Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 6, June 2016, pg.174 – 181

VIDEO BASED HUMAN ABNORMAL BEHAVIOUR DETECTION USING HOFO AND CLASSIFIERS

Sudhamani M[1], Praveen Kumar Konda[2], Dayanand Gatti K[3]
[1]VI SEM Part-time MTech in Digital Electronics, Sahyadri College of Engineering & Management, Mangalore, Karnataka, India
Sudhasurathkal15@gmail.com
[2]Assistant Professor, Department of Electronics and Communication, Sahyadri College of Engineering & Management, Mangalore, Karnataka, India
Praveen.ec@sahyadri.edu.in
[3]Selection Grade Lecturer, Department of Electrical & Electronics, Karnataka (Govt) Polytechnic, Mangalore, Karnataka, India
dayanandgatti@hotmail.com

***Abstract:** The video based human abnormal behavior detection using HOFO and classifiers are to detect the abnormality in the real time situations. In this project, the human activity prediction problem is solved from a novel sparse representation method. The moving objects are detected first and background subtraction is done followed by classification and analysis. As it is a fully automatic method the problem of manual errors which might happen as in the case of semi automatic surveillance is avoided. Public and commercial security, supervision in banks, business centers, airports, railway and bus stations, private properties are the major areas where "intelligent surveillance system" can be adopted. First, in the video a cuboid is extracted at every point. For the extracted cuboids, the histogram of oriented gradients (HOG) is determined. Also histogram of flow (HOF) descriptors is determined for the same extracted cuboid. Then the two are combined to form a one-dimensional vector. The visual code words are generated from the clustered cuboids using K-Means clustering method. Finally, these code words generate histogram for each cuboid. In this project, video based human abnormal behavior detection is evaluated on the PETS dataset.*

***Keywords:** Histogram of oriented gradients, histogram of flow, Background subtraction, Classification*

1. Introduction

In this project an automatic video surveillance with motion detection, tracking and classification is presented. The moving object recognition is done by the use of adaptive background mixing models. Every pixel is estimated as combination of Gaussians an on-line approximate is considered to revise the model in use. The Gaussian resolution is found out. Later the comparison is done with the backdrop colors to see that which mixture matches with it. Pixel rate that do not match the background scatter are considered foreground until there is a Gaussian plane that comprises them with most, consistent proof supporting it. This system go through to deal toughly with fast changes, common motions of frame elements, tracing through mixed regions, slow-moving things, and fixing or taking off objects from the scene. Segmenting travelling pixels from the constant backdrop of the video, the tracing algorithm trace the found out objects in corresponding frames by using a successive based matching method. It also comes across multi-occlusion cases where some things might be fully changed by the others. It uses 2Dimensional object properties such as position, centroid and size to match analogous objects from different frames. It has color histograms of detected things in order to come across object reorganization after a part of an

occlusion cluster. The tracing algorithm doesn't categorize between objects while tracing, which means the algorithm sees both human and non-human, like a bag, an object, alike. As the initial objective of this project is to find bag abandoning situation, there must be a classification step. The last stage of this whole automatic system is finding of bag, and to find the abnormal condition which is bag abandoning.
The features for classification are aspect ratio and compactness, defined as the ratio of area by perimeter. The determination part, help us in classifying the existence of a human. From this one can decide whether the person is recognized or not. This process is done depending on space constraint and time constraint.

2. Literature Survey

There have been a number of studies about object recognition, tracking, classification and activity investigation in the literature. The survey which is presented here covers only the works that are in the context of this project.
A video processing framework for smart algorithms is used as the base for this project with some of the steps interchanged. This gives a good formation for the discussion throughout this brief survey.

2.1 Moving Object recognition

Each application that gains benefit from smart video processing has different requirements thus needs different treatment. But Moving object is a common thing.
The analysis steps can be simplified by detecting regions which correspond to the moving objects like people and vehicles in the video.
The reliability of the motion detection depends upon the dynamic transformations in nature scene frames such as sudden change in the brightness of light and change in weather etc. From the survey it is understood that commonly used techniques for finding the abnormalities in video frames are optical flow, time differencing, statistical steps and background subtraction.

2.1.1 Background Subtraction

Backdrop subtraction is a commonly used method for motion segmentation in stationary scenes. The average of the frames is found to produce the backdrop. Pixel to pixel comparison is done to find the moving regions. A threshold value is fixed. In the comparison if the difference exceeds this threshold value it is considered as foreground. This generates the frontend pixel map. After this, morphological post processing such as erosion, dilation and closing are done. This reduces the noise and improves the identified regions. The reference backdrop is updated with new frames over the time and comparison is done for the new frames.

There are several approaches for the comparison of backdrop with foreground, and post-processing.

Heikkila and Silven said that a pixel location (x, y) in the current image at that instant is considered as foreground only if there is no equality,

$$|I_t(x,y) - B_t(x,y)| > \tau$$

Where T is a pre-defined threshold.

The background image B_t is updated by the use of a first order recursive filter

$$B_{t+1} = \alpha I_t + (1-\alpha)B_t$$

Where α is an adaptation co- efficient. The idea is to combine the new incoming frame into the present background frame. The backdrop frame is modified faster if the scene is superior. However, it cannot exceed certain limits, which might result in "tails" which are formed behind the moving scene. After creating foreground pixel map regions in smaller size are eliminated with morphological closing
Whenever there is a sudden change in the illumination or any scenes the performance of background subtraction is not very accurate. But it is well at the pixel mining from moving images.
Feature extraction:
Extraction of data from an image is a method known as feature extraction. Lesser depiction of unrefined scene data is called as feature in computer visualization. Hence a good feature should convert the unrefined scene data to a very smaller image. For this, only the required data to be collected out of the unneeded and nondescribing data. The definition of data is different for different feature extraction methods.

2.2 Present methods to find abnormalities

With the stationary cameras, abnormalities in video streams can be found by tracking and a region based methods.

2.2.2 Tracking:

Blob tracking, kernel-based tracking and contour tracking are the different methods used. These methods compute the speed and direction. The classifier indicates the abnormality is present or not. The object is there in the scene or not is found with more frame rates in tracking methods. Hence it is difficult to get better results with tracking systems during a crowded scene with much occlusion.

2.2.3 Region-based:

In videos with smaller frame rate region based method is used.
Consider a video in which two cars crossing a bridge. If we consider the interval between two frames in the video as 10 seconds stacking is not possible because cars could have crossed the bridge faster than this. Therefore region-based technique can be used.
While using optical-flow, to compute the algorithm, the item should be stationary and the frame rate is more.
In region based, each region is given by feature vectors which are used to guide the usual actions. The region and the feature descriptor should discriminate themselves from the standard action. This should be recognized by the classifier which outputs an abnormal event.

3. Methodology

The below listed subjects are used in this project which are explained in the following sections.
- Background subtraction.
- Motion: Optical flow.
- Texture: Gabor filter.
- Probability mass function from samples.
- Kernel density estimation.

3.1 Background subtraction

By doing foreground segmentation, while extracting the feature only the foreground is considered and to achieve this background is to be subtracted. There are several methodologies to do this and many of them have been used efficiently in the computer vision.

Necessary aspects for background subtraction are:
- Real-time presentation.
- Reliable.
- Adaptive to changes in the frame.

As the above methods work on per pixel basis , the camera used should be still.

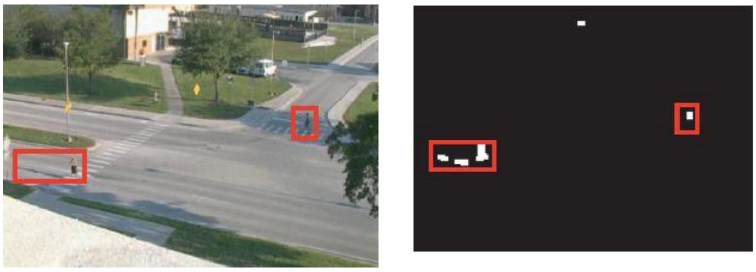


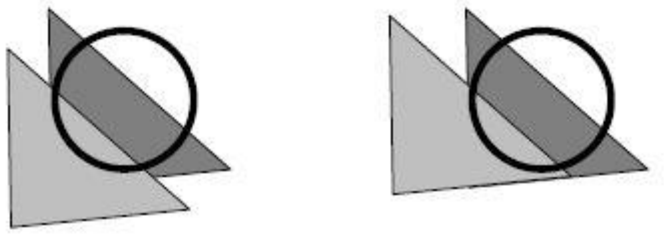
Figure 1. Background subtraction

3.2 Motion: Optical flow

Optical flow is the change in motion of the objects from one video frame to the next frame, depending on the observer. It is possible to use conditions for obtaining an optical flow field, but this is not the same as obtaining the actual motion in the scene. There is a difference between a motion field and an optical flow field as illustrated. The mapping of 3d background to 2D is done in optical flow and it is a rough calculation of the actual movement of the image or frame. Optical flow is often implemented on pixel level and will conclude in a flow field which is a field of vectors that could be used to change one video frame into the next one. Optical flow can in this project both be used to describe speed and direction, two descriptors that are very useful for explaining the behavior in a video.
The different methods used to estimate optical flow are

1. Lucas-Kanade,
2. Pyramid Lucas-Kanade and
3. Horn–Schunck

Aperture problem:
An aperture problem comes because of the movement through the aperture is inconclusive as illustrated in figure 3.3. In the figure the displacement between two triangles is different, but it looks alike through the aperture.



Lucas-Kanade method:
The method was developed by Bruce D. Lucas and Takeo Kanade in 1981.
The following assumptions were made:

- there is spatial coherence between video frames
- The brightness constancy equation.
- the motion in a few neighborhood regions is equivalent

The difficulty in the aperture problem is overcome by assuming that the motion in a small region is the same i.e., u and v are constant in a taken window of m x m. This will lead to an over-decisive. This method also has its own drawbacks. If the matrix A^TA is invertible then only it is possible to solve the least squares problem i.e; have non-zero Eigen numbers. To understand better, some situations are, where the matrix are near to zero, different or high Eigen values.

Homogeneous area:
In an area if each pixel has equal concentration derivative in 2D yields to zero. This makes each scalar in the AT^A matrix null. Hence Eigen values become null and the matrix cannot be invertible.

3.3 Textured areas:
The textured areas in the frame with high difference, the concentration derivative will not be null and hence the matrix can be inverted. Therefore optical flow is calculable with the combination of textured region and Lucas-Kanade method.

3.4 Kernel density estimation

Kernel density estimation is a methodology for reduce the problem of trying to build a population that is based on a finite set. While starting a common histogram there can be high data near two bins. This will disturb the histogram to reveal the actual population. By using the kernel density estimation data i.e, frame is smoothed by a kernel which takes neighboring pixels into consideration. This can be made by using different types of kernels.

Classifier:

K-Nearest Neighbors: The k-nearest neighbor (KNN) is machine learning method for segregating things with respect to the most nearby training datasets in the feature space.KNN algorithm is a kind of lazy learning, in which no training sample is constructed before recognition. Therefore, it is known as the easiest computer learning algorithms.

Given a test data, the distance between the test data and all the training data in the training dataset are estimated based on some distance metric. Hamming distance and Euclidean distance are some usual choices implemented. The distances are used to represent the common between training data and test data. The least the distance is, much nearness is shared between the sets.KNN method determines the k nearest neighbors with the k least distances. The group which wins the maximum voting on the k nearest neighbors is allocated to the class of the training set. When the k is 1, KNN algorithm finds the nearest neighbor and the group of the training set is identified as its nearest neighbor's class. The choice of KNN method is that it is easy and simple to implement. However, when the training dataset is more, the method needs huge memory and the nearness accuracy can quickly less when the number of sets grows. The confusion matrix for the clustered KNN to test the classification exactness of clustered KNN for particular activity. Compared with the performance of some activities of standing, running and sitting the classifier shows bit worse results for walking where it is sometimes determined as standing or running. However, the overall exactness for clustered KNN is nearly 70% accurate considering all actions.

Support Vector Machine: Support Vector Machine (SVM) is the best supervised learning methodology.SVM is initially a binary category. Given a scene of feature vectors with indications such as {-1; +1} SVM aims to predict model and configuration from the guiding datasets and give across the labels of given test datasets. Mathematically, implementation of SVM is to get a individual hyper plane with much margin to segregate the training data sample into dual parts. Optimization problem can be solved by using SVM.

$$min_{w,b,\varepsilon} \frac{1}{2} w^T w + C \sum_{i=1}^{t} \varepsilon_i$$

$$Subject\ to: y_i(w^T \emptyset(x_i) + b) \geq 1 - \varepsilon_i(\varepsilon_i > 0)$$

Example of hyper plane for categorizing point set (SVM)

In this project, we are utilizing a linear SVM model to categorize the human activities. To identify different human activities from multiple activity classes, the single-against multi approach is used to perform multi-category classification.

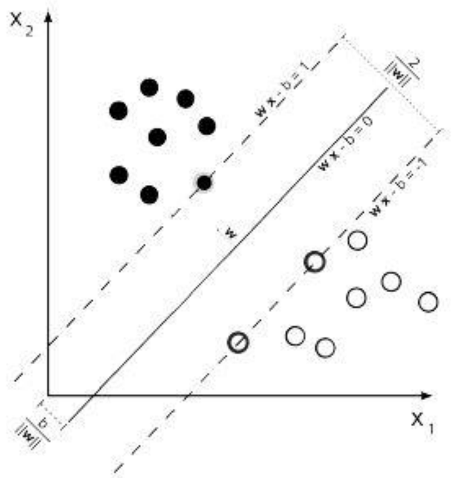


Figure 2 Example of hyper plane for categorizing point set (SVM)

4. Approach

4.1 Introduction to test and results

This section deals with the methods detailed in the design part will be verified. This should be given to what type the methods perform and what the difficulties are. Test will be made on three different datasets. PETS dataset for testing the size, texture and motion classifiers and features. A synthetic database and a real time dataset or any video are for testing feature, direction and classifier created. After every test a small summery will be made regarding the outcome. For the optimal performance the method is not suitable. Because the method makes use of stationary variables which can be used to test the general presentation.

For the PETS dataset the following will be tested for motion classifier and feature at different motion thresholds:

- The Number of histogram bins
- Region size and area.
- The effect of kernel density calculation.
- Optical flow method.

The texture and size classifier will be verified at different thresholds with dedicated Gabor filter parameters and dedicated Gaussian kernel for size averaging. Last an unsupervised training data test and a pixel-level estimation test will be made and the methodology will be verified with other methods using the same data samples.

4.2 The Direction test:

Below test will be conducted for synthetic database:
Incrementing method and Optical flow algorithms.

For the real-time database at the different direction thresholds:

- Region size and area.
- The Incrementing method.
- Optical flow algorithms.

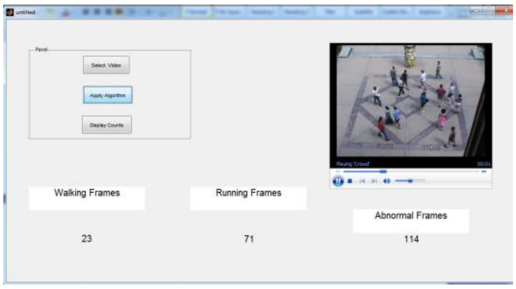
4.3 Dataset for testing motion, size and texture

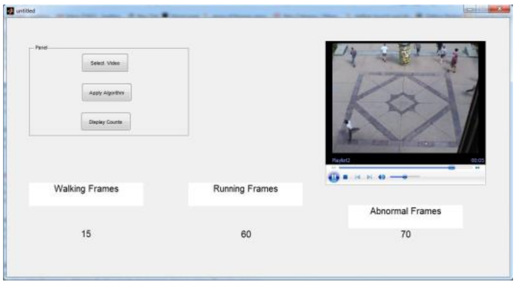
The PETS dataset is created for unusual activity detection. This is used for testing size, motion and texture classification and features. The dataset has two subsets, Ped1 and Ped2. Both data sets have testing and training parts.
The PETS dataset has both moving and texture/size abnormalities. The dataset will be tested upon individually for each classifier. This will make it easy to determine what all methods within the classifiers works good. The tests will be implemented on various parameters, but the interpretation will always be on EER. PED1 data set is used for the motion detection and PED2 dataset is used to detect the size and texture. The classifier will, in two cases, be trained with all the training data sets and tested with every testing data samples. If a video frame in the testing video has one or more abnormal activity it will be decided as abnormal.

4.4 Motion tests

To test the motion classifier and feature, several parameters will be varied and outcome of these will be determined.
The variables that will be tested are:

- Optical Flow algorithm: Standard Lucas-Kanade, Horn–Schunck and Pyramid Lucas-Kanade
- Histogram bins: Amount of histogram bins
- Region size: The area of each region.
- Kernel density calculation: If using kernel density calculation will have an positive effect on the final result.
- Region size: The size of each region.
- Histogram bins: Number of histogram bins.
- Kernel density determination: using kernel density determination of a positive outcome in the output.





4.5 Comparison of KNN and SVM:

Activity	KNN(Frames)	SVM(Frames)
Walking	15	23
Running	60	71
Abnormal	70	114

5. Conclusions

The video based human abnormal behavior detection using HOFO and classifiers are effectively used to detect the abnormality in the real time situations. In this project, the human activity prediction problem is solved from a novel sparse representation method. As more accurate methods like histogram and optical flow are used to detect the abnormality, in real time this algorithm can be used without any delay and errors in the output.

The two classifiers KNN and SVM are compared and it can be concluded that the frame identification and evaluation is more accurate in SVM technique than KNN.

This algorithm can be effectively used as "intelligent surveillance system", wherever the human abnormal behavior detection is required, in real time as well as off the line.

References

1. Ekholm, J., Fabre, S.: Forecast: Mobile data revenue, worldwide, 2010-2015. In: Gartner Mobile Communications Worldwide. (July 2011)

2. Cook, D.J., Das, S.K.: Pervasive computing at scale: Transforming the state of the art. Pervasive and Mobile Computing 8 (1) (February 2012) 22{35

3. Allen, F.R., Ambikairajah, E., Lovell, N.H., Celler, B.G.: Classication of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models. Physiological Measurement 27 (10) (2006) 935

4. Rodrguez-Molinero, A., Perez-Martnez, D., Sama, A., Sanz, P., Calopa, M., Galvez, C., Perez-Lopez, C., Romagosa, J., Catala, A.: Detection of gait parame- ters, bradykinesia and falls in patients with parkinson's disease by using a unique triaxial accelerometer. World Parkinson Congress, Glasgow (2007)

5. Mannini, A., Sabatini, A.M.: Machine learning methods for classifying human physical activity from on-body accelerometers. Sensors 10 (2) (2010) 1154{1175

6. Ravi, N., D, N., Mysore, P., Littman, M.L.: Activity recognition from accelerometer data. In: In Proceedings of the Seventeenth Conference on Innovative Applications of Arti cial Intelligence(IAAI, AAAI Press (2005) 1541{1546

7. Kwapisz, J.R., Weiss, G.M., Moore, S.A.: Activity recognition using cell phone accelerometers. SIGKDD Explor. Newsl. 12 (2) (March 2011)

8. LeCun, Y., Jackel, L., Bottou, L., Brunot, A., Cortes, C., Denker, J., Drucker, H., Guyon, I., Mller, U., Sckinger, E., Simard, P., Vapnik, V.: Comparison of learning algorithms for handwritten digit recognition. In: International Conference on Artificial Neural Networks. (1995)

9. Ganapathiraju, A., Hamaker, J., Picone, J.: Applications of support vector machines to speech recognition. Signal Processing, IEEE Transactions on 52 (8) (aug.2004)
10. Wawrzynek, J., Asanovic, K., Morgan, N., Member, S.: The design of a neuro-microprocessor. VLSI for Neural Networks and Arti cial Intelligence 4 (1993)
11. Anguita, D., Gomes, B.A.: Mixing oating- and xed-point formats for neural network learning on neuroprocessors. Microprocess. Microprogram. 41 (10) (May1996) 757
12. Advantages of surveillance cameras in schools. ttp://www.ehow.com/facts_5615866_advantages-surveillance-cameras-schools.html.
13. Automatic number plate recognition. http://en.wikipedia.org/wiki/Automatic_number_plate_recognition.