

# Multi View Action Detection and Recognition of Human Activities

Harshitha D  
Students, Dept. of ECE,  
GSSSIETW, Mysuru

Keerthirani H V  
Students, Dept. of ECE,  
GSSSIETW, Mysuru

Pallavi S  
Students, Dept. of ECE,  
GSSSIETW, Mysuru

Sindhu Shree K M  
Students, Dept. of ECE,  
GSSSIETW, Mysuru

Spoorthi Y  
Assistant Professor,  
Dept. of ECE,  
GSSSIETW, Mysuru

**Abstract**—The problem of human activity recognition can be approached using spatio-temporal variations in successive video frames. In this paper, a new human activity recognition technique is proposed using multi-view videos. Initially, a naive background subtraction using frame differencing between adjacent frames of a video is performed. Then, the motion information of each pixel is recorded in binary indicating existence/nonexistence of motion in the frame. A pixel wise sum over all the difference images in a view gives the frequency of motion in each pixel throughout the clip. Detection of suspicious activities in public transport areas using video surveillance has attracted an increasing level of attention. In this project, we introduce a framework that processes raw video data received from a fixed color camera installed at a particular location, which makes real-time inferences about the observed activities.

## I. INTRODUCTION

Human Action Recognition (HAR) is the significant research area in vision of computer and human action/interaction. HAR using computer vision involves the understanding of human motion, which is a complex and challenging task in non-rigid body articulation, loose clothing, and mutual obstruction of the breath passage in the articulation of a speech sound as well as noise present in the image by shadow. For example, outdoor recognition activities are significantly influenced by atmosphere and lighting conditions [1].

HAR is an emerging research area in the field of video analytic system. It mainly involves the following two steps

1. Low-level vision processing such as segmentation, tracking, poses recovery, and trajectory estimation.
2. High level processing tasks such as body modeling and representation of action.

The sequences of Image Processing have progressed from simple structure paradigm of motion to the modeling, classification and recognition of human actions / interactions as events.

## II. LITERATURE SURVEY

Literature survey is abounding with many research works in the field of Human Activity Recognition.

### 1. Deep Learning Fusion Conceptual Frameworks for Complex Human Activity Recognition Using Mobile and Wearable Sensors

In 2018, Nweke Henry Friday, Ghulam Mujtaba, Mohammed Ali Al-garadi, Uzoma Rita Alo, analysed to recognize

activities using mobile or wearable sensor, data are collected using appropriate sensors, segmented, needed features extracted and activities categories using discriminative models (SVM, HMM, MLP etc.).

### 2. Improving human action recognition with two-stream 3D convolutional neural network

In 2018, Van-Minh Khong, Thanh-Hai Tran, They have proposed a method that exploits both RGB and optical flow for human action recognition. Specifically, we deploy a two stream convolutional neural network that takes RGB and optical flow computed from RGB stream as inputs.

### 3. Information Fusion for Human Action Recognition via Biset/Multiset Globality Locality Preserving Canonical Correlation Analysis

In NOVEMBER 2018, Nour El Din Elmadany , Student Member, IEEE, Yifeng He, Member, IEEE, and Ling Guan , Fellow, IEEE proposed two novel information fusion techniques for fusing the information from multisets. The first technique is biset globality locality preserving canonical correlation analysis (BGLPCCA), which aims to learn the common feature subspace between two sets.

### 4. Data Fusion and Multiple Classifier Systems for Human Activity Detection and Health Monitoring: Review and Open Research Directions

In 2018, Henry Friday Nweke, Teh Ying Wah, Ghulam Mujtaba, they have focused of this review is to provide in-depth and comprehensive analysis of data fusion and multiple classifier systems techniques for human activity recognition with emphasis on mobile and wearable devices.

### 5. Facial Recognition System for Suspect Identification Using a Surveillance Camera

In 2017- 2018, V. D. Ambeth Kumar, V. D. Ashok Kumar, S. Malathi, K. Vengatesan and M. Ramakrishnan, They have proposed a model is proposed for facial recognition to identify and alert the system when a person in search has been found at a specific location under the surveillance of a CCTV camera.

III. METHODOLOGY

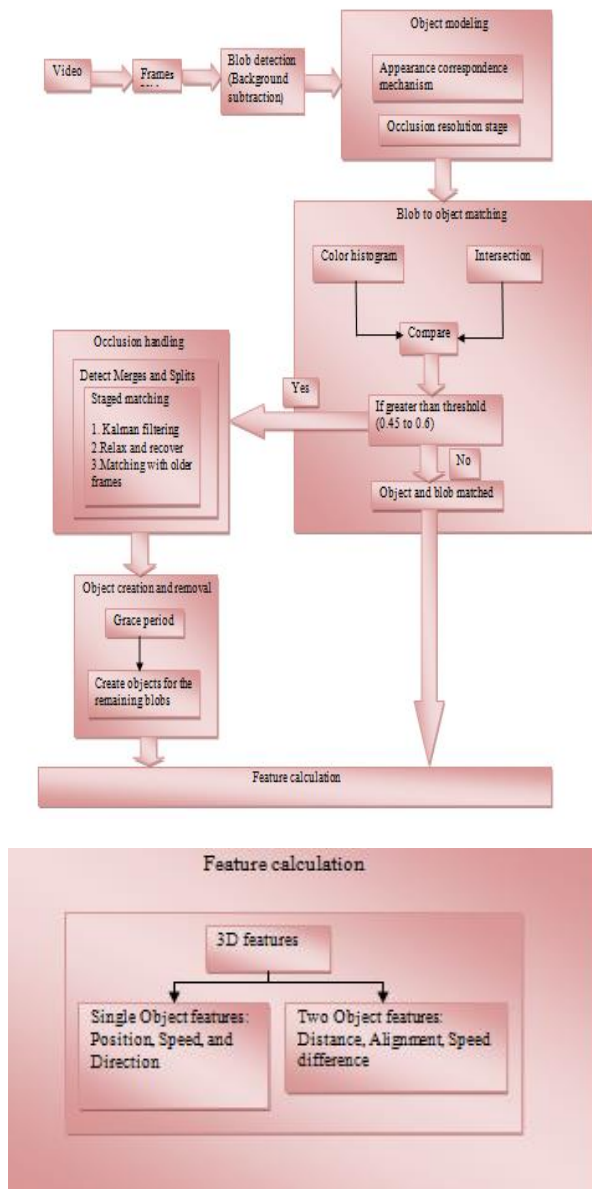


Fig. 1. Flowchart of the proposed human activity recognition technique.

The proposed system focuses on automatically flagging suspicious behavior in public transportation systems. First, the proposed framework obtains 3-D object-level information by detecting and tracking people and luggage in the scene using a real-time blob matching technique. Based on the temporal properties of these blobs, behaviors and events are semantically recognized by employing object and inter-object motion features.

Our framework performs object tracking in an average time of 11 ms per object per frame, whereas behavior recognition averages just about 1 ms per frame. In addition to the single-object features, the inter-object features between every combination of two objects are also stored in historical sequence.

**Hardware Requirements:**

Laptop core i5 Processor Windows Xp

**Software Tool Used: MATLAB**

MATLAB is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation. MATLAB software is used to study and test the performance of activities to Human Action Recognition system.

**Object Modeling And Blob-To-Object Matching:**

In our object tracking approach, at each frame, a list of objects is updated by matching blobs in the current frame with objects from the previous one. This matching process is not necessarily one to-one. Cases of object splits, merges, one-to-one matches, creation, and deletion.

To match blobs and objects in two consecutive frames, color histograms and spatial information are used. The color histograms are all examined to ensure a correct update. The color histograms are adaptively updated at 5 fps using

$$Histogram_{object,t} = \alpha Histogram_{object,t-1} + (1 - \alpha) Histogram_{blob,t} \quad (1)$$

where  $\alpha$  is the learning rate (empirically set by experimentation to approximately 0.6).

A basic form of histogram intersection is given by the following:

$$Intersection(Hist_1, Hist_2) = \sum_{i=1}^n \min(Hist_{1,i}, Hist_{2,i}). \quad (2)$$

Here, we normalize the intersection by the largest of the histograms. The formula for histogram intersection then becomes:

$$Intersection(Hist_1, Hist_2) = \frac{\sum_{i=1}^n \min(Hist_{1,i}, Hist_{2,i})}{\max(\|Hist_1\|, \|Hist_2\|)}. \quad (3)$$

**Occlusion Handling:**

Occlusion handling is a critical task because it bears on the robustness of object tracking and coherence. In this framework, the issue of which objects are occluding which is completely ignored, and we adopt the position that all merged objects form a pool (the blob) with no particular occluding/occluded relationships being noted.

**Object Creation and Removal:**

After all blob-to-object matching cases of merging, splitting, and one-to-one matches have been processed, some of the remaining blobs and objects may still remain unmatched. An unmatched blob is ideally a new object that has just appeared in the scene. Therefore, a new object is created for each unmatched blob.

**Behavior Semantics Recognition:**

**Abandoned and Stolen Objects:** We defines an abandoned object as “a stationary object that has not been touched by a person for some time threshold.” Integrating the object ownership into this statement, showing our refined definitions of abandoned and stolen luggage, which are based on the motion features.

**Loitering:** Loitering is useful for detecting a number of public transit situations such as drug dealing. It is defined as “the presence of an individual in an area for a period of time longer than a given time threshold.”

**Fighting:** Our experiments indicate that the most reliable means for defining fights is in terms of the frequency of object splitting, merging or a presence of high dynamic level of activity.

**Meeting and Walking Together:** Although generally not considered to be suspicious, meeting and walking together may be useful in certain surveillance scenarios. This would be particularly the case were face recognition included as a feature. For example, it might be pertinent for security purposes to flag individuals that meet with a suspicious individual.

**Fainting:** In this project, we use a camera calibration method to resolve the alignment issue in 3-D and account for any nonlinearity in the camera parameters. Assuming the person to be standing, the hypothesized 2-D location of the feet on the floor is computed. To verify this assumption, this location is compared with the actual detected location of the feet in 3-D.

illumination conditions. The experimental results demonstrated successful detection of the various activities of interest.

REFERENCE

- [1] S. Reisman, "Measurement of Physiological Stress", Proceedings of the IEEE, 23<sup>rd</sup> Northeast Bioengineering Conference, pp. 21-23, May 1997.
- [2] R.R. Cornelius, "Theoretical approaches to emotion", Proc. Int. Speech Communication Association (ISCA) Workshop on Speech and Emotion, Belfast, Ireland, 2000.
- [3] A. Savran, K. Ciftci, G. Chanel, J.C. Mota, L.H. Viet, B. Sankur, L. Akarun, A. Caplier, and M. Rombaut, "Emotion Detection in the Loop from Brain Signals and Facial Images", Final Project Report, eINTERFACE'06, Dubrovnik, Croatia, 2006.
- [4] S.A. Hosseini, and M.A. Khalilzadeh, "Qualitative and Quantitative Evaluation of EEG signals in Emotional state with through higher order spectra", 3<sup>rd</sup> Iranian Congress on Fuzzy and Intelligent Systems, July 2009. (Article in Persian).
- [5] A. Bashashati, R.K. Ward, G.E. Birch, M.R. Hashemi, and MA. Khalilzadeh, "Fractal Dimension-Based EEG Biofeedback Slystem", Proceedings of the 25<sup>th</sup> Annual International Conference of the IEEE EMBS Cancun, Mexico, September 2003.

IV. RESULTS

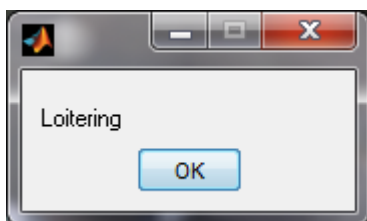
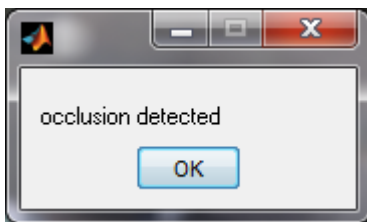
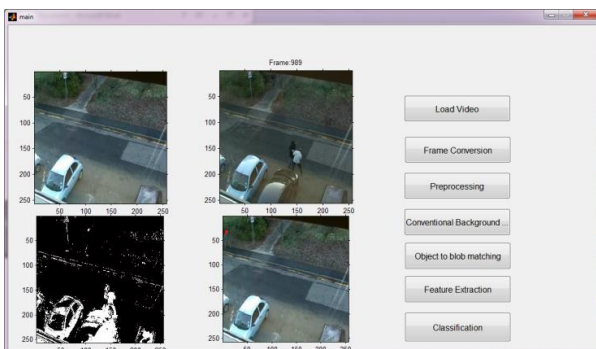


Fig: Screen shots of the output

V. CONCLUSION

In this paper, a complete semantics-based behavior recognition approach that depends on object tracking has been introduced and extensively investigated. Our approach begins by translating the objects obtained by background segmentation into semantic entities in the scene. These objects are tracked in 2-D and classified as being either animate (people) or inanimate (objects). Experimentation was carried out on multiple standard publicly available data sets that varied in terms of crowd density, camera angle, and