Recognition of Human Actions in Videos using Computer Vision Techniques

B. Hemanth Kumar
M.Tech (Computer Science and technology), Department of Information Technology,
VR Siddhartha Engineering College,
Vijayawada, Andhra Pradesh, India.

M. Ashok Kumar
Asst. professor, Department of Information Technology,
VR Siddhartha Engineering College,
Vijayawada, Andhra Pradesh, India.

V. Radhesyam
Asst. professor, Department of Information Technology,
VR Siddhartha Engineering College,
Vijayawada, Andhra Pradesh, India.

B. L. N. Phaneendra kumar
Asst. professor, Department of Information Technology,
VR Siddhartha Engineering College,
Vijayawada, Andhra Pradesh, India.

Abstract-A primary phase now undertaking this issue is to achieve automatic recognition of human actions that are typical. In order to make human-computer interfaces genuinely likely, we need to develop a technology that tracks human effort, body behavior and interpret these movements in an effective way. The goal of the action recognition is to analyze the ongoing events automatically from video data. A reliable system which is capable of recognizing various human actions has many important applications that include surveillance systems, health-care systems, and a variety of systems which involve interactions between persons and electronic devices such as human-computer interfaces.

Keywords: Recognition, Action, Sequence, Testing, Training.

I. Introduction

An activity of human is a blend of a few miniaturized scale activity groupings performed by one or more body parts of the human. An occasion is set of activities performed in a grouping. Human activity acknowledgement intends to perceive the activities and objectives of one or more specialists from a progression of perceptions on the operators' activities. Human activity acknowledgement is an essential region of PC vision examination and applications [5]. The objective of the activity acknowledgement is a mechanized investigation or translation of continuous occasions and their connection from video information. Its applications incorporate reconnaissance frameworks, persistent checking frameworks, and an assortment of frameworks that include cooperation's amongst people and electronic gadgets, for example, human-PC interfaces.

A. Action:

An act that one consciously wills and that may be characterized by physical or mental activity is said to be an action. We can recognize human actions by capturing local events in video and can be adapted to the size, the frequency and the velocity of moving patterns. An observer can gain a comprehensive description of the purposes of actions by watching the other person's detailed body movement. Action recognition has traditionally studied processing fixed camera observations while ignoring non-visual information.

B. Types of Actions:

The different types of actions from various datasets are: Boxing, Hand Waving, and Hand Clapping, Walking, Jogging, Running, Answerphone, Drive Car, Eat, Fight Person, brush hair, catch, clap, climb stairs, golf, jump, kick ball, pick, pour, pull-up, push, run, shoot ball, shoot bow, shoot gun, sit, stand, swing baseball, throw, walk, wave, GetOutCar, Handshake, Hug Person, Kiss, Run, Sit-down, Sit-up, and Standup[4].

II. RELATED WORK

A. Hollywood2 Dataset:

Hollywood2 dataset has 12 classes of human actions and 10 classes of scenes distributed over 3669 video clips and approximately 20.1 hours of video in total. The dataset intends to provide a comprehensive benchmark for human action recognition in realistic and challenging settings. The actions in the Hollywood2 dataset are: Answerphone, Drive Car, Eat, Fight Person, GetOutCar, Handshake, Hug Person, Kiss, Run, Sit-down, Sit-up, and Standup.

Figure 1: Human actions in Hollywood-2 Data set

B. KTH Dataset:

KTH video database contains six types of human actions (walking, jogging, running, boxing, hand waving and hand clapping) performed several times by 25 subjects in four different scenarios. The actions in the KTH dataset: *Walking, Jogging, Running, Boxing, Hand Waving, and Hand Clapping.*

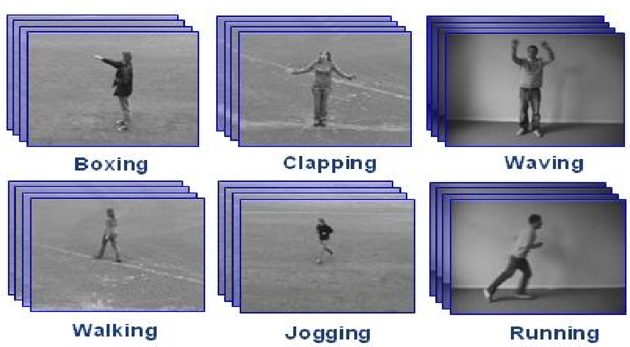


Figure 2: Human actions in KTH Data set

C. JHMDB Dataset:

The HMDB51 database contains more than 5,100 clips of 51 different human actions collected from movies or the Internet. Annotating this entire dataset is impractical so J-HMDB is a subset with fewer categories. The result contains 21 categories involving a single person in action: *brush hair, catch, clap, climb stairs, golf, jump, kick ball, pick, pour, pull-up, push, run, shoot ball, shoot bow, shoot gun, sit, stand, swing baseball, throw, walk, wave.*

Figure 3: Human actions in J-HMDB Data set

Table-1 Description of video data sets for action recognition

S.NO	Database name	No of classes	Description	Algorithms	Constraints
1	KTH	6	600 videos (192 training, 192 validation, 216 testing). Resolution = 160x120. Black and white videos. Static camera.	Recognition: multi-class recognition accuracy.	Homogeneous indoor/outdoor backgrounds. Performed by 25 persons, 4 scenes, and 6 verbs. Provided annotations: labeled temporal segments.
2	HOLLYWOOD 2	12	2859 videos from clean labeled. For actions: 823 training, 884 testing. For scenes: 570 training, 582 testing	Recognition: (mean) average-precision for action, scene, and the combination of two.	Short sequences from 69 movies. An expansion of the Hollywood dataset. Provide videos of scene for both training and testing.
3	J-HMDB	21	5,100 videos of 51 different human actions	Recognition: pose, action and the combination of both	This selection-and cleaning process results in 36-55 clips per action class with each clip containing 15-40 frames.

Table -2 Human Action recognition literature survey

S.NO	Title	Authors
1	Real-Time Human Action Recognition Based on Depth Motion Maps	Chen Chen, Kui Liu, and Nasser Kehtarnavaz

<table>
<tr><td colspan="3">Description: A human action recognition method by using depth motion maps (DMMs). Each depth frame in a depth video sequence is projected onto three orthogonal Cartesian planes [1]. The absolute difference between two consecutive projected maps is accumulated through an entire depth video sequence forming a DMM. An l2-regularized collaborative representation classifier with a distance-weighted Tikhonov matrix is then employed for action recognition.</td></tr>
<tr><td>2</td><td>Generic Human Action Recognition</td><td>Hae Jong Seo and Peyman Milanfar</td></tr>
<tr><td colspan="3">Description: A novel human action recognition method based on space-time locally adaptive regression kernels and the matrix cosine similarity measure. [8]The computation of the so-called local steering kernels as space-time descriptors from a query video, which measure the likeness of a voxel to its surroundings.</td></tr>
<tr><td>3</td><td>Human Action Recognition: Contour-Based and Silhouette-Based Approaches</td><td>Salim Al-Ali, Mariofanna Milanova, Hussain Al-Rizzo and Victoria Lynn Fox</td></tr>
<tr><td colspan="3">Description: Human action recognition in videos is a desired field in computer vision applications since it can be applied in human computer interaction, surveillance monitors, robot vision, etc. [7]Two approaches of features are investigated in this chapter. First approach is a contour-based type. The second approach is a silhouette-based type. The classification is achieved using two classifiers: K Nearest- Neighbor (KNN) and Support Vector Machine (SVM).</td></tr>
<tr><td>4</td><td>Human Action Recognition</td><td>Rajendra Kumar</td></tr>
<tr><td colspan="3">Description: Human action recognition is an important topic of computer vision research and applications. The goal of the action recognition is an automated analysis of ongoing events from video data. A reliable system capable of recognizing various human actions has many important applications[4]. The applications include surveillance systems, health-care systems, and a variety of systems that involve interactions between persons and electronic devices such as human-computer interfaces.</td></tr>
<tr><td>5</td><td>Human Action Recognition by Representing 3D Human Skeletons</td><td>Raviteja Vemulapalli, Felipe Arrate, and Rama Chellappa</td></tr>
<tr><td colspan="3">Description: Cost-effective depth sensors coupled with the real-time skeleton estimation algorithm of Shotton et al. have resulted in a renewed interest in skeleton-based human action recognition. Most of the existing skeleton-based approaches use either the joint locations or the joint angles to represent a human skeleton.</td></tr>
</table>

III. Proposed system

The input for this method can be either a video file or a video stream from the camera supported by OpenCV camera interface. The code detects Space-Time Interest Points (STIPs) and computes corresponding local space-time descriptors. This simplification appears to produce similar (or better) results in applications (e.g. action recognition) while resulting in a considerable speed-up and close-to-video-rate run time. The code detects extraction points using edge detection technique and from the extraction points classify the action. By the classification gainn the knowledge from the videos and recognize the what type of action it is.

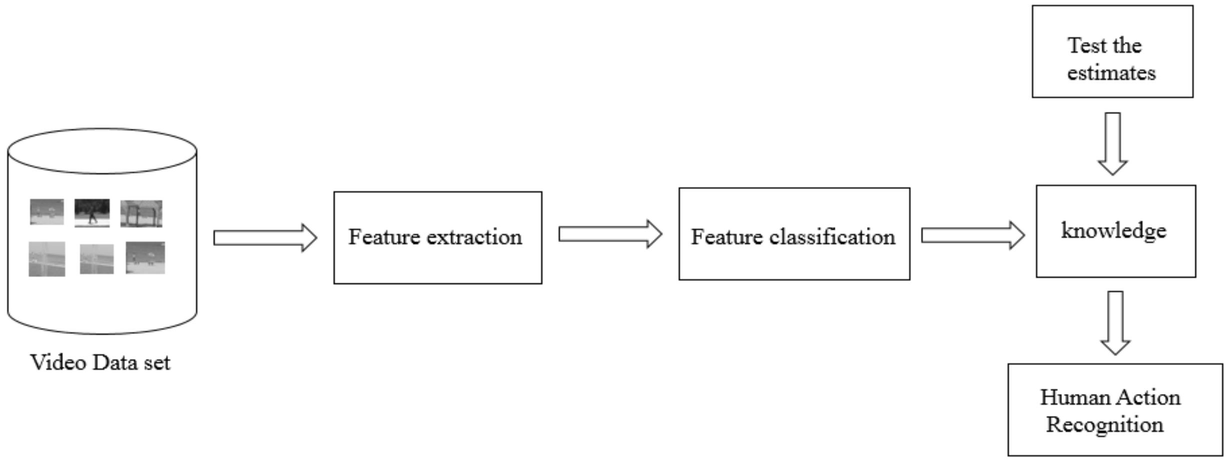


Figure 4: Architecture for Human action recognition

Classifiers

A. Space-Time Interest Points (STIP):

- ✓ STIP is an image descriptor for image-based matching and recognition.
- ✓ The STIP descriptor is invariant to translations, rotations and scaling transformations in the image domain and robust to moderate perspective transformations and illumination variations.
- ✓ The image is represented in the form of linear scale representation by convoluting the image function with Gaussian Kernel.
- ✓ The significant points are obtained as μ = second moment matrix integrated over Gaussian weighting function.
- ✓ The Eigen values of μ are calculated which are the descriptors and are represented by λ.
- ✓ The histogram of oriented gradients (HOG) is a feature descriptor used in computer vision and image processing for the purpose of object detection.

B. K nearest neighbor (KNN):

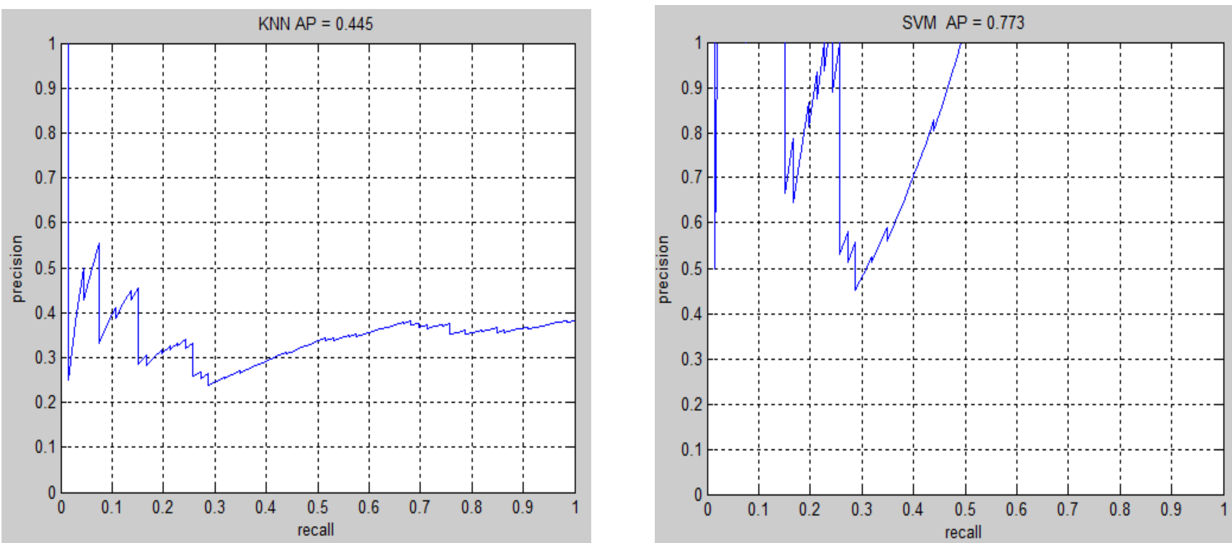
- ✓ The k-means algorithm is an algorithm to cluster n objects based on attributes into k partitions, where $k < n$. k-means algorithm is implemented in 4 steps:
- ✓ Partition objects into k nonempty subsets.
- ✓ Compute seed points as the centroids of the clusters of the current partition. The centroid is the center (mean point) of the cluster.
- ✓ Assign each object to the cluster with the nearest seed point.

C. Support Vector Machine (SVM):

- ✓ SVM is a classification method for both linear and nonlinear points.
- ✓ It uses a nonlinear mapping to transform the original training points into a higher dimension points.
- ✓ With the new dimension, it searches for the linear optimal separating hyper plane (i.e., “decision boundary”).
- ✓ SVM finds this hyper plane using support vectors (“essential” training videos) and margins (defined by the support vectors).

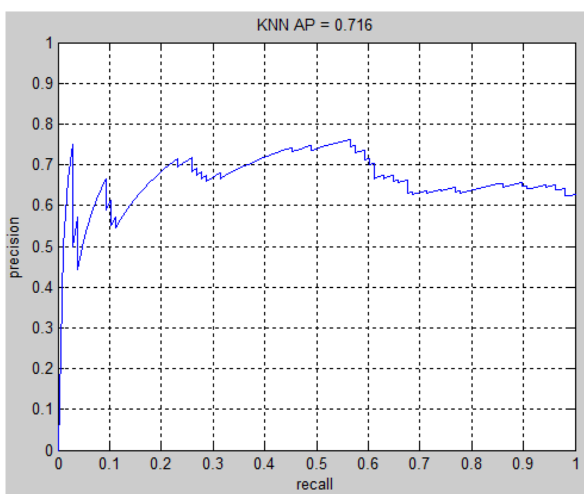
IV. PROPOSED SYSTEM

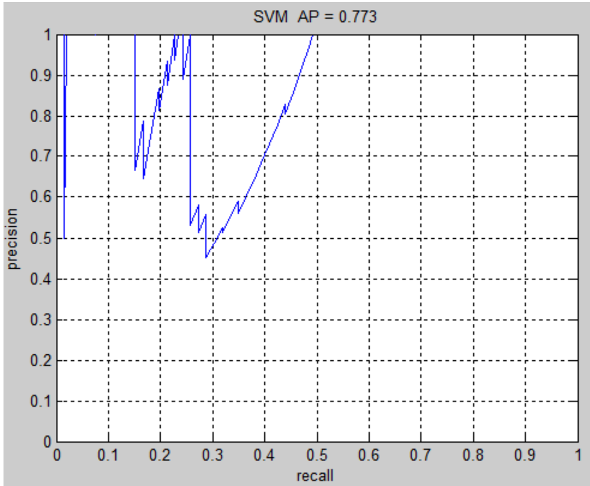
In this paper we show the results of the human action events in videos. By the KTH dataset we just train the videos and from the HOLLYWOOD-2 dataset we done with testing and classifying. Displays the videos as a sequence of the frames. Finds the interest points using the STIP (Space-Time Interest point's) algorithm. Shows the obtained interest points in the video. The obtained interest points are clustered using the K-means clustering algorithm. KNN & SVM classification is applied on the clustered points.



(a) (b)

Figure 5: A graph plotted average points for Hug person action video accuracy of KNN and SVM





(a) (b)

Figure 6: A graph plotted average points for sit down action video accuracy of KNN and SVM

Table -3 Accuracy percentage for different actions using different classifier

	Human Actions		
Classifier	**Hug Person**	**Sit Down**	**Running**
KNN	0.445	0.716	0.545
SVM	0.773	0.773	0.687

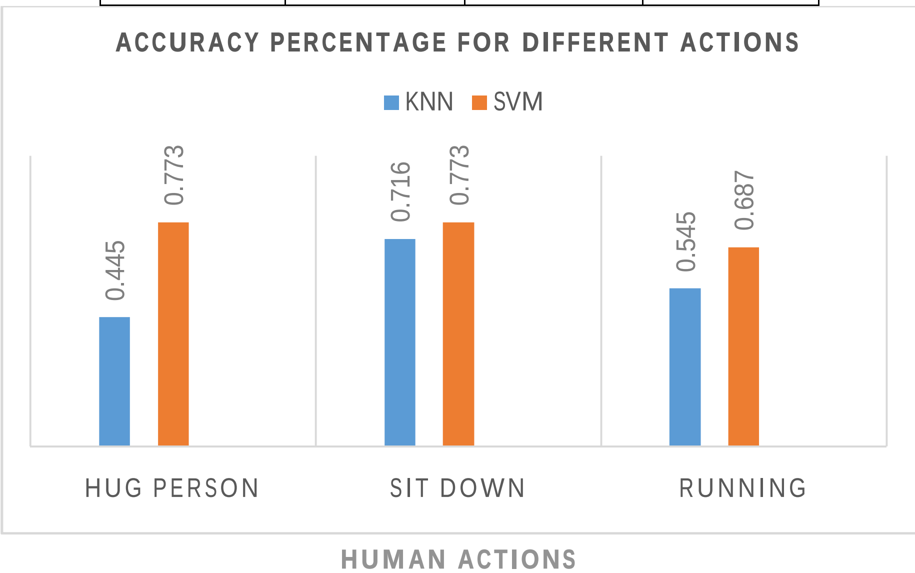


Figure 7: Bar graph representation for different actions

V. CONCLUSION AND FURTHER WORK

In this paper we evaluate the performance of human action recognition. We use different classifiers to find the accuracy of different actions based on space time interest points (STIP). To classify the human actions we use K-Nearest Neighbor (KNN) and Support Vector Machine (SVM). The future work is to improve the technique for surveillance purposes and to extend the implemented algorithm in a way that the results are obtained in a detailed manner and test the method on a variety of data sets. STIP effectively captures the local structure in spatio temporal dimensions of the video sequence. For better performance and better use CNN features for action recognition.

REFERENCE

[1]. Ivan Laptev- http://www.di.ens.fr/willow/events/cvml2010/materials/practical-laptev
[2]. F. Perronnin, J. S´anchez, and T. Mensink. Improving the fisher kernel for large-scale image classification. In ECCV.2010
[3]. Hao Zhang Alexander C. Berg Michael Maire Jitendra Malik Computer Science Division, EECS Department Univ. of California, Berkeley, CA 94720 {nhz,aberg,mmaire,malik} @eecs.berkeley.edu
[4]. Fei-Fei Li; Perona, P. (2005). "2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)" 2: 524. doi:10.1109/CVPR.2005.16. ISBN 0-7695-2372-2.
[5]. N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. In ECCV, pages 428–441. 2006. 4323
[6]. Hae Jong Seo · Peyman Milanfar Generic Human Action Recognition University of California at Santa Cruz 1156 High Street, Santa Cruz, CA, USA.
[7]. Raviteja Vemulapalli, Felipe Arrate and Rama Chellappa Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group Center for Automation Research UMIACS, University of Maryland.
[8]. Chen Chen • Kui Liu • Nasser Kehtarnavaz Real-time human action recognition based on depth motion maps The University of Texas at Dallas, Richardson, TX, USA.
[9]. Salim Al-Ali, Mariofanna Milanova, Hussain Al-Rizzo and Victoria Lynn Fox Human Action Recognition: Contour-Based and Silhouette-Based Approaches Department of Computer Science, University of Arkansas at Little Rock, 2801 S. University Avenue, Little Rock, AR 72204, USA