A Generic System For Automated Detection Of Activities In Surveillance Videos

Abhishek Sinha Ashwin Kumar

June 20, 2012

1 Objective

In a two phase process, we aim to build a system capable of detecting human activities in a video and tag unusual activites in it. In the first phase, we aim to build a generic system capable of detecting human activities across multiple cameras based on a hierarchial model of FSMs (finite state machines).
The salient features of this system are

(a) The hierarchy of FSMs is completely extendible.

(b) The system will be capable of handling queries such as "list of all activities done between two frames", "list of people who performed a particular activity"etc.

The limitations of this system are

(a) It can only handle sparse situations (maximum of 8 persons currently).

(b) It cannot handle occlusions.

(c) At present, our system does not make any use of body movements in activity detection.

(d) The system will be only as good as it is programmed, i.e. the activites will have to be carefully programmed based upon the setup in which it is being used.

To overcome the above limitations and rigidity of a pre-programmed system, in the second phase we will look at unsupervised learning based techniques for activity detection. In this case, activities will be learnt as probability distributions over the lower level feature space and unusual activities will be tagged by virtue of being low probability events over this distribution. Such a model will also have a hierarchial structure with the feature space for the individual activities being the low level movements we get by computing the optical flow vectors across video frames and the feature space for the group activities being the individual activites [1].

2 Approach

In the first step, we divide the set of cameras into subsets. We group a set of cameras in one subset if they are capturing videos in a common plane. We assign a unique plane index for each such subset. Now, for each individual camera field of view, we compute the homography matrix for transforming image coordinates to global world coordinates (here we assume that the world coordinates provided for cameras in the same subset are consistent). Our global coordinates consist of the local plane coordinates and the plane index.

Now, we define a set of ROIs(regions of interest) in each camera's field of view. These ROIs include office doors, entry points, exit points and any other areas which the user may find interesting [2]. Next, we divide the ground plane of each camera into a rectangular grid of locations. Based upon the transitions of a particular person from one location to another in the grid, the physical layer of the hierarchial FSM detects whether that person is standing, walking in a certain direction, entering a certain ROI or exiting from a certain ROI. These informations are passed to the higher levels of FSMs.

In order to track multiple persons in the multi-camera setting, we use colour histogram techinque for person identification. We use opencv's haar detector to detect full bodies of persons, after which we transform the RGB image to LAB image [3, Sec II.B] and then compute the colour histogram of A and B channels (leaving out L channel to remove illumination issues). We store a fixed number of histrograms corresponding to each person. When a person is detected in a scene, we compute its histogram and compare it with the existing histograms of current persons using *bhattacharya* metric. If the match is close enough, we tag it as that person, else we malloc a new person.

As described earlier, the hierarchial FSM module receives the directional information and information about the appearance or disappearance in a particular ROI from the physical layer. This directional information is used to determine the ROI towards which the person is currently headed and acts as a status signal while the ROI information is passed to the hierarchial FSM. The hierarchial FSM has a layered structure in which activities are defined at each level. An activity FSM defined at any level takes the activity status of the predecessor level as its input alphabet (the predecessor of the lowermost level is the physical layer) and transitions are made on the basis of the particular character recieved. For e.g. at the lowermost level, infomation about the presence of a person in a particular ROI will cause all activity FSMs in this level to transition to the state corresponding to the particular ROI if they have this among their next states or transtion to an error state if they don't. The data structure used to represent FSMs is basically an adjacency list structure in which pointers to the next state are stored.

Once an activity gets detected, we store the pointer to the starting frame and ending frame of the video in a mysql database which is later used in answering the queries asked by users.

As is the problem with many other vector-distance based measures, the colour histogram based approach fails with larger number of people. In our testing thus far, it started giving false matches in a sample of 8 people. It would also fail if two people wear very similar colored clothes. The FSM system detected all activities correctly on the very small dataset we tested. On the basis of time measurements we have done, the FSM system will be able to handle a capacity of around 50 people, 4 levels of FSM and 10 activites at each level well within a processing time of 1 ms/frame (does not include image processing module).

In order to improve the accuracy of the lower level system, we plan to use a set of techniques other than colour histogram comparisons such as shape histogram based methods described in [3, Sec II.C].

3 Resources Required

- Four high resolution PTZ type cameras for our testing purposes.
- Ethernet cables of various length.
- Mounting stands for the cameras.
- One way network switches.

References

[1] Tanveer Afzal Faruquie. *Automated Analysis of Surveillance Videos using Latent Topic Models*

[2] Dhruv Mahajan, Nipun Kwatra, Sumit Jain, Prem Karla, Subhashis Banerjee. *A Framework for Activity Recognition and Detection of Unusual Activities*

[3] Zhengqiang Jiang, Du Q Huynh, William Moran, Subhash Challa, Nick Spadaccini. *Multiple Pedestrian Tracking using Colour and Motion Models*