel is described by a Gaussian distribution in terms of color values. At initialization, the background model is used to identify a foreground region with pixel values other than expected given the background model. A model-building process follows where blobs are placed over the foreground region. This process is guided by a 2-D contour shape analysis that attempts to identify various body parts using heuristics. Tracking involves a loop of predicting the appearance of the person in the new image, determining for each pixel the likelihood that it is part of one of the blob models or background model, assigning it to one of the models, and updating the statistical models.

Yoo et al. (2002) estimate hip and knee angles from the body contour by linear regression analysis. Then trigonometric-polynomial interpolant functions are fitted to the angle sequences and the parameters so-obtained are used for recognition. In Lee and Grimson (2002), human silhouette is divided into local regions corresponding to different human body parts, and ellipses are fitted to each region to represent the human structure. Spatial and spectral features are extracted from these local regions for recognition and classification.

In these model-based approaches, the accuracy of human model reconstruction strongly depends on the quality of the extracted human silhouette. In the presence of noise, the estimated parameters may not be reliable. To obtain more reliable estimates, Tanawongsuwan and Bobick (2001) reconstruct the human structure by tracking 3D sensors attached on fixed joint positions. However, their approach needs lots of human interaction.

Wang et al. (2004) build a 2D human cone model, track the walker under the Condensation framework, and extract dynamic features from different body part for gait recognition. Zhang et al. (2007) used a simplified five-link biped locomotion human model for gait recognition. Gait features are first extracted from image sequences, and are then used to train hidden Markov models for recognition.

In Nattapon et al. (2008), an approach for automatic human action recognition is introduced by using the parametric model of human from image sequences using motion/texture based human detection and tracking.

**Model-Free Approaches**

Bobick and Davis (2001) interpret human motion in an image sequence by using motion-energy images (MEI) and motion-history images (MHI). The motion images in a sequence are calculated via differencing between successive frames and then thresholded into binary values. These motion images are accumulated in time and form MEI, which are binary images containing motion blobs. The MEI is later enhanced into MHI, where each pixel value is proportional to the duration of motion at that position. Moment-based features are extracted from MEIs and MHIs and employed for recognition using template matching.
Rajagopalan and Chellappa (2000) described a higher-order spectral analysis-based approach for detecting people by recognizing human motion such as walking or running. In their proposed method, the stride length was determined in every frame as the image sequence evolves.

Vega and Sarkar (2003) offered a novel representation scheme for view-based motion analysis using just the change in the relational statistics among the detected image features, without the need for object models, perfect segmentation, or part-level tracking. They modeled the relational statistics using the probability that a random group of features in an image would exhibit a particular relation. To reduce the representational combinatorics of these relational distributions, they represented them in a Space of Probability Functions (SoPF). Different motion types sweep out different traces in this space. They also demonstrated and evaluated the effectiveness of that representation in the context of recognizing persons from gait.

Heisele et al. (1997) used groups of pixels as basic units for tracking. They grouped the pixels by clustering techniques in spatial \((x, y)\) dimensions and combined color \((R, G, B)\); the motivation for this was that adding spatial information makes clustering more stable than using only single color information. The obtained pixel groups were adapted iteratively from one image to the next image using \(k\)-means clustering algorithm. Because of the fixed number of pixel groups and enforced one-to-one correspondence over time, tracking these units became straightforward. But there was no guarantee that units will remain locked onto the same physical entity during tracking, but initial results on tracking pedestrians appeared promising.

Oren et al. (1997) performed object detection in static images. They used (Haar) wavelet coefficients as low-level intensity features; these coefficients are obtained by applying a differential operator at various locations, scales, and orientations on the image grid of interest. Many coefficients can be part of this representation. Once it has been established which wavelet coefficients to use as features, a support vector machine (SVM) classifier is applied to the training set. During the detection stage, one shifts windows of various sizes over the image, extracts the selected features, and applies the SVM classifier to verify whether the desired object is present or not.

**PROPOSED METHODOLOGY**

Visual surveillance systems play a very crucial role in the circumstances where continuous patrolling by human guards is not possible like international border patrolling, nuclear reactors etc. Demand for automatic surveillance systems in civilian applications like monitoring a parking lot, shopping complexes etc. is also increasing heavily. It is difficult and manpower intensive to monitor the data collected from various cameras continuously and it gives a rise to the necessity for automatic understanding of human actions and building a higher level knowledge of the events occurring in the scene by the computer vision system. The objective of this paper is to propose an efficient method for foreground extraction, human activity analysis and feature extraction analysis such that these methods can be integrated in a framework, to implement a system for recognizing human activities.

The proposed technique of human activity recognition is based on the foreground extraction, human tracking, feature extraction and recognition. Figure 5 shows the framework of the introduced human activity recognition system using Gait to identify four basic human activities (i.e. walking, running, jogging and jumping as shown in Figure 6). The proposed method has following main steps: Foreground Extraction, Human Tracking, Feature Extraction and Activity Recognition. In this framework, the video is given as an input to the system from the activity database and frames are extracted from that video. The parametric model of human is extracted from image sequences using motion/texture based human detection and tracking. After that the results are displayed as the recognized activities like walking, running,
foreground extraction; and finally the performance of the method is tested experimentally using the datasets under indoor and outdoor environments.

**Foreground Extraction**

The first step is to provide a video sequence of an activity as an input in the proposed system from the dataset. That video contains a number of continuous frames. After that background subtraction technique is used to separate moving object present inside those frames. But these frames contain some noises which may lead to incorrect foreground subtraction. So first of all, we remove these noises. Some of the small noises are removed by using morphological image processing tools such as Erosion, Dilation, or Gaussian Filters. Generally, an object might be detected in several fragmented image regions. In that case, a region-fusion operation is needed. Two regions are considered to be the same object if they are overlapped or their distance less than a specific threshold value. With these constraints, the method is again very sensitive to light condition, such as shadow, contrast changing and sudden changes of brightness.

Intuitively, introducing some special characteristics of object, for instance texture properties, will probably improve the better results. Therefore, in the fusion process the color probability density of object’s texture is additionally applied for computing the similarity between regions using Mean-shift algorithm (Cheng, 1995). This mixture of motion and texture of object for detection and tracking can reduce significantly noises and increases

---

Figure 5. Framework of proposed system of human activity recognition
consequently the effectiveness of our tracking algorithm. However, there are always additive noises superposed with detected objects that will be eliminated later by human model constraints.

**Mean Shift Clustering**

The mean shift algorithm is a nonparametric clustering technique which does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters. Hence, mean shift represents a general non-parametric mode finding/clustering procedure. In contrast to the classic K-means clustering approach (Duda et al., 2001), there are no embedded assumptions on the shape of the distribution or the number of modes/clusters. Mean shift was first proposed by Fukunaga & Hostetler (1975), later adapted by Cheng in 1995 for the purpose of image analysis and more recently extended by Comaniciu et al. (2001), to low-level vision problems, including, segmentation, adaptive smoothing (Comaniciu and Meer, 2002) and tracking (Comaniciu et al., 2003).

The main idea behind mean shift is to treat the points in the d-dimensional feature space as an empirical probability density function where dense regions in the feature space correspond to the local maxima or modes of the underlying distribution. For each data point in the feature space, one performs a gradient ascent procedure on the local estimated density until convergence. The stationary points of this procedure represent the modes of the distribution. Furthermore, the data points associated (at least approximately) with the same stationary point are considered members of the same cluster.

Given n data points $x_i, i = 1, ..., n$ on a d-dimensional space $R^d$, the multivariate kernel density estimate obtained with kernel $K(x)$ and window radius $h$ is

Figure 6. Example Sequences from KTH action dataset (a), (b), (c) and Weizmann classification database (d)

(a) Walking  (b) Jogging  
(c) Running  (d) Jumping
\[ f(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K \left( \frac{x - x_i}{h} \right) \]  

where, \( h \) (termed the \textit{bandwidth} parameter) defines the radius of kernel. The radially symmetric kernel is defined as,

\[ K(x) = c_k x^d \]  

where, \( c_k \) represents a normalization constant which assures \( K(x) \) integrates to 1. Taking the gradient of the density estimator and some further algebraic manipulation yields,

\[ \nabla f(x) = \frac{2c_k d}{nd^{d+2}} \sum_{i=1}^{n} g \left( \frac{x - x_i}{h} \right) \left( \sum_{i=1}^{n} x_i g \left( \frac{x - x_i}{h} \right) \right) - x \]  

where, \( g(x) = -k_0(x) \) denotes the derivative of the selected kernel profile. The first term is directly proportional to the density estimate at \( x \) (computed with the kernel \( G(x) = c_k x^d \)). The second term, i.e. term 2, called the \textit{mean shift} vector, \( \mathbf{m} \), points toward the direction of maximum increase in density and is proportional to the density gradient estimate at point \( x \) obtained with kernel \( K \).

The mean shift procedure for a given point \( x_i \) is as follows:

1. Compute the mean shift vector \( \mathbf{m} \left( x_i \right) \).
2. Translate density estimation window:
   \[ x_i^{t+1} = x_i^t + \mathbf{m} \left( x_i^t \right) \].
3. Iterate steps 1 and 2 until the gradient of density function is zero.

The mean shift clustering algorithm is a practical application of the mode finding procedure (Figure 7 shows the mean shift mode finding process):

1. Starting on the data points, run mean shift procedure to find the stationary points of the density function,
2. Prune these points by retaining only the local maxima.

The set of all locations that converges to the same mode defines the basin of attraction of that mode. The points which are in the same basin of attraction are associated with the same cluster. The most computationally expensive component of the mean shift procedure corresponds to identifying the neighbors of a point in space (as defined by the kernel and its bandwidth); this problem is known as multidimensional range searching in the computational geometry literature. This computation becomes unwieldy for high dimensional feature spaces. Proposed solutions to this problem include, embedding the mean shift procedure into a fine-to-coarse hierarchical bandwidth approach (DeMenthon, & Megret, 2002) and employing approximate nearest-neighbor hashing-based search (Georgescu et al., 2003).

**Human Tracking and Activity Recognition**

In this phase, we apply Hu-moments (Hu, 1962) for shape analysis in which Zero- to third-order moments are used for shape recognition and orientation as well as for the location tracking of the shape. Hu-moments are invariant to translation, rotation and scaling.

Hu derived expressions from algebraic invariants applied to the moment generating function under a rotation transformation. They consist of groups of nonlinear centralized moment expressions. The result is a set of absolute orthogonal (i.e. rotation) moment invariants, which can be used for scale, position, and rotation invariant pattern identification. The advantage of using Hu invariant moment is that
it can be used for disjoint shapes. In particular, Hu invariant moment set consists of seven values computed by normalizing central moments through order three. In terms of central moment the seven moments are given as below:

\[
M_1 = \eta_{20} + \eta_{02}
\]
\[
M_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2
\]
\[
M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2
\]
\[
M_4 = (h_{30} + h_{12})^2 + (h_{21} + h_{03})^2
\]
\[
M_5 = (h_{30} - 3h_{12})[(h_{30} + h_{12})^2 - 3(h_{21} + h_{03})^2] + (3h_{21} - h_{03})[(3h_{30} + h_{12})^2 - (h_{21} + h_{03})^2]
\]
\[
M_6 = (h_{20} - h_{02})[(h_{30} + h_{12})^2 - (h_{21} + h_{03})^2] + [4h_{11}(h_{30} + h_{12})(h_{21} + h_{03})]
\]
\[
M_7 = (3h_{21} - h_{03})[(h_{30} + h_{12})^2 - 3(h_{21} + h_{03})^2] + (3h_{21} - h_{03})[(3h_{30} + h_{12})^2 - (h_{21} + h_{03})^2]
\]

These seven values given by Hu are used as a feature vector for centroid in the human model.

**Feature Extraction**

We employed a model based approach to extract the features. The extracted foreground that supposed to be a human is segmented into centroid and two leg components. We use Mean-shift algorithm again for computing the similar regions below the centroid of the human body for each leg components that will serve for tracking legs. We assume that with only these three components of human model the four basic actions could be identified correctly. The human model constraints are used for noise suppression. The three components namely centroid, left leg and right leg (i.e. vm1, vm2, vm3 respectively), are used in order to model parametric approach.

The threshold concept is also used along with the defined method. Threshold calculation is applied as follows: Video sequences from the KTH and Weizmann datasets are normalized on the basis of number of frames and the time of a particular sequence for an activity. The threshold is calculated on the basis of a case study given in (Gazendam & Hof, 2007) and it is illustrated in Table 1.
To recognize the actions, we establish the features of each action from the parameters of human model as follows:

1. **Walking Feature:** In case of walking action, every part of human move generally and approximately in the same direction and speed. Therefore, the walking activity can then be identified by the velocities of all components superior to zero but lesser than a predefined threshold for walking. The threshold is calculated on the basis of previous criteria as shown in Table 1. Note that the significant difference between running and walking strides is that at least one of the feet will be in contact with the principal axis (ground) at any given time as shown in Figure 8.

2. **Jumping Feature:** In case of jumping activity, every part of human moves only vertically and in the same direction either up or down. Therefore, jumping action can be identified by the velocities of all the three components to be near or equal to zero in horizontal direction but greater than zero in vertical direction as shown in Figure 9.

3. **Jogging Feature:** The only differences between jogging and running activities were that travelling speed of running is greater than jogging and other difference is of distance ratio between the leg components to the axis of ground as shown in Figure 10.

4. **Running Feature:** Similarly in case of running activity, speed of travelling is greater than jogging and the other difference is of distance ratio between leg components to the axis of ground as shown in Figure 11.

### Algorithm for Human Activity Recognition

**Step1:** Input is fed to the system as a single video sequence.

**Step2:** Frames are extracted from the input video, which are used for further processing.

Table 1. Threshold for activity classification

<table>
<thead>
<tr>
<th>Activity</th>
<th>Normalized time duration for a single sequence</th>
<th>Number of frames for that duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jogging</td>
<td>1.80–2.74 sec</td>
<td>27–42 frames</td>
</tr>
<tr>
<td>Running</td>
<td>0.00–1.79 sec</td>
<td>0–26 frames</td>
</tr>
<tr>
<td>Walking</td>
<td>2.75–5.80 sec</td>
<td>43–87 frames</td>
</tr>
</tbody>
</table>

Figure 8. Silhouette pattern for walking
Figure 9. Silhouette pattern for jumping

Figure 10. Silhouette pattern for jogging

Figure 11. Silhouette pattern for running
Step 3: Background subtraction technique is implemented to subtract background from the frames in order to obtain the foreground moving object.

Step 4: Morphological operators are used to remove additional noises in the frames.

Step 5: Mean-shift algorithm is used to track the human; based on the texture similarities in the frames.

Step 6: Hu-moments are calculated to recognize the centroid of the tracked human. Again the Mean-shift algorithm is used to recognize each leg components of the model.

Step 7: For feature extraction, model based approach is employed. The extracted foreground that supposed to be human is then segmented into centroid and the two leg components i.e., total three components.

Step 8: The features of each action from the parameters of human model acts as the features for classifying all four activities (walking, jumping, jogging and running).

Step 9: The features depend on the following criteria:

- **Walking**: It can be identified as the velocities of all components that are greater than zero but lesser than a predefined threshold and at least one of the feet will be in contact with the principal axis at any given time.

- **Jumping**: It can be identified by the velocities of all components to be near or equal to zero in horizontal direction but greater than zero in vertical direction.

- **Jogging and Running**: The only differences between jogging and running activities were that travelling speed of running is greater than jogging and other difference is of distance ratio between the leg components to the ground axis and there is a period of time during each stride in which neither of your feet is in contact with the ground.

**RESULTS AND DISCUSSIONS**

This section analyses the various aspects of the proposed method. In activity recognition through gait, feature requirement is the main issue to model the human according to the parameters to fulfill the criteria.

**Data Set Used**

In order to evaluate our proposed approach of human activity recognition, we have used two datasets: (1) KTH Human Actions dataset (http://www.nada.kth.se/cvap/actions) and (2) Weizmann Actions dataset (http://www.wisdom.weizmann.ac.il/~vision/SpaceTime-Actions.html).

**KTH Human Actions Dataset**

KTH video dataset uses six types of human actions such as “walking”, “jogging”, “running”, “boxing”, “hand waving” and “hand clapping”, which were performed by 25 subjects in different scenarios with different clothing conditions as well. The video sequences are down sampled to 160*120 pixels and an average length varying from 4 to 41 seconds. This dataset contains 2391 activity sequences. All videos are having static background with 25 fps. We use walking, jogging and running sequences of KTH actions data set for evaluation. Examples of some sequences are shown in Figure 12.

**Weizmann Actions Dataset**

Weizmann Actions dataset uses ten types of natural human actions such as “run,” “walk,” “skip,” “jumping-jack”, “jump-forward-on-two-legs”, “jump-in-place-on-two-legs”, “gallop sideways”, “wave-two-hands”, “wave-one-hand”, or “bend” which are performed by 9 different people in different scenarios with different clothing conditions as well. The video sequences are down sampled to 184*144 pixels and an average length varying from 2 to 4 seconds.
This dataset contains 90 low resolution activity sequences. All the videos are having static background and running with 50 fps. We use walking, jogging and jumping sequences of Weizmann Actions dataset in this paper.

Examples of some sequences are shown in Figure 13.

We have used templates of Mean Shift Clustering and Hu-Moments for jogging, running, walking and jumping activities as shown.

Figure 12. Examples of some sequences of KTH dataset for (a) walking, (b) jogging and (c) running

Figure 13. Examples of some sequences of Weizmann Actions dataset for (a) walking, (b) jogging and (c) jumping
Experimental Results

We have performed the human activity recognition experiments, with the proposed technique, on several videos, captured in outdoor and indoor environment. We have used two standard dataset namely KTH action dataset and Weizmann action dataset. In this paper, we have performed the experiments considering both indoor and outdoor scenario using KTH action dataset. But we have performed on only outdoor images of Weizmann action dataset.

Results on KTH Dataset

Figure 15, 16 and 17 show the different frames of experimental results at different time instances on a standard KTH actions dataset. In Figure 15, first image of frame 5 shows that a human is walking. Second image of frame 5 shows the corresponding recognition result as walking with good accuracy. In Figure 16, first image of frame 10 shows that a human is jogging. Second image of frame 10 shows the corresponding recognition result as jogging. In Figure 17, first image of frame 3 shows that a human is running. Second image of frame 3 shows the corresponding recognition result as running with good accuracy.

Results on Weizmann Dataset

To validate the robustness of our proposed method, we experimented on a standard Weizmann dataset. Figure 18, 19 and 20 shows the frame by frame result analysis of different human activity on this dataset at different time instances.

In Figure 18, first image of frame 5 shows that a human is walking in outdoor environment. Second image of frame 5 shows the corresponding recognition result as walking with good accuracy. In Figure 19, first image of frame 7 shows that a human is running in outdoor environment. Second image of frame 1 shows the corresponding recognition result as running with good accuracy. In Figure 20, first image of frame 1 shows that a human is jumping in outdoor environment. Second image of frame 1 shows the corresponding recognition result as jumping with good accuracy.

Figure 14. Templates of (a) jogging, (b) running, (c) walking, and (d) jumping for human activities
Result Analysis

Accuracy of proposed method is measured based on the number of frames recognized and number of frames not recognized by the following formulae:

\[
\text{Accuracy (\%)} = \frac{\text{No. of frames correctly recognized}}{\text{Total no. of video frames in a sequence}} \times 100
\]

Table 2 shows the accuracy of the recognition method using Gait over two large datasets.

Figure 15. Result on standard KTH dataset from of walking; first image shows input frame, second image shows corresponding output image; at the end, it recognize human activity as “Walking”
with encouraging results; up to 95.01% of activities are recognized correctly in KTH dataset and 91.36% of activities are recognized correctly in Weizmann dataset. We have calculated the accuracy in both indoor and outdoor scenarios in the case of KTH dataset.
CONCLUSION

An efficient human activity recognition using gait technique based on model based approach is introduced in this paper which uses Mean shift clustering algorithm and Hu-Moments to construct the activity templates. This method has a promising execution speed of 25 frames per second and good activity recognition accuracy. The experimental results demonstrate...
that the proposed method accurately recognizes different activities in various video frames considering both indoor and outdoor scenarios while maintaining a high recognition accuracy rate. Currently our method determines key poses of each activity independently using parametric model only. Different activity classes may give similar key poses which may cause confusion.
and redundancy in recognition. More discriminative key poses can be applied jointly using some more refined and sophisticated algorithms such as Support Vector Machine (SVM).
Figure 20. Experimental result on standard Weizmann dataset of jumping; first image shows input frame, second image shows corresponding output image; at the end of each sub-sequence it recognize human activity as “Jumping”

Table 2. Table shows the result analysis of proposed method on KTH Human Actions dataset and Weizmann Actions dataset on the basis of frames

<table>
<thead>
<tr>
<th>Name of Dataset</th>
<th>Environment condition</th>
<th>Human Activities</th>
<th>Number of Frames</th>
<th>Number of Frames recognized</th>
<th>Recognition rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>KTH Dataset</td>
<td>Outdoor Walking</td>
<td>1443</td>
<td>1434</td>
<td>99.3%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Indoor Walking</td>
<td>1415</td>
<td>1383</td>
<td>97.7%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Outdoor Jogging</td>
<td>1525</td>
<td>1425</td>
<td>93.4%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Indoor Jogging</td>
<td>1218</td>
<td>1157</td>
<td>94.9%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Outdoor Running</td>
<td>1089</td>
<td>980</td>
<td>89.9%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Indoor Running</td>
<td>1137</td>
<td>1080</td>
<td>94.9%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Avg. %</td>
<td></td>
<td></td>
<td></td>
<td>95.01%</td>
</tr>
<tr>
<td>Weizmann Dataset</td>
<td>Outdoor Walking</td>
<td>678</td>
<td>650</td>
<td>95.8%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Outdoor Running</td>
<td>588</td>
<td>552</td>
<td>93.8%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Outdoor Jumping</td>
<td>756</td>
<td>642</td>
<td>84.5%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Avg.%</td>
<td></td>
<td></td>
<td></td>
<td>91.36%</td>
</tr>
</tbody>
</table>
REFERENCES


