ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 2, Issue 2, February 2013

# SUSPICIOUS MOTION DETECTION IN SURVEILLANCE VIDEO

V.Padmavathi[1] Dr.M.Kalaiselvi Geetha[2]

***ABSTRACT***

*Video Surveillance systems are used for any type of monitoring. Action recognition has drawn much attention over the past decade in the computer vision community which is critical for a wide range of applications, such as video surveillance, digital entertainment etc. In action recognition, bag of visual words based approaches have been shown to be successful, for which the quality of dataset is critical. This project proposes a novel approach for key pose selection, which models the descriptor space utilizing a manifold learning technique to recover the geometric structure of the descriptors on a lower dimensional manifold. With the obtained dataset, each action can be represented with a histogram of the key poses. To solve the ambiguity between some action classes, a pairwise subdivision is executed to select discriminative datasets for further recognition.*

## I. INTRODUCTION

Human activity recognition is an important area of computer vision research and applications. The goal of the activity recognition is an automated analysis of ongoing events and their context from video data. **Activity recognition** aims to recognize the actions and goals of one or more agents from a series of observations on the agents actions and the environmental conditions.

Application Areas are Visual Surveillance, Human-Computer Interaction, Sign Language Recognition, News, Movie, Personal Video Archives, Social Evaluation of Movements etc.,

**Video Motion Analysis**

**Video Motion Analysis** is the technique

used to get information about moving objects from video. Examples of this include gait analysis, replays, speed/acceleration calculations, team/individual sports, task performance analysis. The motions analysis technique usually involves a digital movie camera, and a computer that has software allowing frame-by-frame playback of the video.

This paper proposes a novel approach called manifold learning technique for key poses selection. It proposes a PageRank based measure, which assigns relative scores to all nodes in the graph via the recursive principle, to select a key poses according to the geometric structure. A key pose is selected from the manifold in each step and the remaining model is modified to maximize the discriminative power of selected codebook. It models the descriptor space utilizing manifold learning to recover the geometric structure of the descriptors and put the poses on a lower dimension manifold. This approach performs better than the typical k-means clustering based approaches.

## II. SYSTEM ARCHITECTURE

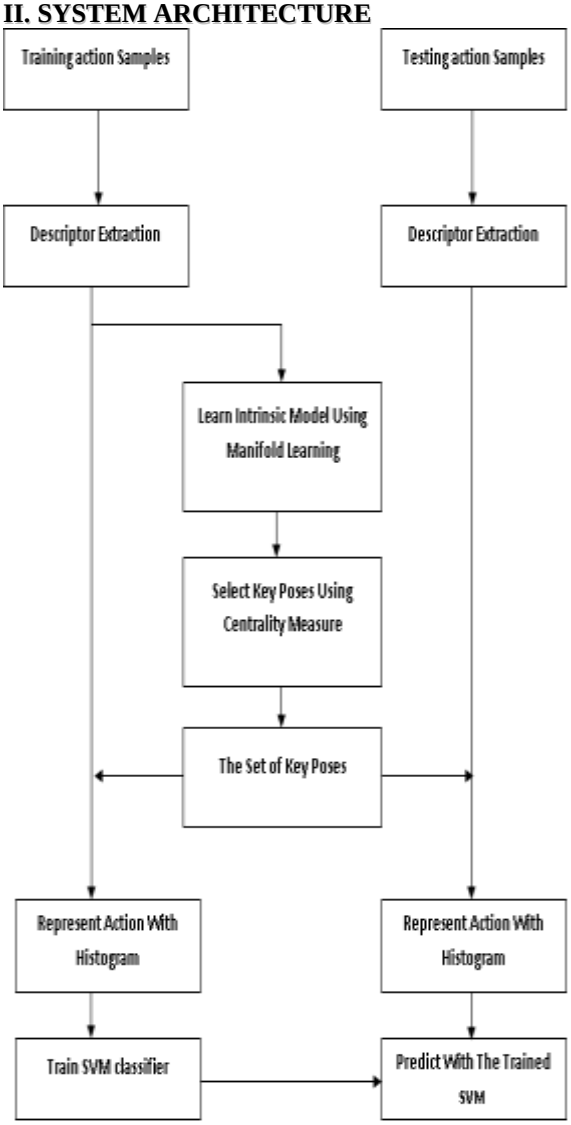



**Figure: Key Pose Identification System**

## III. METHODOLOGIES

### 1. DESCRIPTOR EXTRACTION

In this method the certain descriptors are extracted from the action videos to catch usable information and then a vocabulary called codebook / Data directory is obtained from extracted descriptors. Thus, an action can be represented as a collection of words from the codebook and finally, a discriminative model is build to categorize actions into groups using the extracted descriptors.

### 2. KEY POSE SELECTION

In this method the descriptors have been extracted, a compact set called codebook needs to be learned. In order to catch the relationship between the high-dimensional descriptors and find typical key poses, a refined model for the feature space is needed. Hence a manifold learning algorithm, which learns an internal model of the input data and projects the high-dimensional descriptors to low-dimensional manifold. The intrinsic dimension of descriptor space is found and a dimensional reduction is made in the modeling procedure.

### 3. MULTICLASS CLASSIFICATION

Here each and every frame of the action video corresponds to a descriptor, which can be matched to the nearest neighbor in the codebook. Thus, an action translates to a series of words in the codebook. The frequency of the words in the series can be accumulated into a histogram to represent the action and be used for recognition.

### 4. PAIRWISE SUBDIVISION

A pairwise algorithm in which an additional subdivision is employed for ambiguous couples. In subdivision process, manifold learning is employed to build special model for the given couple. Thus, the centrality measure tends to select key poses which catch the intrinsic difference between given couple and form more powerful codebook for special subdivision problem.

## IV. TECHNIQUES

### i. PAGERANK BASED MEASURE

PageRank is a probability distribution used to represent the likelihood that a person randomly clicking on links will arrive at any particular page. PageRank can be calculated for collections of documents of any size. The PageRank computations require several passes, called iterations, through the collection to adjust approximate PageRank values to more closely reflect the theoretical true value.

### ii. MANIFOLD LEARNING TECHNIQUE

It is a popular recent approach to nonlinear dimensionality reduction. Algorithms described as a function of only a few underlying parameters. The data points are actually samples from a low-dimensional manifold that is embedded in a high-dimensional space. Manifold learning algorithms attempt to uncover the parameters in order to find a low-dimensional representation of the data. The features of such a data set contain much overlapping information, it would be helpful to somehow get a simplified, non-overlapping representation of the data whose features are identifiable with the underlying parameters that govern the data. This suspicion is formalized using the idea of a manifold.

### COMMON DATASETS

In this section, publicly available datasets, which are widely used for activity recognition evaluation. It contains 6 types of action. They are boxing, hand clapping, hand waving, walking, running and jogging.

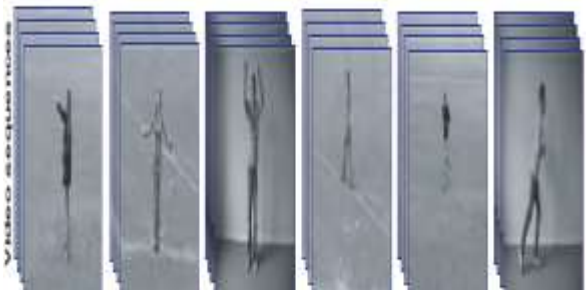


**Figure: A Sample dataset**

## V. RESULTS AND DISCUSSION

Motion detection is a continuous video stream. All of them are based on comparing the current video frame with one from the previous frames or with something that will call background. The application supports the following types of video sources are AVI files and Local Capture Device

ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 2, Issue 2, February 2013

## 1. DETECTION PROCESS

**Background modeling**, which resembled by a morph filter combines the background as an overlay image, and the current frame to decrease the difference with the background, which can be taken as updating the background.
**Temporal variance**, which is accomplished by a Connected Component Labeling Algorithm. That takes connected labeled pixels, which assembles a region in the image, and combines them into object.

## 2. MOTION DETECTION ALGORITHM

The background to be the first frame is received, it have no motion at all, and further process the background by applying a Grayscale filter and a Pixellate Filter. The pixellete filter here used to reduce the pixels count and emphasize the overall color distribution of the image; then the extract the image dimensions to use in further processing. Now the initial background image is ready.

### Updating The Background Image

Update the background image by moving the pixels intensity towards the pixels, intensity of the current frame by one level, to decrease difference with overlay image - source image is moved towards overlay image. The update equation is defined in the next way:
*Result = src + min( abs( ovr - src ), step ) * sign( ovr - src )*
*Result - Updated background image, which will be the background for the next frame.*
*src - Current frame image.*
*ovr - Current background image.*
*step - The maximum amount of changes per pixel in the source image.*

### BLOB Extraction And Counting

Detection of connected components between pixels in binary images is a fundamental step in segmentation of an image objects and regions or *blob*. Each blob is assigned a unique label to separate it from other blobs. All the pixels within a blob of spatially connected 1's are assigned the same label. It can be used to establish boundaries of objects, components of regions, and to count the number of blobs in an image. Its applications can be found in automatic inspection, optical character recognition, robotic vision, etc.

### The Basics

A pixel p at coordinate (x, y) has four direct neighbors $N_4(p)$ and four diagonal neighbors $N_D(p)$. Eight-neighbors $N_8(p)$ of pixel p consist of the union of N4(p) and $N_D(p)$.

To establish connectivity for pixels p and q can be considered:
4-connectivity-connected if q is in $N_4(p)$;
8-connectivity-connected if q is in $N_8(p)$;
m-connectivity-connected if q is in $N_4(P)$, or if q is in $N_D(P)$ and $N_4(p) \cap N_4(q)=\phi$;

## 3. A CONNECTED COMPONENT LABELING ALGORITHM

***Step 1**: Initial labeling.*
***Step 2:** Resolve equivalences.*

To obtain transitive closure the Floyd_Warshall (F-W) algorithm is *used.*
for j = 1 to n
for i = 1 to n
if L[i,j] = 1 then
for k = 1 to n
L[i,k] = L[1, k] OR L[j,k];

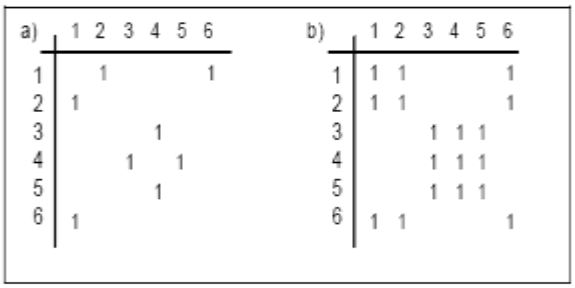



Figure. Equivalence relations in terms of binary matrix.
a) Matrix before applying the F-W algorithm.
b) Matrix after applying reflexivity and the F-W algorithm.

## 4. FAST CONNECTED COMPONENT LABELING ALGORITHM WORKING FLOW

The main idea in this algorithm is to divide the image into NxM small regions. The large equivalence array is the main bottleneck in the original algorithm, but NxN small equivalence arrays can be found in greatly reduced time.

| | | |
|---|---|---|
| Region[1] | Region[2] | Region[3] |
| Region[4] | Region[5] | Region[6] |
| Region[6] | Region[7] | Region[8] |

Figure. Division of original image into 3x3 regions.

***Algorithm***

***Step 1***- Divide the given image into NxN small regions and set Total_ Index = 0

***Step 2:*** For each region i = 1 to NxN

1- apply *Step 1*of the original algorithm;

2- Allocate memory for the array pointed to by Label_List[i] as maximum no. of labels for Region[i];

3- Use F-W algorithm resolve the equivalences within Region[i].

4- For j=1 to size of an array for Region[i] do

Label_List[i][j] = Total_Index + Ibl

// lbl is a label to its equivalence class after equiv. resolution .

5- Total_index= Total_index + maximum{lbl}

6- if (i > 1) then call *Merge( i ):*

// to update labels in bordering area between regions.

***Step3:*** For each region i = l toNxN do

can image in Region[i] from left to right, top to bottom and replace all local label value k with Label_List[i][k];


Figure: Connected Component Labeling Algorithm Working Flow

## VI. TOTAL SYSTEM RESULT

### 1. Based on frame difference

In the simplest way, it takes the difference between two frames and highlight that difference, this method is poor in object recognition because the difference don't emphasize the object shape as in Figure1.


### 2. Based on edge detection

The second best approach is based on edge detection, in spite of its good shape recognition, But its lake in speed and takes a lot of hardware resources shows in Figure2.


### 3. Based On Edge Detection Using Pixellete Filter

A little enhancement in the previous algorithm by adding a pixellete filter obtain a good representation of the object and a fast performance, due to the reduced number of pixels blocks to process as seen by following figure3.


**Using BLOB counter**

After all I eliminate the bounders and replace them by a surrounding rectangle applied to each detected object, because the idea is to detect and track motion not to recognize object shapes in Figure 6.4.

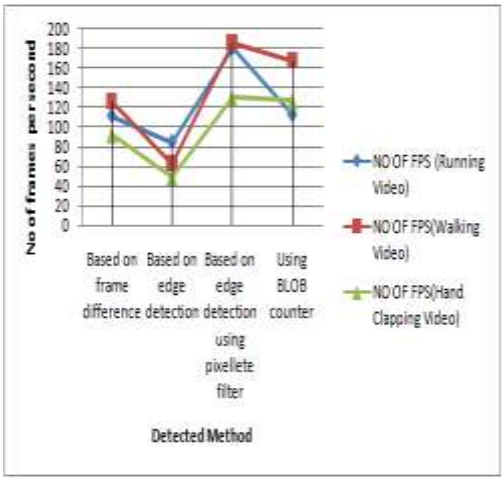

## COMPARISON OF TECHNIQUES

The Comparison of various techniques presented in the following table. Based on edge detection using pixellete filter gives the better result compared to others.

| DETECTOR NAME | NO OF FPS (Running Video) | NO OF FPS(Walking Video) | NO OF FPS(Hand Clapping Video) |
|---|---|---|---|
| Based on frame difference | 111.25 | 125.8 | 92.25 |
| Based on edge detection | 84.6 | 62.82 | 49.15 |
| **Based on edge detection using pixellete filter** | **180.5** | **185.75** | **130** |
| Using BLOB counter | 112.5 | 167.5 | 126.33 |

**Table 5.2 Comparison of techniques**

| N | 4 | 5 | 10 | 15 | 20 | 25 |
|---|---|---|---|---|---|---|
| **CPU(secs)** | 274.57 | 160.22 | 12.47 | 4.01 | 2.75 | 2.47 |




## CONCLUSION

This paper has proposed a novel bag of words based approach for action recognition However, manifold learning is distinct from the methods mentioned earlier. In this proposed framework, manifold learning is employed to generate the codebook. It not only reduces the dimensionality, but also provides a model of the descriptors on a lower dimensional manifold. Thus, PageRank-based centrality measure can be executed to select key poses and obtain the final vocabulary. It is used to evaluate the extracted descriptors. Typical nodes from different regions are required to compose a powerful codebook.

## REFERENCES

[1] R. Navigli and M. Lapata, "An experimental study of graph connectivity for unsupervised word sense disambiguation," IEEE Trans. Pattern Anal. Mach. Intell., Apr. 2010.

[2] J. Liu, M. Shah, B. Kuipers, S. Savarese, *Cross-View Action Recognition via View Knowledge Transfer*, Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2011.

[3] Daniel Barbar´a, Maurizio Filippone, Detecting Suspicious Behaviour in Surveillance Images, Proceedings of the IEEE International Conference on Computing & Processing, 2008.

[4] G Gordon2.5. Jingen Liu and Mubarak Shah, Learning Human Actions via Information Maximization, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[5] Jingen Liu and Mubarak Shah, Learning Human Actions via Information Maximization, IEEE

International Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[6] Mohamed, F., 2006, Integrated Motion Detection and Tracking for Visual Surveillance Center for Automation Research (CfAR) UoM.
[7] M Lewis, Face recognition using video clips and mug shots, Proceedings of the Office of National Drug Control Policy (ONDCP) International Technical Symposium, 1995.
[8] Alan, J. Lipton, 2000, Moving target classification and tracking from real-time video, The Robotics Institute., Carnegie Melon University.
[9] Collins R. T., 2000, A system for video surveillance and monitoring, Carnegie Mellon Univ., Pittsburgh.
[10] Hu W., 2004, A survey on visual surveillance of object motion and behaviors, IEEE Trans. on systems, man, and cybernetics part C: Applications and Reviews, Vol.34, NO.3, August.
[11] Jung-Me Park, 2001, Fast Connected Component Labeling Algorithm Using A Divide and Conquer Technique, Computer Science Dept. University of Alabama.
[12] Lumia. R., Shapiro. L., 1983, A New Connected Components Algorithm for Virtual Memory Computers, Computer Vision, Graphics, and Image Processing.
[13] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: A local svm approach," in *Proc. Int. Conf. Pattern Recognit.*, 2004, pp.

**AUTHORS**

V.Padmavathi is working as an Assistant Professor / CSE in A.V.C College of Engineering, Mayiladuthurai from December 2008. Previously, she worked as a Junior Programmer for a period of one year in D'zine Garage, Chennai. Now she is pursuing her M.E Computer Science in Annamalai University, Chidambaram. . She published nearly 10 papers in various Journals and Conferences. Her area of interest is Object Oriented Programming, Image Processing and Web Technology.

Dr.M.KalaiselviGeetha is working as Assoc.Prof in the department of computer science and Engineering at Annamalai University, Chidambaram. She published nearly 30 papers in various Journals and Conferences. Her area of interests are Pattern Recognition, Image Processing and Artificial Intelligence.